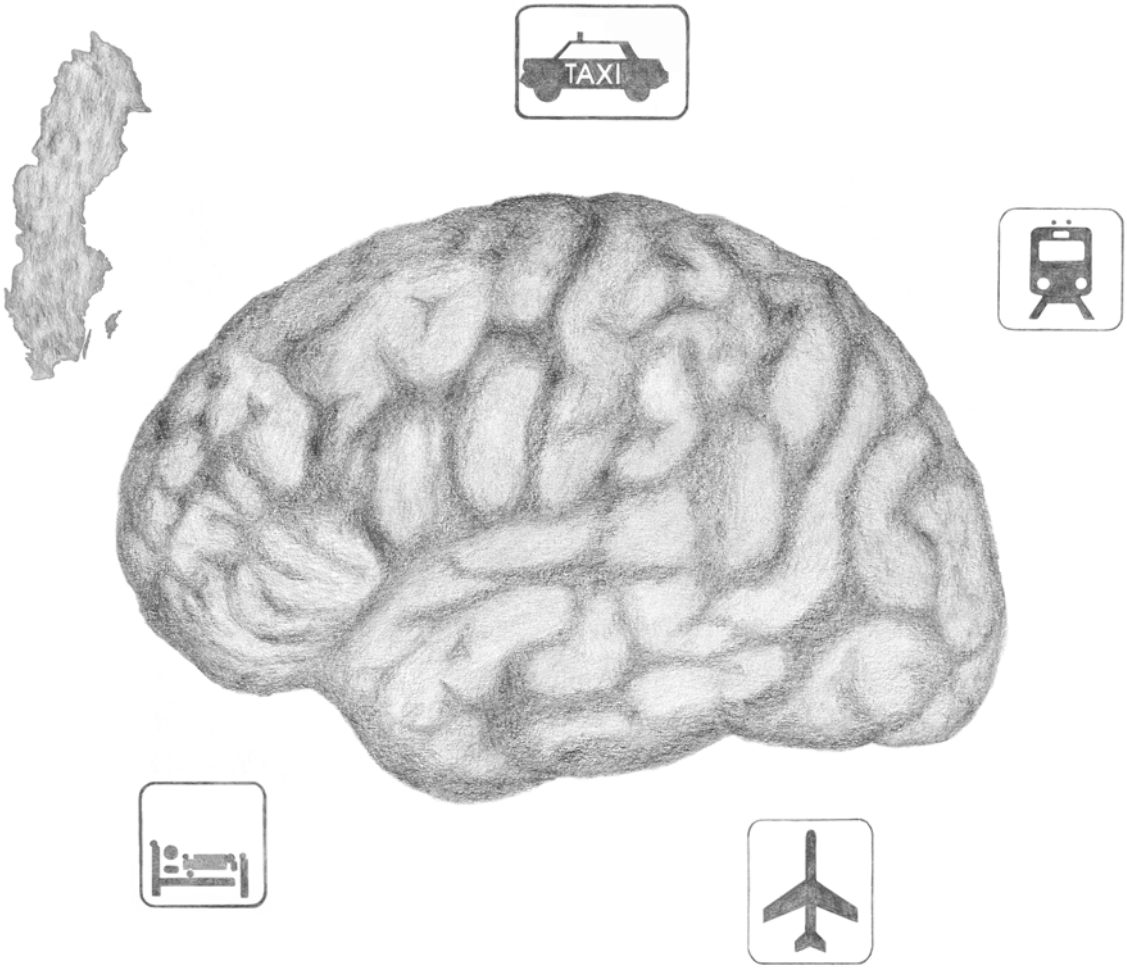
Disfluency in Swedish human–human and human–machine travel booking dialogues



Robert Eklund

Linköping Studies in Science and Technology
Dissertation No. 882
2004

This page intentionally left blank.

Disfluency in Swedish human–human and human–machine travel booking dialogues

Robert Eklund

Department of Computer and Information Science
Linköping Studies in Science and Technology
Dissertation No. 882
2004

Nota bene! Corrected version, different from print version.

Cover illustration: *The brain plans a business trip in Sweden.*
Pencil drawing by the author (Robert Eklund).

Figure 2.1 adapted and published with kind permission by MIT Press.

ISBN 91-7373-966-9
ISSN 0345-7524

Printed by Unitryck, Linköping, Sweden, 2004.

Disfluency in Swedish human–human and human–machine travel booking dialogues

Abstract

This thesis studies disfluency in spontaneous Swedish speech, i.e., the occurrence of hesitation phenomena like *eh*, *öh*, truncated words, repetitions and repairs, mispronunciations, truncated words and so on. The thesis is divided into three parts:

PART I provides the background, both concerning scientific, personal and industrial–academic aspects in the *Tuning in* quotes, and the *Preamble* and *Introduction* (**chapter 1**).

PART II consists of one chapter only, **chapter 2**, which dives into *the etiology of disfluency*. Consequently it describes previous research on disfluencies, also including areas that are not the main focus of the present tome, like stuttering, psychotherapy, philosophy, neurology, discourse perspectives, speech production, application-driven perspectives, cognitive aspects, and so on. A discussion on terminology and definitions is also provided. The goal of this chapter is to provide as broad a picture as possible of the phenomenon of disfluency, and how all those different and varying perspectives are related to each other.

PART III describes the linguistic data studied and analyzed in this thesis, with the following structure: **Chapter 3** describes how the speech data were collected, and for what reason. Sum totals of the data and the post-processing method are also described. **Chapter 4** describes how the data were transcribed, annotated and analyzed. The labeling method is described in detail, as is the method employed to do frequency counts. **Chapter 5** presents the analysis and results for all different categories of disfluencies. Besides general frequency and distribution of the different types of disfluencies, both inter- and intra-corpus results are presented, as are co-occurrences of different types of disfluencies. Also, inter- and intra-speaker differences are discussed. **Chapter 6** discusses the results, mainly in light of previous research. Reasons for the observed frequencies and distribution are proposed, as are their relation to language typology, as well as syntactic, morphological and phonetic reasons for the observed phenomena. Future work is also envisaged, both work that is possible on the present data set, work that is possible on the present data set given extended labeling and work that I think should be carried out, but where the present data set fails, in one way or another, to meet the requirements of such studies.

Appendices 1–4 list the sum total of all data analyzed in this thesis (apart from Tok Pisin data). **Appendix 5** provides an example of a full human–computer dialogue.

Robert Eklund

Linköping 2004

LINKÖPINGS UNIVERSITET

0 Acknowledgements

0.1 Introduction

When finishing a major work like this, it is not easy to decide whom to thank for help, since help is a multifaceted phenomenon. I have been working within linguistics for so many years now, and so many people have taught me things, helped me out, given me insights and thoughts and so on. It is hard to evaluate how much I have learned from different people, but I will try to include *some* of them/you, with the—soon to be obvious—basic stance "rather ten too many than one too few."

First, I would like to say something about the present volume. This thesis contains about a tenth of what I *planned* to include (you should see the list of references that didn't make it), but it (still) probably contains more than ten times more than it *should* include. My first teacher in linguistics, **Peter af Trampe**, pointed out that "nothing can be about everything", something I seemingly have failed to understand, despite my habit of referring to that quote every now and then. However, even if this thesis perhaps should have benefited from even harder pruning, I would like to extend my thanks to those who more or less succeeded in making me cut out some of the things I deemed relevant at the time, and in that way made the present tome much more focused on the issue proper. That being said, here goes!

0.2 First and foremost

There are a small number of people who have been somewhat "extra central" in my involvement in linguistics in general and in the writing of this thesis in particular, and I would like to mention them first, rather than employing the famous "last but not least" algorithm.

I commenced my studies in linguistics with a genuine interest in language, but I am not sure I would be where I am today without two people who exhibited both an extreme interest in the subject proper, but also an interest in getting me involved and engaged in it. Consequently, my deepest thanks go out to **Benny Brodda** and **Gunnel Källgren** (†) at Stockholm University for support and a burning passion for linguistics!

After a couple of years at the Department of Linguistics at Stockholm University, I was hijacked to Telia (then Televerket, now TeliaSonera) by **Bertil Lyberg** to help create the first concatenative synthesizer for Swedish. Later, he offered me full-time employment to work within the *Spoken Language Translator* project, which was a lifetime experience. I have traveled widely with Bertil around the world, and how many times have we not solved the "Riddle of Speech" over a glass of good champagne and single malt whisky. Thanks for boosting my interest in this field!

Thanks to my coach, colleague, co-author, brother-in-arms (at Linköping University), fellow musician and travel companion **Anders Lindström** for much fun, interesting discussions and research collaboration—ranging from glimpse-in-the-eye and tongue-in-cheek to the deeply serious—over the years. Also, during the writing of this thesis, Anders did his best to divert as much of my attention as possible towards the equally fascinating field of *xenophones*, a baby we share. We are eagerly awaiting his forthcoming thesis on that particular phenomenon. Right, Anders?

In December 1997, I was traveling in the USA in connection with the 134th ASA meeting in San Diego. As part of that trip, I included a visit to SRI International in Menlo Park, mainly to see Patti Price, with whom I was collaborating within the SLT project (more about that, and her, later). Patti (to whom I also extend thanks) introduced me to **Elizabeth Shriberg**, whom I already knew about, and had seen giving talks at conferences, but had never spoken with, and I had a consultation talk with her about my work, which then mainly focused on prosodic aspects (and even semantic focus… shiver!), and at that time only included disfluencies as a peripheral part. Our discussion and her advice influenced me profoundly at several levels, both concerning science in general and my thesis topic in particular, which I subsequently steered towards disfluency. Much of this work is the result of that conversation (and other, later talks I've had with her). Liz, my deepest thanks!!!

Finally—although I am not sure that he realizes this—my supervisor **Lars Ahrenberg** at Linköping University has been central, crucial, instrumental and a sine qua non in making this happen. He provided an array of down-to-earth, in-your-face, insightful, astute and distinctly no-nonsense comments that made this work much more stringent than it would have been (if it would have been at all!) without his help. So, Lars, thanks!

0.3 Telia Research[1]

Most of the work presented in this thesis was done under a PhD program contract between Telia Research AB and Linköping University. I would like to thank **Claes Nycander** for signing this contract on behalf of Telia Research AB. Without that signature, things would have been much harder, if not impossible,

Other colleagues who have been helpful over the years are (in a roughly chronological order) **Jaan Kaja**, **Per Sautermeister**, **Mats Ljungqvist**, **Åsa Rydenius** (née Hällgren), **Eva Öberg**, **Camilla Eklund**, **Catriona Chaplin** (née MacDermid), **Joakim Gustafson** and **Linda Bell**.

[1] The company I've worked for has changed its name a couple of times over the years so I don't know how to refer to it. I decided to call it Telia Research AB, since most of the thesis writing was done when I worked at that particular company (circumventing a discussion regarding the distinction between denotation and connotation). It is called TeliaSonera Sweden now.

Also thanks to **Hans Ellemar** for helping in keeping a running Unix environment, which was a prerequisite for me when doing the transcription, labeling and analysis work. During the last months he also helped me to run Unix on a PC, where—to my great surprise—all my Unix shell scripts still worked! (What do you know!)

Warm thanks to **Martin Eineborg** for on-site and off-site fun, and for being a good companion in the gym, where we have been challenging each other to top ourselves in the bench-press over the years—an almost required pastime when spending so many hours crouched-up before a computer.

Finally, I want to extend my deepest thanks to **Ingalill Ankarberg**, **Anita Karlsson** and **Lisbeth Forsberg** at Telia's InformationsCenter (information and library service), who have executed my literature orders swiftly, diligently and skillfully. The luxury of being able to just type in a reference in an email, and a couple of days or weeks later find the article or book (or even microfilm!) in my pigeon hole was an indescribable luxury.

0.4 Linköping University

Of course, first in line to be thanked are **Curt Karlsson** and **Sture Hägglund** at Linköping University for signing the aforementioned PhD contract. Then, sincere and deeply felt acknowledgements are extended to my supervisors **Lars Ahrenberg** and **Nils Dahlbäck**, who at several occasions over the years (seemingly) played "good cop, bad cop" on me (in a special version with two bad cops). Also thanks to my third (co-)supervisor, **Jan Anward**, for providing insights from the field of general linguistics. Thanks to **Lillemor Wallgren** for everything and anything administrative.

Also, thanks the people at the Department of Computer and Information Science, Linköping University, **Arne Jönsson**, **Lars Degerstedt**, **Magnus Merkel**, **Genevieve Gorrell**, **Annika Flycht-Eriksson**, **Pernilla Qvarfordt**, **Lena Santamarta**, **Håkan Sundblad**, **Pontus Johansson**, **Sonia Sangari**, **Mustapha Skhiri** and also **Aseel Berglund** (née Ibrahim, whom I include in that group).

0.5 Stockholm University

Most things have an historical backdrop, and for me, my first course mates, teachers and colleagues in linguistics constitute a decisive factor in making me realize how fun and interesting this field is and for making me want to remain within it.

In no particular order, I extend my thanks to **Janne "Beb" Lindberg**, **Qina Hermansson**, **Carin Lindberg** (née Svensson), and **Malin Ericson** (who all spent some time at Telia), **Don Miller**, **Helen Kåselöv**, **Gunnar Eriksson**, **Britt Hartmann**, **Ljuba Veselinova** and **Eva Lindström**. I also extend thanks to thanks to **Lars Wallin** for discussing disfluency in sign language with me.

Finally, in case you did not know it, the Department of Linguistics at Stockholm University is blessed with the nicest and friendliest reception staff in the world (this is a true fact, confirmed by many, unpublished, scientific studies), and it is impossible not to feel like the 'welcomest' person on the planet when you meet **Cilla Nilsson**, **Linda Habermann** and **Lotta Voulethe**.

0.6 Kungliga Tekniska Högskolan, KTH

I would like to thank **Rolf Carlson** at the Royal Institute of Technology (Kungliga Tekniska Högskolan), Stockholm, for support in arranging the disfluency workshop, *DiSS'03*, *Disfluency in Spontaneous Speech Workshop* in Göteborg, 5–8 September 2003, and for promoting disfluency research in general. Also thanks to **Mattias Heldner** and **Jens Edlund** for discussions and for sundry help.

0.7 Göteborg University

Thanks to **Jens Allwood** at Göteborg University for interesting discussions on disfl… sorry, I mean *Own Communication Management*, and for nice co-organization of *DiSS'03*. While we're talking about *DiSS'03*, I would also like to extend my deepest thanks to **Åsa Wengelin** for being the best co-organizer on this planet, besides being a fantastic friend. Thanks!

0.8 Lund University

I would like to thanks **Petra Hansson** and **Merle Horne**, Lund University, for sharing data with me, and for being helpful in general over the years.

0.9 International

So, let's leave Sweden and turn to the rest of the world.

0.9.1 SRI International, Menlo Park (USA) / Cambridge (UK)

The data on which this thesis is based were all collected during the Spoken Language Translator (SLT) project. Naturally, several people involved in SLT were helpful over the years, and I would like to list some of them. At SRI, Menlo Park, I extend my thanks to **Patti Price**, **Horacio Franco** and **Leo Neumeyer**. At SRI, Cambridge, I would like to thank **Manny Rayner**, **Dave Carter**, **Ralph Becket** and **Ian Lewin** for nice collaboration and insightful discussions.

0.9.2 Conferences

Over the years, I have attended a large number of international conferences where I've met an even larger number of wonderful people who have in various ways made my life richer personally, but who have also cheered me on concerning my thesis work. In a somewhat chronological order I would like to mention **Juliette Waals**, **Saskia Te Riele**, **Laura Dilley**, **Ellen Gerritts** and **Mirjam Wester**, all good friends and great fun! Also thanks to **Michael Kiefte**, **Matthew Aylett**, **Yolanda Vazquez Alvarez**, **Peter Heeman**, **Sieb Nooteboom**, **Rocky Bellini**, **Andreas Stolcke** and **Sherri Page** for hospitality, much fun, interesting discussions and encouragement.

0.9.3 Papua New Guinea

Some of the data mentioned (in the periphery) in this thesis were collected at the Kavieng Airport, New Ireland, Papua New Guinea. I would like to thank the Air Niugini travel agents at the Kavieng Airport, **Loris Levy**, **Nianne Kelep** and **Liza Gabriel** for their kind cooperation.

0.9.4 NASA / Ames, Moffet Field, California

I would like to thank **Beth Ann Hockey** for inviting me (twice) to present my work at NASA/Ames at Moffet Field, California. Also, thanks for comments and proof-reading papers written over the years.

0.9.5 Academia Sinica, Taipei, Taiwan

I would like to thank **Shu-Chuan Tseng** for inviting me to present my work at the Academia Sinica, and to **Yifen Liu**, **Tzu-Lun Lee**, **Ya-Fang He**, **Shu-Huang "Becky" Chiu** and **Yun-Ju Huang** (and all the other students there) for a wonderful week and nice collaboration.

0.10 Thesis-related acknowledgements

While some help has been more indirect in the writing of this work, there are people who have provided direct input, in various ways. I would like to thank the following for this:

0.10.1 ToBI

When I first set up my analysis tool (downloaded from the ToBI site), I received friendly advice and help from **Mary Beckman**, **Gayle Ayers Elam** and **Colin Wightman**.

0.10.2 Statistics

Stats guru par excellence **Per Näsman** was always available to answer my questions on statistical analyses. Another person always willing to try to understand my analysis problems was **Åsa Wengelin**, who helped keep me on track in this respect.

0.10.3 Literature

When the Telia library service (mentioned above) was shut down, **Gunilla Thunberg** at Stockholm University was very helpful in finding articles for me. **Eva Lindström** also helped out in urgent cases, as did **Sara Johansson**, **Elizabeth Shriberg**, **Jens Edlund**, **Petra Hansson** and **Michael Kiefte**.

Also, thanks to my present colleague **Joakim Gustafson**, who—besides being a stimulating sounding board within the applied field—made me aware of interesting work on dialog system development, as well as providing an array of interesting and valid comments on the analyses and results section. Also, every time I thought that I had "closed" the references section, a new, interesting, article was sitting on my desk, and I always knew who put it there—obviously a man of the same ilk as yours truly.

0.10.4 Comments on draft versions

Several people have read draft versions (galore) of this thesis, and subsequently provided valuable comments on sundry parts of this thesis, which caused me to make a lot of and improvements.

I would like to thank the following people, in no particular order:

Mats Wirén provided comments on the structure of this work, as well as pinpointed opinions on wordings here and there. **Johan Boye** reminded me that *soft* AI exists, and was the source of much intellectual stringency (as always), as well as pointing out a few sections where I was possibly leading the readers down the garden path.

Christina Samuelsson and **Janne "Beb" Lindberg** made valuable comments on the stuttering section.

Åsa Wengelin made ever-clever comments on most parts and aspects of this work, besides being helpful concerning *SPSS* menus et simile.

Eva Lindström, always opinionated, provided a bona-fide and breath-taking avalanche of comments on just about everything, ranging from the pixel resolution of the Linköping University logotype (I am not kidding!) to the basic structure of chapters. Her comments resulted in much-needed pruning, and made most paragraphs (or even sentences) in this work more stringent and readable. Also worth pointing out, Eva is the only person I know whose *comments* commonly exceed in quantity[1] the *commented*.

Elizabeth Shriberg read chapter two in its quasi-pseudo-antepenultimate version, and made me much more confident in the not-to-be-taken-for-granted assumption of mine that I was on the right track, and that it was worth reading. For this, I am truly grateful.

Peter af Trampe provided insightful and valuable comments on the speech production section, both concerning methodological aspects as well as philosophical implications.

Thanks to **Martin Eineborg** who made a few but qualitatively crucial comments that saved some sections from disaster. Technical information was also provided by **Magnus Wåhlberg**.

At the very last minute, **Joakim Gustafson** almost drowned me in valid points concerning my results, which prompted me to make amendments and additions. He was kind enough to produce a couple of figures for me, since he—unlike me—is an Excel expert par excellence.

Finally, thanks to **Michael Kiefte** for proofing and final comments—the only guy I know who speaks English, French and Swedish, *and* knows his way around statistics like there's no tomorrow and who performed the task over and beyond the call of duty.[2]

0.11 Sundry and private

First of all, I would like to thank my old and close friend **Kristian Simsarian**, for many stimulating discussions over the last decade, and for being such a good host in the Bay Area. Kristian already has his PhD, and provided a good example. On a somewhat related (Kristian's relatives, that is) note, I would like to thank **Gordon and Carol Laughlin** for inviting me to wrap up my writing in their little guest house in the mountains overlooking Los Gatos. Although I did not actually do that, I will always regret that I didn't. Thanks for being great hosts and for cheering me on!

[1] And oftentimes also quality.

[2] Although Michael told me to blame him, I take full responsibility for whatever disfluency remains in this thesis, be it language, figures, calculations, typos or punctuation.

My parents **Hilding** and **Ingabritt** have always been helpful in miscellaneous ways, and my brother **Roger** and his girlfriend **Maria Holmvall** always provided good company when I needed a break or two (preferably by watching around ten **Simpsons** episodes in a row—the peak of living).

Some people (I can hear you) would say that I have kept them waiting for this thesis to be finished. Well, there are things you can wait even longer for. I want to extend my thanks to luthier **Michael Lowe** of Wootton-by-Woodstock, Oxon, England, who timed the making of my 11-course baroque lute (after Hans Frey) perfectly for me to present to myself as PhD gift. When did I order it? Well, back in 1984. Twenty years. Thanks, Michael! **Jacques Bittner** and **François Dufault**—my favorite French baroque (17th century) lute composers—will finally get the rendering they deserve!

Speaking of which, other prominent musical breaks during nightly writing sessions were also provided by **Johnny Cash** (especially *American III* and *IV*), **Eminem** (*The Eminem Show*), as well as **Andy Williams** (sundry live recordings). Simply breath-taking!

Also, during the last few months when I wrote this up, I did not see huge amounts of living people, being secluded in my home. In fact, the person I probably saw (read: watched) the most was **Jack Lord** in nocturnal reruns of *Hawaii 5-0*. I've read somewhere (I won't provide references here, and by the way, I've forgotten where I read it) that people who watch a lot of TV think they have more friends than people who don't watch TV do, and Jack certainly kept me good company.

0.12 Finally

Thanks to my beloved and super-humanly patience-endowed busy bee **Miriam Oldenburg** and her lovely cats **Sasha** and **Misha**!

Contents

Abbreviations

ASL	American Sign Language
ASR	Automatic Speech Recognition
BP	Bereitschaftspotential
CNS	Central Nervous System
CNV	Contingent Negative Variation
cps	Cycles per second
DAF	Delayed Auditory Feedback
DAT	Digital Audiotape
DPS	Duration Pattern Sequence
EEG	Electroencephalogram
EMG	Electromyogram
EET	Explicit Editing Term
ERM	Explicit Referential Message
ERP	Event-Related Potential
fMRI	Functional Magnetic Resonance Imaging
F_0	Fundamental Frequency
GSR	Galvanic Skin Response
ICM	Interactive Communication Management
L1	Native language
L2	Second, nonnative language
LRP	Lateralized Readiness Potential
NIST	National Institute of Standards and Technology
M	Reported time of awareness of movement

MI	Rolandic motor cortex
MP	Mispronunciation
MSO	Modified Standard Orthography
OCM	Own Communication Management
PET	Positron Emission Topography
PR	Prolongation
REP	Repair
RP	Readiness Potential
RP1	Readiness Potential with associated preplanning
RP2	Readiness Potential without preplanning, i.e. fully spontaneous
SIT	Speech Initiation Time
SLT	Spoken Language Translator
SLT-1	Spoken Language Translator, first phase
SLT-2	Spoken Language Translator, second phase
SLT/DB	Spoken Language Translator/Database
SM	Speech Management
SMA	Supplementary Motor Area
SOT	Slip-of-the-Tongue
SPET	Single Photon Emission Tomography
TMS	Transcranial Magnetic Stimulation
TOT	Tip-of-the-Tongue
TR	Truncation
TTS	Text-To-Speech
UP	Unfilled Pause
VCV	Vowel–Consonant–Vowel sequence
VOT	Voice Onset Time
VRT	Voice Reaction Time
W	Willed (decision to move awareness)
WOZ	Wizard-of-Oz
WOZ-1	Wizard-of-Oz corpus number 1 (1996)
WOZ-2	Wizard-of-Oz corpus number 2 (1997)

List of plates

List of figures

List of tables

PART I

Tuning in…

[S]ound has no independent existence. It is merely a disturbance in a medium.

> Bob Berman. 2004.
> Space: A Very Noisy Place.
> *Discover*, February 2004, vol. 25, no. 2, p. 30.

Once the tongue started moving during speech, it presented a whole new situation with regard to motor control.

> Roger S. Fouts & Gabriel Waters. 2003.
> Unbalanced human apes and syntax.
> *Behavioral and Brain Sciences*, vol. 26, no. 2, p. 221.

'Perfect' fluency and 'normal' fluency are often confused.

> Curtis Tuthill. 1946.
> A Quantitative Study of Extensional Meaning with Special References to Stuttering.
> *Speech Monographs*, vol. 13, p. 96.

[N]o speaker is as fluent as an old mill stream.

> Wendell Johnson et al. 1948.
> *Speech Handicapped School Children*.
> New York: Harper & Brothers Publishers, p. 181.

Fluency has probably received less attention and study than any of the other dimensions and processes involved in verbal communication.

> Martin R. Adams. 1982.
> Fluency, Nonfluency, and Stuttering in Children.
> *Journal of Fluency Disorders*, vol. 7, p. 171

It is normal to be fluent. This is not true of other sequential behaviors. A musician who plays an instrument with the same level of skill that is normal for speech is a very talented and advanced musician. Most human beings become this talented in speech performance.

C. Woodruff Starkweather. 1987.
Fluency and Stuttering,
Englewood Cliffs, New Jersey: Prentice-Hall, p. 11.

[E]rrors do not just happen, but are caused.

John Morton. 1964.
A Model for Continuous Language Behaviour.
Language and Speech, vol. 7, p. 41

Man differs from a linear electronic or mechanical system, however, in that he sometimes varies his standard of relative precision for a movement at the same time as he varies its amplitude.

Paul M. Fitts. 1954.
The information capacity of the human motor system in controlling the amplitude of movement.
Journal of Experimental Psychology, vol. 47, no. 6, p. 390.

Whatever we may want to say, we probably won't say exactly *that*.

Marvin Minsky. 1985.
The Society of Mind.
New York: Simon & Schuster, p. 236.

[T]here is no such thing as 'actual linguistic behavior' which can be accepted unscreened as the empirical basis for linguistic theory.

Jens Allwood. 1976.
Linguistic Communication as Action and Cooperation.
PhD thesis, Göteborg University, Sweden, p. 24.

[T]he need for the future is not so much for computer-oriented people as for people-oriented computers.

R. S. Nickerson. 1969.
Man–Computer interaction: a challenge for human factors approach.
Ergonomics, vol. 12, pp. 515.

To improve speech recognition applications, designers must understand acoustic memory and prosody.

Ben Schneiderman. 2000.
The limits of speech recognition.
Communications of the ACM, vol. 43, p. 63.

It has become generally accepted that a large, perhaps even a major part of our mental activities can take place without our being consciously aware of them.

> Benjamin Libet. 1965.
> Cortical activation in conscious and unconscious experience.
> *Perspectives in Biology and Medicine*, vol. 9, p. 77.

The time is past when philosophers *ex cathedra* could issue naïve views on the nature of knowledge, "brain and mind," reality and appearance, and similar concepts without penetrating the physiological aspects of these problems in detail.

> Lord Brain. 1963.
> Some reflections on brain and mind.
> *Brain*, vol. 86, pt. 3, p. 382.

One has to watch out for the distinction between making a decision response and then being consciously aware of it.

> Benjamin Libet. 1966.
> Brain Stimulation and the Threshold of Conscious Experience.
> In: John C. Eccles (ed.), *Brain and conscius experience. Study Week September 28 to October 4, 1964, of the Pontifica Academia Scientiarum*, Città del Vaticano. New York: Springer-Verlag, ch. 7, p. 178.

But why is it so important to feel that we are in control of our actions when this experience has such little effect on the actual control of action?

> Chris Frith. 2002.
> Attention to action and awareness of other minds.
> *Consciousness and Cognition*, vol. 11, p. 484.

[K]nowledge is not necessarily understanding[.]

> Mark Onslow. 1995.
> A Picture Is Worth More Than Any Words.
> *Journal of Speech and Hearing Research*, vol. 38, no. 3, p. 587.

Certitude propels conversion by the sword, and the defeated must profess the mythologies of the victors. /…/ What is needed, of course, is a strong injection of humility into belief, the skepticism that is the bedrock of science.

> Robert W. Doty. 1998.
> The five mysteries of the mind, and their consequences.
> *Neuropsychologia*, vol. 36, no. 10, p. 1074.

Preamble

This thesis is formally a work within computational linguistics. Consequently, one could, or would, perhaps expect it to be full of formalisms, different kinds of brackets, arrows, box-and-pointer diagrams, flow charts and so on. That is also pretty much the way I started out when entering the field of speech technology around a decade ago when I was "kidnapped" from Stockholm University to Telia Research AB. I spent my first years at Telia doing speech synthesis, speech recognition and speech-to-speech translation, as well as other tidbits like phone set expansion and to some extent face animation. It was a sort of finger-in-every-pie experience. Doing this put me in contact with a plethora of people of sundry backgrounds, a bona-fide cornucopia of knowledge areas thitherto unknown, or at least opaque, to me, which made me realize, and also emphasized, the truly interdisciplinary characteristics of speech technology, and that computational linguistics was so much more than the formalization of grammar rules. When creating systems for human–machine interaction, most things, at most levels, have consequences for most (other) things, at most (other) levels. So, after having spent some of my linguistic "youth", academically speaking, writing tagging formalisms or grammar rules, I've come to consider myself more and more of a speech technologist in general, rather than labeling myself a computational linguist, mainly as an attempt to acknowledge the previously mentioned interdisciplinary trait this field exhibits. If I had to pinpoint (at gunpoint) one area of exceptional importance within speech technology, I would have to mention behavioral psychology, which in a way trickles down through all the nooks and crannies of human–machine interaction at every possible level. From my point of view, this has been, and still is, very rewarding, and very humbling.

This thesis is about disfluencies in human–machine telephone conversations. Few things I've dealt with prior to this have in any way been nearly as interdisciplinary. It is possible to find disfluencies treated in the literature all over the place, from all possible angles and stances. Freud mentions disfluencies. They are studied in stuttering research. Psychologists, computer scientists, engineers, neurologists, physicians, physicists, philosophers, computational linguists, general linguists and phoneticians have all studied disfluencies over the years from different perspectives and for different reasons.

The starting point for writing this thesis was mainly technical, with the more or less explicit objective to enhance the performance of human–machine applications. However, in the process of writing, I found it well-nigh impossible to avoid delving into the core of the phenomenon. Very soon, the burning issue became, what is disfluency. Really.

My personal stance in approaching this problem is similar to what Alphonse Chapanis wrote in a paper in 1971 on the role of the engineering psychologist:

> The starting point for an engineering psychologist is not a deduction from someone's theory or a self-generated hypothesis, but a real-world question, a question such as /… / : What do we need to know to build a computer that would communicate like HAL?[1] (Chapanis, 1971, p. 951.)

My own approach has not been to confirm or rebut a(ny) theory, but instead I have attempted to describe, as objectively as possible (being well aware of that conundrum of objectiveness), the structure of a specific linguistic phenomenon typical of spontaneous, spoken language, which in this work will be called disfluency or disfluencies. Thus, my own "real-word" question would be: "What does disfluency in spoken Swedish human–machine, telephone conversation look like?"

Consequently, what the reader will find here is a three-part book, where the first part introduces the area in general, the second part tries to answer the etiological question, i.e., what disfluency is in a deeper sense (I like clear definitions, or at least attempts to explain what something is about), followed by the third part, an excruciatingly detailed account of how 116 Swedish-speaking people were disfluent in 661 dialogues with what they either *believed* was a machine, actually *was* a machine, or, in a few cases, were human beings. These observations are then discussed in the light of previous observations reported within fields as varied as speech production—with its bearings on consciousness research—stuttering research, speech act theory, linguistic morphology and syntax, cross-linguistic comparisons and so on and so forth. There are issues galore, I can assure you.

Hopefully this book is readable and interesting, and I hope that the reader will know more about disfluency after having read it than they did before, and also that they will find disfluency more interesting after the last page. It is always a basic tenet of mine that "things are never that simple", and putting disfluencies in context will hopefully illustrate how much wider the horizon is than is perhaps evident from the results reported in this work.

Summing up, I have found it utterly rewarding to devote a relatively large chunk of my life to this book, both as regards the new (to me) literature and research I've been exposed to, but also, and perhaps even to a larger extent, the many people I have met all over the world who all, in one way or another, work on the same problem, and who all contribute various bits and pieces of the larger "jigsaw puzzle". Their knowledge, insights, views, opinions and comments have made an already interesting quest so much richer. For this I am very grateful.

Robert Eklund,
Västerhaninge, April 2004.

[1] HAL is of course the conversant, chess-playing, lip-reading (and so on) super-computer featured in the Stanley Kubrick and Arthur C. Clarke film *2001: A Space Odyssey* (Clarke, 1968).

1 Introduction

1.1 Spontaneous speech

Spontaneous speech is indeed a wondrous thing. While written language has existed for perhaps something like 5000 years, humans have been speaking for a much longer time than that, although all figures given are mere conjectures given the elusive character of speech,[1] which makes it go away the very instant we hear it, unless standing in e.g. a cave with a lot of echo, of course. There are even claims that ancient cave paintings and petroglyphs found around the world were made in places where the echo is stronger than at neighboring non-decorated locations, which made speech (and other sounds) linger, which may have been interpreted as the presence of gods or spirits (Waller, 2002). An extraordinary claim is made by Jaynes (1976/2000, 1980), who suggests that humans beings were all "unconscious", in the modern sense of the term, until around the 5th century B.C., and obeyed hallucinated voices produced by the right hemisphere of the brain, something which still occurs in schizophrenics (Jaynes, 1986, 1990; Hamilton, 1985; Frith, 1979, 1987, 1999). The power of these voices is immense, and most often perceived by schizophrenics as "gods", or at the very least, something one should obey. Be that as it may, the sheer power of the spoken word cannot be ignored. It is there, and it influences our lives on a daily basis. Speech "speaks" to us, as it were.

Seen in the light of all this, it is striking how much literate individuals tend to blur the distinction between speech and its written form, thinking that the conventions agreed upon concerning how to represent language in print, in some way represents "true" language. This is ubiquitous in letters to the editor in newspapers or magazines, or in open microphone shows on the radio, where people often voice their extreme concern whenever (other) people "don't speak the way it is spelled!". Alas, would it were that simple!

This thesis is about *spoken* language, for one simple reason. Recent years have seen a boom in launching automatic (computerized) applications. They are all around us, and in the industrialized world, it is more and more common to have some kind of conversation with a computer. The rationale for such systems is of course the assumption that communication

[1] Holloway (1976) believes that language may have begun early in the hominid evolution, "perhaps two to three million years ago" (Holloway, 1976, p. 330). For a more recent discussion on the dawn of language, see Greenfield (1991).

with machines through normal, spoken language is much easier than communication with the help of keyboards or similar artifacts. Indeed, already the first *International Joint Conference on Artificial Language* in 1969 included a paper with the title "Talking with a robot in English" (Coles, 1969).[1] However, although the aforementioned assumption concerns spoken conversation with machines, speech-based automatic systems are mainly rooted in our knowledge of *written* language, once again for a very simple reason: we know much more about written language, as it appears in text-book grammars, and the crux is that the thing closer to us, *spoken* language, is something that eludes us more, something we know much less about when it comes to describing it, analyzing it, or representing it formally.

Another feature of spoken language is that it is very hard indeed to even understand it in writing (which once again emphasizes the point that spoken and written language constitute different modes of conveying language). A good example is given by Pinker (1995) on the Watergate transcripts:

> The Watergate tapes are the most famous and extensive transcripts of real-life speech ever published. When they were released, Americans were shocked, though not all for the same reason. Some people—a very small number—were surprised that Nixon had taken part in a conspiracy to obstruct justice. A few were surprised that the leader of the free world cussed like a stevedore. But one thing that surprised everyone was what ordinary conversation looks like when it is written down verbatim. Conversation out of context is virtually opaque. (Pinker, 1995, p. 224.)

Indeed, it has even been claimed that the reasons we understand each other is not so much the information conveyed in the things we say, but rather the information we "convey" in everything besides speech that is transferred in human–human communication, everything which is not an explicit part of the speech string but is still transferred. This is sometimes called *exformation* (Nørretranders, 1993/1995), and is related to another buzzword term in the area, *world knowledge*. The main reason automatic systems are having problems with human speech, and will continue having problems with human speech, is not so much that they cannot process the speech string proper, i.e. parse and interpret the information embedded in the words as such, but rather that they do not possess any ability whatsoever to interpret the exformation. This is related to the so-called *AI Problem* (for Artificial Intelligence), at least its hard version (e.g. Kurzweil, 1999), and will not be discussed much more in this work, however interesting I find it. Suffice it to say that it is related to the work described in this thesis.

Back to the differences between spoken and written language. Yet another difference between spoken and written language is how editing appears. While for any author or writer (like myself right now), written language provides the opportunity to revise, rephrase, and ponder wordings ad infinitum (modulo deadlines!), before final versions are published, spoken language is by definition real-time and on-line, and once something is said, there is very little opportunity to take it back, however attractive that would be every now and then. Mostly, this is not an obstacle in spoken conversation, but could be problematic when the interlocutor has limited world knowledge, as is the case with young children or current automatic applications (i.e. computerized systems).

And now we are homing in on the focus of this work.

[1] However, the "talking" referred to in this work was using a teletype device.

One of the major differences between spoken and written language is that the former is not so well-rehearsed as we are led to believe when we go to the movies or the theater, read quotes in newspapers, or read novels. In fact, given an over-all figure, some 5% of what we say are things like *err*, *eh*, *uh*, *uhm*, truncated words, restarts, mispronunciations, "editing terms" like *oops*, *sorry*, *no, I mean* and so on. This phenomenon, so typical of spoken language, will in this book be referred to as **disfluency**, but has often been referred to in the literature as *dysfluency*, *nonfluency*, *disturbance*, and *discontinuity*, just to mention the more common terms.

More specifically, this work is about disfluencies in telephone conversations between native speakers of Swedish and what they believed was a computer, or what was in fact a computer, or another human being, also a native speaker of Swedish. Even more specifically, the only thing they talk about is the reservation of business travels in Sweden, including rental cars, hotel reservations and so on. More about that later on.

Recent technological developments have made speech come into the fore in the design of human–computer systems. The rationale for this is, as mentioned above, that speech being the most human of all forms of communication, it should be the easiest, most natural, and quite often most efficient to use, even when communicating with non-animate systems, like computers. Granted, this quest appears to be very much less esoteric than the Jaynesian program, but at the very basis of this approach, this difference might prove to be something of a chimera. Irrespective of whether you try to explain the origins of human consciousness as we know it, or if you simply try to design easy-to-use modern-day automatic human–machine interfaces, observations and decisions tend to trickle down to some form of insight that speech in a very profound way constitutes a very central part in what it is to be human.

1.2 Disfluency: different approaches

Disfluency can be, and has been, studied from different angles and with different objectives. For example, Freud discussed disfluencies from a psychological perspective as something that reveals our inner selves. More recently, cognitive psychologists and psycholinguists like Levelt and Nooteboom have studied disfluencies in order to understand how human speech is produced in the brain. Philosophers like Dennett link speech production to human consciousness in general. Within stuttering research, speech therapists, psychologists and speech pathologists have tried to pinpoint what the difference is between pathological speech, like stuttering (or stammering), and normal disfluencies, typical of all speakers of human languages. Disfluencies have been studied from a discourse perspective, e.g. by Allwood and Clark, who point out that disfluencies should not (always) be seen as a detriment to communication, but instead constitute a linguistic cue or signal that helps structuring conversation between human speakers and listeners, and are thus beneficial both from a speaker and a listener perspective. Disfluencies have also been studied from a gender perspective, linked to body language and gestures, studied from a purely linguistic perspective, analyzed from a phonetic and/or acoustic, or even physiological point of view and, once again, more recently, studied from an engineering or computational perspective, in order to enhance the performance of automatic, or computerized, speech-based applications.[1] These different fields and approaches will be described and discussed in chapter 2 of this thesis.

[1] For references not provided in this chapter, the reader is referred to chapter 2.

One of the things I found fascinating with disfluency is its truly interdisciplinary character. The second chapter, *the etiology of disfluency*, can be seen as an attempt to convey to the reader how much is actually embedded in what is perhaps seen as an ephemeral phenomenon, like saying *eh* every now and then. Although the survey of previous research will be central to the present approach, the motivation has been to widen the horizons for everyone and anyone interested in disfluency, whatever their perspective might be, and to show how results, observations, discussions and hypotheses within other fields could shed additional light on one's own research, in one way or another.

However, let me first briefly outline my own rationale, and goals, for doing this study.

1.3 Disfluency: the approach here

It has long been acknowledged within the linguistic community that disfluencies are more than detriments in the speech produced by human interlocutors. Disfluencies have further been shown to signal a truly magnificent array of different phenomena, ranging from mental state, to conversation planning, physiological or mental stress, and so on and so forth. However, only a small number of languages have been devoted more thorough studies with regard to disfluency, either from a functional or structural point of view. This thesis aims at providing a large such study of disfluency in Swedish. How, you ask? And why would the industry be interested in such an undertaking?

Current automatic speech recognition (ASR) and human–computer dialogue systems have attained a technological level that allows use in every-day commercial applications, at least so long as the tasks are sufficiently constrained, and so long as the users employ fairly "disciplined" speech. In order to allow more open-ended speech input, which assumedly would be so much more attractive and would facilitate use of automatic services, certain phenomena typical of spontaneous speech need to be described, understood and modeled. One such phenomenon is disfluency. To obtain basic knowledge of how disfluencies appear, a first necessary step is to study them in contexts that are *as natural as possible* (although it is possible to study, or even elicit, disfluencies within certain areas of research).

In general, in order to collect application-like data, one needs to collect and study speech data tuned to the particular conversational situation one has in mind. As the domain of the project within which this work was carried out was travel reservations, the focus is on how disfluencies occur in travel reservation contexts. Another reason for focusing on one very specific domain is that keeping that one feature constant, other potential differences as to frequency, distribution, type of disfluency and so on, will not run the risk of being the result of uncontrolled-for causes. Moreover, keeping the domain and channel constant, it is easier to study differences between human–machine and human–human dialogues.

1.4 Scientific goals

This thesis has two main scientific goals. The first is to delve into the **etiology of disfluency**, i.e., what it looks like, where it appears, why it appears and looks the way it does, what the possible causes might be, what the relationship to other phenomena, like stuttering, might be and so on. To date, no detailed such description exists, despite decades of research carried out within a number of different fields. Indeed, the very reason no such synopticon has been written is likely an effect of research being done in parallel, with little or no cross-breeding

between various disciplines. Consequently, this thesis aims to "bring it all together", so that we all know what we are dealing with, at least to some degree.

The second goal is the more specific quest of providing a detailed description of what disfluencies look like in **spoken Swedish**. The focus will be on **description** of disfluencies in Swedish human–computer travel booking dialogues on the telephone, from a sentence, lexical and morphological perspective, with some glances at higher levels, such as *Speech Act Theory*. Since this is the first major work that exclusively describes disfluencies in Swedish,[1] emphasis is placed on **structure**, **categorization**, **distribution** and **comparison** with other, more well-researched languages. Moreover, the observations and conclusions made here will also provide the basis for our understanding of speech as a phenomenon, which is a sine qua non for future model-building and theory creation within the field of human speech production, as well as constituting a solid foundation for the design of more natural automatic human–machine interfaces.

It has been shown that conversation is affected by the channel used (e.g. face-to-face as compared to telephone), as well as the respective roles of the speakers/listeners, which is why this work has tried to keep the channel fixed (telephone) to enable studies of differences between the interacting speakers/listeners, i.e., humans, make-believe computers, and actual computers. Since the domain also affects speech and language, this has been kept fixed throughout this study, although there are some differences between different speech corpora. Given the problems thus outlined, it is only natural that some part of this work will be devoted to **methodological issues** in connection with the analysis of human spoken language phenomena, in particular the problem of corpus collection of spontaneous speech.

The observations made on the Swedish data set will then be compared to the findings reported in the overview. Some space will be devoted to **cross-linguistic comparison**, to the extent that is possible, given the fairly sparse literature on the subject. To study how disfluencies occur within and between languages is important in order to gain insight concerning the deeper levels of human speech production, and is also interesting from a purely linguistic, typological, point of view, in that disfluencies might differ as a function of the type of language in which they appear.

In summary, then, the main quest has been to bring to the fore exactly how widely interesting disfluency is. It occurs as an object of study within fields that on the surface may appear only distantly related, like psychotherapy, (neuro-)philosophy, linguistic speech act theory and human–machine air travel booking. That categorization, crossbreeding of results, hypothesis generation and so on appear and reappear within all these fields, and that observations are often repeated and corroborative in nature within a different array of disciplines is something, in my view, that should be more widely acknowledged. Consequently, a spin-off goal has been to **highlight the interdisciplinary character of disfluency**. To my knowledge, results and perspectives from as wide a range of disciplines have not been brought together in a single volume before now.

[1] Related work has been carried out by Allwood et al. at Göteborg University, Sweden, but the focus of their work is on linguistic function, rather than structure.

1.5 Technological goals

Besides the scientific goals described in the previous paragraphs, there are also technological goals associated with this work. From a rather superficial perspective, a mere distributional description of disfluency occurrence in speech could easily help create language models for speech recognizers or dialogue systems that would enhance their performance in conversation with human users. As will be shown, even basic knowledge of disfluency distribution can be incorporated into speech application systems, both at morphological, syntactic acoustic levels. Thus, one of the goals of this work is to provide a **first basis for improved language models** for inclusion in automatic dialogue systems. As has already been hinted, a down-to-earth consequence of this work will be the facilitation of improved **spoken-interface design** in automatic human–machine services and applications. However, given the limited space, no such practical work will be carried out here, but the data will be there to use, especially seen in the light of how such data have already been incorporated in other systems.

1.6 The contribution

As should be clear from the above, the contribution of this thesis is two-fold. First, this book will provide an extensive description of the **etiology of disfluency**, summarizing research from a wide variety of different disciplines. Second, this work constitutes a large study of the description and categorization of **disfluency in Swedish spontaneous speech**. This division serves the purpose of not presenting Swedish data out of context, but aims at showing how Swedish is similar—or dissimilar—to other languages, and whether the particular data set studied here adheres to or runs counter to previously reported observations. However, it should be pointed out already here that given the veritable cornucopia of different fields and research angles within which disfluency has been the object of study, far from everything that is described in the background chapter will find its counterpart in the results chapter in this work, which will constitute but a proper subset of all possible investigations that could be bestowed this field.

1.6.1 What is covered?

Turning to the Swedish data specifically, the main focus is on **structure**, **categorization** and **distribution** of disfluencies. Although some other areas are described in the backdrop part of this work—notably speech production, psychological studies and discourse functions of disfluencies and so on—the results and analyses presented here will by necessity be limited to a few areas, and will mainly be based on observation of raw data. The main reason is that the data studied in this work are constrained by the way they were collected, and do not lend themselves to all kinds of investigation. Thus, in a way, one could say that the focus of this thesis will be on **mere data**. However, where appropriate, comments and reflections that stray outside structure, categorization and distribution proper will be inserted.

1.6.2 What is not covered?

As is mentioned above, disfluency constitutes a truly interdisciplinary field, and it should come as no surprise only that most of the aspects associated with disfluencies cannot be discussed in this work of limited scope. However, I feel that it is important to mention some of the more important areas that will not be covered in the results and analyses section of this work—although they in some cases are described in the second chapter, in some cases even in some detail—or will only be briefly covered in the etiological section of this thesis. That the

following phenomena are not included does not mean that I consider them irrelevant or unimportant, only that I had to draw the line somewhere, and that the following areas were "included out".

1.6.2.1 Pathology

Speech disfluencies can occur for a variety of reasons of medical nature. Besides the obvious case of stuttering—which will be discussed in some detail—speech can also be disfluent for more obviously pathological reasons, which is the case with conditions like **aphasia**, **cluttering**, **dyslexia**, **spastic dysphonia**, or even **depression**, just to mention a few. These will not be treated in any detail in this work, but will only be mentioned in passing.

1.6.2.2 Interruptions in general

Conversation is also interrupted by a variety of non-speech—or meta-speech—phenomena like **laughter**, **inhalations** (to suck air into the lungs, as distinct from pulmonic ingressive speech, mentioned en passant at sundry places in this work), **coughing**, **clearing of the throat** and so on. These, and similar, phenomena are not discussed at all.

1.6.2.3 "Well, kinda, like, knowhaddamean…"

In some studies, commonly employed words and phrases like *well*, *y'know*, *kinda*, *sorta*, *like* and so on are included in disfluency counts, especially if there is a category **filler words**, or **interjections**, including *eh*, *uh* and *uhm*. Although there is good reason to believe that some of these words may often serve the same linguistic function as e.g. filled pauses, I have chosen not to include them here.

1.6.2.4 Prosody

What is and what is not regarded as disfluent, e.g. from a perceptual point of view, depends to a large degree on the intonational realization of the utterance in question, and work has been devoted to the interaction between disfluencies and prosody. Also, phenomena like prosodic words and prosodic phrases have been shown to play a significant rôle in disfluency production and perception. Since prosody presents additional, and rather different, methodological problems, both from a theoretical and practical (e.g. labeling) point of view, **prosodic aspects** are not covered in this thesis (other than duration proper).

1.6.2.5 Higher-level linguistic phenomena

Although it is clear that most phenomena that appear in spontaneous speech are affected by phenomena like **discourse** realization, **speech acts**, situational **setting**, **channel**, **context**, **interaction** between speaker and listeners and their respective **social roles**, and so on, these fields will only be briefly discussed in this work.

1.6.2.6 Paralinguistic phenomena

Phenomena like **voice quality** (creaky, breathy etc.), **glottalization**, and other meta-linguistic factors will not be discussed, although their occurrence is most probably related to disfluency, and well worth studying within the framework of human communication.

1.6.2.7 Extralinguistic phenomena

Speech is also affected by things like **cognitive load**, amount of **stress**, **social and societal conventions**, **fatigue**, **inebriation**, **biological cycles** and other phenomena not directly linked to normal, spontaneous speech. Some of these will be described in the overview section, but will not be covered in the analyses of the Swedish data, once again since most of this information was not recorded in the data collections.

1.6.2.8 Multimodal communication

It has been known for a long time that **multimodal communication** is different from voice-only communication, and that **facial expressions**, **gestures** and so on contribute to human, or indeed human–machine, message exchanges. This will not at all be covered in this work for the obvious reason that the dialogues studied were telephone dialogues, and thus by definition voice-only.

1.6.2.9 Sundry phenomena

Some areas will not be discussed simply because the present data set does not lend itself to such analysis. This includes **EEG** or **EMG** activity, **galvanic skin responses**, **personality mapping** (of the subjects, or their parents) and so on.

1.7 Backdrop: the Spoken Language Translator project(s)

Ere we commence, some pinpointed project-related acknowledgements need be made. The work on which the present work is based has been carried out as a part of industrial activities at Telia Research AB, Sweden, during a number of years. This means that much work is the results of group efforts, rather than accomplishments by a "lone scholar", in this case the author (me). This could be regarded as an inevitable feature of industrial research, for good or bad (good, methinks). The following paragraphs list (most of) the people involved in different stages of the projects that form the foundation of this work, with the focus on people with whom I interacted personally at various stages of the project. Lest diligent hands go unrewarded.

1.7.1 The Spoken Language Translator

The *Spoken Language Translator* (SLT) was a joint project between **Telia Research AB** (Sweden), **The Swedish Institute of Computer Science** (SICS, Sweden), **SRI International** (Menlo Park, CA), **SRI International** (Cambridge, UK). I will not attempt to list *everyone* that contributed, partly since I am probably not even aware of everyone, e.g. students who helped out during shorter periods, especially those who worked in other countries. However, since this thesis would not have seen the light of the day without SLT, the least I could do is to list the people with whom I collaborated personally during the different stages of this project are listed below.

1.7.1.1 Telia Research AB, Sweden

The project was first proposed by Bertil Lyberg and Ken Ceder, and accepted by Conny Björkvall and Bengt Hagström. In creating the Swedish concatenative synthesizer that was included in SLT, I worked with Barbro Ekholm and Tomas Svensson.[1] Language work (e.g. grammar and translation) was carried out by Ivan Bretan, Johan Boye, Martin Eineborg and Mats Wirén. Work on speech was done in collaboration with Jaan Kaja and Per Sautermeister. Alongside the aforementioned persons, who were all full-time employees at Telia Research AB, a number of hired hands (mostly students) also contributed: People who helped collect Swedish speech data at around 40 locations around Sweden, involved Anita Andersson, Johanna Etzler, Qina Hermansson, Inge Karlsson, Carin Lindberg, Janne "Beb" Lindberg, Jaan Pannel, Curth Svensson and Tomas Svensson.[2] Translation and evaluation work included Anita Andersson, Maria Arnstad, Jens Edlund, Malin Ericson, Beata Forsmark, Nathalie Kirchmeyer, Maria Kronberg, Carin Lindberg, Janne "Beb" Lindberg, Eva Lindström, Tove Mathis, Don Miller, Thierry Reynier, Sara Rydin, Jennifer Spenader, John Swedenmark and Matilda Wernström. Several people helped transcribing the data, either as officially working within SLT, or outside the SLT project proper. These activities involved Eva Holmberg and Carin Lindberg, who also helped evaluate early versions of the transcription tool as I was launching beta versions of it. Beata Forsmark and Rósa Guðjónsdóttir labeled disfluencies in WOZ-1 (Switchboard style), and although I later relabeled the entire corpora according to the scheme described in **chapter 4**, Beata and Rósa provided me with interesting and valuable comments. At a later stage, Annika Asp provided extensive help with the mind-numbingly tedious work of providing orthographic transcriptions of parts of WOZ-1 and WOZ-2 (two of the speech corpora I have analyzed).

1.7.1.2 SICS, Sweden

People at SICS were involved during SLT-1 and the beginning of SLT-2. These included Ivan Bretan (who later moved to Telia Research AB), Björn Gambäck, Mikael Eriksson, Jussi Karlgren and Christer Samuelsson.

1.7.1.3 SRI International, Menlo Park, CA

People with whom I interacted at SRI, Menlo Park, included Harry Bratt, Vassilis Digalakis, Horacio Franco, Leo Neumeyer, Patti Price and Fuliang Weng.

1.7.1.4 SRI International, Cambridge, UK

Among those working at SRI in England, I mainly collaborated with Ralph Becket, David Carter, Martin Keegan, Ian Lewin, Steven Pulman and Manny Rayner.

1.7.1.5 Nyman & Schultz, Sweden

Besides serving as interview victims, Carina Ekedahl and Lennart Svanfeldt at Nyman & Schultz served as the agents in the Nymans human–human corpus.

[1] Not identical with Tomas Svensson[2].
[2] Not identical with Tomas Svensson[1].

1.8 Previous publications

This work draws on previous work that has been published over a number of years, concurrent with the projects listed above. Some of the articles describe how the data studied in this thesis were collected—e.g. **Bretan Eklund & MacDermid (1996)**, **Bretan, Eklund, Kaja, MacDermid, Rayner & Carter (2000)** and **Eklund, Kaja, Neumeyer, Weng & Digalakis (2000)**—while other articles have presented preliminary observations from a disfluency perspective. The first related article (abstract only) was **Eklund (1997)**, which provided a first outline of the labeling method and categories. At the time, labeling and analysis of prosody, focus and GIVEN–NEW information analyses were included. Also, people other than me carried out labeling (above all for scientific and methodological reasons). **Eklund & Shriberg (1998)** compared Swedish and American English human–human and human-machine data, and was among the first articles with focus on crosslinguistic comparison. **Eklund (1999)** could be seen as the seminal version of this thesis, being the first article published on the entire data set, although it was not fully transcribed at the time. As to analyses and results, the present work can be seen as an exhaustive version of the 1999 article. **Eklund (2000a, 2000b)** pursued the crosslinguistic theme, comparing Swedish and Tok Pisin authentic human–human travel booking data. **Bell, Eklund & Gustafson (2000)** compared the Telia telephone data with multimodal data collected at KTH (Royal Institute of Technology, Stockholm). It also included a discussion on the status of unfilled pauses, and an analysis of the rôle of speech acts from a disfluency production point of view. **Eklund (2001)** focused on prolongations, and included a comparison between Swedish and Tok Pisin. Prolongations were analyzed from a distributional point of view, but with regard to phones and location in words. A very short version of chapter 2—*The etiology of disfluency*—is given in the *Preamble* of the *DiSS'03 Proceedings*, i.e., **Eklund (2003)**. Finally, while this thesis was sent to the printer, **Lee, He, Huang, Tseng & Eklund** (on prolongation in Mandarin) was submitted for publication.

1.9 Thesis overview

This thesis is divided into three parts.

PART I—which is just about to end—provided the background, both concerning scientific, personal and industrial–academic aspects in the *Tuning in* quotes, and the *Preamble* and *Introduction* (**Chapter 1**).

PART II—which is waiting around the corner—consists of one chapter only. **Chapter 2** dives into the heart of the problem at focus, *the etiology of disfluency*. Consequently it describes previous research on disfluencies, also including areas that are not the main focus of the present tome, like stuttering, psychotherapy, philosophy, neurology, discourse perspectives, speech production, other cognitive aspects, and so on. A discussion on terminology and definitions is also provided. The goal of this chapter is to provide as broad a picture as possible of what the phenomenon disfluency is, and how all those different and varying perspectives are related to each other.

PART III describes the linguistic data studied in this thesis, with the following structure: **Chapter 3** describes how the speech data analyzed in this work were collected, and for what reason. Sum totals of the data and post-processing method are also described.

Chapter 4 describes how the data were transcribed, annotated and analyzed. The labeling method is described in detail, as is the method employed to do frequency counts. **Chapter 5** presents the analysis and results for all different categories of disfluencies. Besides general frequency and distribution of the different types of disfluencies, both inter- and intra-corpus results are presented, as are co-occurrences of different types of disfluencies. Also, inter- and intra-speaker differences are discussed. **Chapter 6** discusses the results, mainly in the light of previous research. Reasons for the observed frequencies and distribution are proposed, as are their relation to language typology, as well as syntactic, morphological and phonetic reasons for the observed phenomena. Future work is also envisaged, both work that is possible on the present data set and work that is possible on the present data set given extended labeling, but work that I think should be carried out, but where the present data set fails, in one way or another, or meet the requirements of such studies.

Finally, **Appendices 1–4** list the sum total of all data analyzed in this thesis (apart from Tok Pisin data). **Appendix 5** provides an example of a full human–computer dialogue.

PART II

2 The etiology of disfluency

Disfluencies have been studied from many perspectives, and for a wide variety of reasons. In this chapter I will try to summarize the disfluency research carried out within different disciplines over the years, as well as attempt to provide an account for the different rationales for this research. My intention has been to draw the attention to the fact that speech disfluency is a truly multi-faceted phenomenon (or phenomen*a*?), and that it can be studied from almost any conceivable angle.

Although speech disfluency of sorts—especially stuttering—have been mentioned and studied for a very long time, formal study really took off in the 1950s, within three different fields: Within **stuttering research**, Wendell Johnson and his colleagues summarized research carried out since the 1930s and published the first categorization of different kinds of disfluencies, which was then used as the standard for a large number of years. Within **psychotherapy**, George F. Mahl and colleagues carried out extensive research on disfluency as a function of anxiety and developed a first algorithm to evaluate disfluency frequency. Finally, within **general linguistics**, Frieda Goldman-Eisler made extensive studies of pausing and hesitation in spontaneous speech.

However, disfluency has been studied from a much wider variety of perspectives. With the obvious hedge that the ensuing summary must be synoptic rather than detailed, it is my intention to make it clear to the reader that disfluency has bearings on a cornucopia of different disciplines, however far-fetched it may seem at first glance. Also, not all of the included areas deal with etiology *proper* (e.g. application-driven research), but might nevertheless shed light on the causation of disfluency, and are therefore covered.

The main rationale for this chapter is to make the reader aware of the great number of studies that have been devoted to disfluency, as well as the likewise wide range of fields wherein disfluency has been studied. Also, a major point is that despite the great variety of different approaches, the taxonomy of disfluency seems to converge into a fairly delimited set of disfluency categories that most research seemingly, more or less, agrees upon.

2.1 Different perspectives on disfluency

The way I see it, one can differentiate between a few major approaches to the study of hesitation phenomena and other disfluencies in natural language. Of course, the list is neither exhaustive nor even totally fair in all instances. Also, since so much of the research is

overlapping, despite the different perspectives, one could easily have made the division into various fields according to other principles. The following account is just *one* way to do it. It goes without saying that this is clearly not *the* way to do it, but it is my contention (and hope) that it still has a pedagogical value. So, then, how have disfluencies been regarded over the years?

This chapter is divided into an array of different fields, mainly from a **perspective** point of view, rather than a *results* point of view. One of the fascinating things is that so much, completely different, research has yielded the same kinds of results, and led to similar conclusions, even though the research has been carried out for very different reasons, and probably quite often without any (apparent) knowledge about similar research within other disciplines. It must be borne in mind that my division of these areas into separate fields is both necessarily a smidgen ad hoc, and also to a large degree overlapping. This is not surprising, since that although the object of study is, in all cases, disfluency, the rationale for doing the research has differed between the fields. Also, although my main reason for conceiving of the research in the the way described below has been that of different research fields (carried out by research with different competences), there is a **chronological** factor there, too. Thus, both bilingual and crosslingual studies constitute parts of what could be called general linguistics, but such studies appeared much later than the first work on disfluency from a linguistic perspective, whose main object of study was that of slips-of-the-tongue. A final underlying factor is that the fields are somewhat self-defined insofar as the references given within certain fields have been used to delimit what obviously was considered related research by the authors of the papers referred to.

In this chapter, disfluency research within the following areas will be introduced:

- **Stuttering research**
  Research on stuttering has been carried out for decades, and since the objective has been to be able to tell the difference (if there is any) between stuttered and normal speech, stuttering research has often included nonstuttered speech as control groups. Consequently, an enormous amount of work on normal disfluency has been done within this field. Stuttering research will be described in **section 2.2**.

- **Psychotherapy and psychology**
  Formal, exhaustive, research on the role disfluency plays in revealing the inner workings of human mental states began in the 1950s, although spearheaded earlier by Sigmund Freud. From having been viewed mainly as a tool for psychotherapists, it has recently become of central importance with the inclusion of automatic speech recognition in settings where the users can be expected to be under severe mental stress (e.g. fighter aircraft cockpits). An account of psychologically motivated research will be given in **section 2.3**.

- **Physiological factors**
  Speech being a human motor action, like all other bodily activities, it is subject to physiological disturbances that affect performance. A short account for physiological studies on human performance in general, and speech performance in particular, will be given in **section 2.4**.

- **General linguistics**
  After decades of studies of idealized language, the view that performance errors could be interesting in their own right appeared in the 1950s. Consequently, a number of studies of

hesitation phenomena and pausing were carried out, and following the overall trend within general linguistics, with a shift of focus from written language towards spoken language, more and more studies include some kind of nod towards disfluency phenomena. Early linguistic research will be covered in **section 2.5**.

- **Speech production**
  Results and insights obtained within the aforementioned fields, as well as philosophical issues since days of yore, led to the obvious conclusion that disfluencies constitute a means of gaining knowledge of how the human mind, or brain, *makes* language. *How* do we speak? What kinds of processes are at play? What role does consciousness play? Is human language production conscious or automatic? To what extent are we aware of what we are saying, or why, or how? Is speech production simply an example of motor execution in general, or is it in any way special of different? By studying what types of disfluencies that occur, and what disfluencies do *not* occur, one can create models of the inner representation of speech production. By measuring intricate timing relationships associated with disfluencies, one can also make hypotheses with regard to the organization of speech—or more accurately, language—production in the mind/brain. This area exhibits the greatest number of different disciplines, and includes not only general linguistics and stuttering research, but also neuroscience, neurology, psychology, psychiatry, anaesthesia, philosophy, and physics, to mention but a few. Much of this research has far-reaching implications for work carried out within linguistics, and these implications will be discussed in some detail. **Section 2.6** is devoted to speech production.

- **Schizophrenic speech**
  Related, but not equal, to the speech production issues above is what the language of schizophrenics can tell us about what language is, in a deeper sense. Studies have been carried out both concerning hallucinated (covert) and articulated (overt) speech of schizophrenics, which will be described briefly in **section 2.7**.

- **Sign language**
  Speech is not the *only* way humans communicate. Sign languages exist, and provide a valuable source of insight into human language capabilities. Although far less researched than spoken languages, they still have a lot to tell us. Sign language research will be described in **section 2.8**.

- **Application-driven studies**
  With the creation of computers that listen and speak, disfluencies have become of central interest, given the need to incorporate spontaneous speech phenomena in the capabilities of speaking and listening machines. So long as computers need Chomskyan, perfect, reflections of our language competence, they are not likely to be more than marginally accepted or successful. Consequently, a more or less engineering-based perspective on disfluency has seen the light in the last decades. This does not mean that the work is carried out by engineers only, just that the objective within this field has been to formalize disfluency occurrences for inclusion in whatever representation computers employ, both at the acoustic level (speech recognition) or text level (tagging, parsing, semantic analysis). This field will be described in **section 2.9**.

- **Disfluency in a nonnative language**
  Given that disfluencies constitute some of the more common linguistic units in language (the words *uh* or *um* in English make the top-ten list, for example), and given that these are rarely, if ever, taught in language courses, it is of interest to see how disfluencies are

ported between languages, which makes non-native disfluency an area of interest in its own right. This is treated in **section 2.10**.

- **Disfluency and bilingualism**
  Along the same lines, what does disfluency look like in the speech of bilinguals? Given that the so-called filled pause (for example) is not the same "word" in different languages, do bilinguals apply different strategies when speaking different languages? Results from this field are described in **section 2.11**.

- **Crosslingual aspects of disfluency**
  Although it can safely be assumed that disfluency is part and parcel of most human linguistic communication, it cannot be taken for granted that it looks exactly the same in all languages. Disfluency has been studied in a number of languages, but comparatively few studies have explicitly been devoted to crosslinguistic aspects. A few of these will be summarized in **section 2.12**.

- **Gestures**
  Humans do not exclusively communicate by verbal means, but make also use of body language, including hand and arm movements, as well as head nods, gaze and so on. A small number of studies have been devoted to the interaction between spoken language and gesture and head movements, and will be described briefly in **section 2.13**.

- **Disfluency in writing**
  While we have spoken for many thousands of years (despite the variation in the more or less educated conjectures made as to the origin of speech in man), writing has existed only for around five thousand years. Writing, being a motor action, mostly carried out with the arms and hands, reflects language execution in another channel, and is consequently of interest from a disfluency point of view. Writing will be covered in **section 2.14**.

- **Paralinguistic aspects of disfluency**
  Hand in hand with the notion of communication as a greater whole—including gestures and body movements—goes the concept of paralinguistic communication, i.e., the information conveyed besides the semantic meaning we tend to regard as the main objective of language. A few comments on those aspects will be given in **section 2.15**.

- **Disfluency among the elderly**
  While the bulk of stuttering research has focused on the speech of the young, and most other studies mentioned above have focused on young or middle-aged adults, the speech of the elderly has generally been neglected from a research point of view. A few of the existing studies will be covered in **section 2.16**.

- **Effects of disfluency**
  So, given the frequency with which disfluencies occur in all human languages, do they actually matter? Are disfluencies detrimental or perhaps beneficial for language or speech comprehension? Do we even notice them? Should they be regarded as performance errors, or even pathological errors, or do they form a natural part of human language, and actually make language easier to understand and use? This field has been devoted number of studies, which will be accounted for in **section 2.17**.

- **Terminology and definitions**
  Given the number of diverse fields listed above, and given the various amounts of rationales for studying disfluency, it comes as no surprise that there is no *one* set of terms or definitions to be found when discussing disfluency. Instead, although the term **disfluency** is (at the time of writing) the most common (since it was introduced in 1961, as far as I can tell), there are good reasons for criticizing it on the basis of an assumed *fluency*, and the notion that this alleged fluency would sort of be *dis*:ed when speakers of a language utter things like *er* or *uhm*. Of course, the truth is far more complicated than that. Since so much of the work listed above has not been interdisciplinary, there has obviously been very little reason to discuss terminology per se, and as far as I have been able to tell, the only explicit discussion of terminology proper that has been carried out is found within the stuttering community. Terminology—and the associated definitions—both at the top level and regarding finer distinctions, will be discussed in **section 2.19**.

This chapter aims at giving a broad introduction to disfluencies by presenting as many fields as possible where disfluencies play a role, be it small, large or central. The amount of space devoted to these areas does not reflect their relative importance—as if such a thing could be gauged in the first place—or even the amount of work that has been carried out within any particular area, and it is no doubt the case that it is skewed in more than one way. Certain areas include enormous amounts of work (e.g. stuttering research), while other areas are more anecdotal. This is not reflected in this presentation, where the aim has been to include as many different approaches as possible, rather than giving the different areas their justified proportions, as it were. Moreover, no doubt my own interests are surely also reflected in the proportions here given, which may or may not be a good thing.

2.2 Stuttering

Nonfluent speech has been known throughout human history, and it is said that the creator of rhetoric, Demosthenes, initially suffered from a special form of nonfluency, viz. stuttering, which he overcame by filling his mouth with gravel while trying to outvoice the roar of the ocean.[1] Whoever first noticed that some people are less fluent than others, it seems as if at least one fluency disorder has been known in the entire history of mankind: stuttering (or stammering). Despite extensive research on stuttering, researchers still disagree on the causes for it, and I will not take a stand here as to what underlying reasons might be more likely than others. However, what is interesting from our point of view is the enormous amount of research that has been devoted to stuttering. From our perspective, it is interesting since most of the research on stuttering has also been research on normal disfluency. The reason for this is obvious: in order to detect and diagnose stuttering, one needs to be able to tell the difference between what is normal nonfluency in the child, and what is cause for alarm. Consequently, there is a substantial body of research on disfluencies to be found within the field of speech pathology generally, and stuttering research, specifically. Since stuttering mostly appears at an early age, most studies on stuttering have been on children. This means that almost all studies on normal disfluencies in children are found in the stuttering literature.

[1] Johnson and Associates. (1959) mentioned that "such evidence as modern scholarship has yielded appears to indicate that Demosthenes lisped and was concerned with improving his breath control but probably did not stutter as we understand the term" (Johnson et al., 1959, p. 4). D. A. Weiss (1964. *Cluttering*, Englewood Cliffs, New Jersey: Prentice-Hall) cited in St. Louis, Hinzman & Hull, 1985) suggested that Demosthenes was suffering from cluttering.

Another reason for summarizing stuttering research is that any kind of insight into what the possible difference between stuttered disfluencies and normal disfluencies might be, will also be of help in determining the role of disfluencies in spoken conversation.

Perhaps the first question that presents itself when one studies disfluency is how, if at all, it is related to stuttering. Stuttering has been known throughout written history. Stuttering is referred to in the twentieth century BC in hieroglyphics (where the word *nit-nit* is used to describe stuttering) as well as on Mesopotamian clay tablets from the centuries before Christ (Rosenfield & Nudelman, 1987, p. 3). Rosenfield & Nudelman also point out that it is referred to in both the Koran and the Old Testament (Isaiah 28:11). Hippocrates (460–377 BC) included stuttering as one of the phenomena described by the term *trauloi* used to denote articulatory disturbances, and Aristotle (384–377 BC) asserted that stuttering is a defect of the tongue. In the middle of the 19th century, stuttering was still treated by cutting pieces out of stutterers' tongues (Johnson et al., 1948). Galen (131–200 AD) thought that stuttering had many different individual causes, each based on a bad balance between the four elements heat, dryness, moisture and cold (Rosenfield & Nudelman, 1987, p. 3).

So, what is it then? While most laymen are fully aware of the fact that stuttering exists, they cannot produce a good definition of it (Ham, 1990). However, definition escapes even professionals, speech therapists or other (e.g. Ham, 1990, p. 259).

Given the huge amount of disfluency research that has been carried out with the stuttering community, and given that many, if not most, of these studies have included control groups of normal speakers, I will summarize in what ways stutterers and normal speakers resemble or deviate from each other. The present objective is to elucidate the fact that the distinction is not obvious, and that as a result, many findings and observations made in stuttering research are of interest to the study of normal disfluency. For fuller summaries of stuttering research, the reader is referred to e.g. Van Riper (1971/1982), Bloodstein (1969/1987), Starkweather (1987), Silverman (1992), or Alm (1995, in Swedish), to mention but a few.

Finally, the focus of this overview is historical, since the main objective has been to provide an historical backdrop to *disfluency*, not stuttering, research. This means that most of the studies reported are fairly dated, from a stuttering perspective, and I want to stress that the ambition has not been to provide a fully updated summary of the most recent findings within stuttering research proper.

2.2.1 The beginning: Johnson and Associates

While not being the first to study or describe stuttering, one could probably say that the work carried out under Wendell Johnson at Iowa University was the first full-fledged scientific study of stuttering. In a number of publications (Johnson et al., 1948; Johnson, 1955; Johnson and Associates, 1959; Johnson, 1961), a vast number of studies were reported in detail, covering almost all conceivable variables that could in any way be considered to be associated with stuttering. Just to mention a couple, to illustrate the scope of the studies, Darley (1955) investigated parental attitudes, where parents were interviewed concerning e.g. geographical origin, rural–urban background, education, religion, occupation, social adjustment, marital relationship, frequency of hunting, fishing and smoking, and so on and so forth. Other examples are Love (1955), who studied the effect of nembutal and benzedrine on the severity

of stuttering, and Staats (1955), who compared the sense of humor (sic!) of stutterers and nonstutterers.[1]

The set of categories of disfluency that were used by Johnson and colleagues became standard for decades, and were used widely, either wholesale or with minor deviations, by a number of researchers within different disciplines. In Johnson et al. (1948) the following nonfluencies are listed, covering both stuttering and nonfluencies "of the average adult" (Johnson et al., 1948, pp. 180–181):

- Repeated sounds, syllables, words, or phrases.
- Prolonged sounds.
- Pauses.
- Blockages.
- Hesitancies.
- False starts.

Johnson et al. (1959, pp. 134–135) list syllable, word, and phrase repetitions, sound prolongations, silent intervals, pauses, interjections and complete blocks. When used by later researchers, disfluency categories were often referred to as "Johnson's eight categories", as they appeared in Johnson (1961, pp. 3–4), *viz.*:

1. Interjections of sounds, syllables, words or phrases. This category included "extraneous" sounds such as *uh*, *er* and *hmmm*, corresponding to the filled pause of later research.
2. Part-word repetitions.
3. Word-repetitions.
4. Phrase repetitions.
5. Revisions, i.e. instances in which the content of a phrase is modified, or in which there is grammatical modification. This category also included change of pronunciation.
6. Incomplete phrases.
7. Broken words.
8. Prolonged sounds.

As we shall see later, these categories have stood the ravages of time remarkably well. Researchers have differed as to what disfluencies should be included, their respective function, or whether or not they are typical of stutterers, nonstutterers or both, but most of

[1] Since I cannot expose the reader to such a cliffhanger without telling what happened, I can reveal that "[n]o statistically significant differences were found when the median ratings of the nonstuttering control group were compared with those of the stutterers" (Staats, 1955, p. 315).

later categorizations include the above phenomena, although they may be "sliced" in different ways. We will return to Johnson later, but suffice it to say here that the pioneering research carried out during the 1930s, 1940s and 1950s still is of interest, not only for historical reasons.

2.2.2 Loci: the whens and wheres of stuttering

Among the first to point out that stuttering is not randomly distributed in the speech of stutterers was Spencer F. Brown (1937, 1945). In a paper from 1945, he summarized results presented by himself and others in papers from 1935 and onwards, that:

- Certain sounds are more likely to be stuttered than other sounds, mainly consonants, but with wide individual variations as to what particular sounds are problematic, although word-initial sounds are often a major determinant.

- Certain parts of speech are more likely to be stuttered than other parts of speech, *viz*. adjectives, nouns, adverbs and verbs (i.e. words belonging to open word classes).

- The position of a word in a sentence affects the degree of difficulty it presents to the stutterer, the first three words of a sentence being stuttered more often than words occurring later.

- Longer words seem to be stuttered more often than shorter words.

Blankenship (1964) mentioned that stuttering in normal speech occurs more often on lexical words than on function words, thus confirming Brown's results. Silverman & Williams (1967a, 1967b) also corroborated Brown's four points (listed above) for male stutterers (1967a) and male nonstutterers (1967b). Chaney (1969) received similar results for female nonstutterers. Williams, Silverman & Kools (1969b) replicated these findings for school children, and concluded that the four characteristics mentioned by Brown were valid both for adults and children, stutterers and nonstutterers.

Soderberg (1967) reported the contradictory finding than pronouns are more prone to be stuttered. Ellen-Marie Silverman (1974) found that nonstuttering preschoolers were more disfluent on utterance-initial words, pronouns and conjunctions, and argued that this behavior should not be regarded as a sign of early stuttering, but as something typical of young children's speech production in general. Koopmans, Slis & Rietveld (1991) studied spontaneous speech of stutterers, and found that at first-word and second-word positions, function words were more likely to be stuttered, while at third-word positions and later, more stuttering occurred on lexical words.

Hannah & Gardner (1968) argued that a significant factor is whether a linguistic unit appears in post-verbal position, and pointed out that a more detailed syntactic analysis is required in order to describe the loci of nonfluency. Jayaram (1984) studied the role of sentence length and clause position in English and Kannada[1] and found that a clause placed at the beginning of a complex sentence was more likely to be stuttered than the same clause placed at the end of a sentence. This tendency was maintained irrespective of sentence length, which points to

[1] Kannada is a Dravidian language spoken in South India with around 40 million speakers, according to Jayaram (1984). *Ethnologue* (September, 2003), **http://www.ethnologue.com/language_index.asp**, gives the figure 35 million first-language speakers, and 44 million speakers, including second-language speakers.

problems in the motor programming of utterances. No language-dependent effects were observed in this study.

Franklin Silverman (1972) found that while both stutterers and nonstutterers were more likely to be disfluent on long words than short words, this tendency was stronger for the nonstutterers, and that stutterers were relatively more disfluent on short words.

2.2.3 Fluency-enhancing conditions

It has been noted since days of yore that there are several conditions under which stuttering is either reduced or completely suppressed. A brief summary of some of these will be given in the following sections.

2.2.3.1 Sundry studies

Hayden, Adams & Jordahl (1982) tested stutterers' and non-stutterers' speech initiation times (SITs) under pacing and masking and control conditions. Their findings included the following observations: 1. Both groups had faster SITs under pacing conditions. 2. Both groups had slower SITs under masking conditions. 3. Stutterers were always slower than non-stutterers in all conditions.

Stager & Ludlow (1993) tested nonstutterers under the following four fluency-enhancing conditions: 1. Choral reading. 2. Metronome pacing. 3. Delayed Auditory Feedback (DAF). 4. Masking noise. All conditions resulted in significant changes in intraoral pressure and flow. Thus, changes in vocalization occur also in the speech of nonstutterers. Andrews et al. (1982) tested stutterers under 15 different conditions known to reduce or suppress stuttering: 1. Speaking while writing. 2. Speaking with a regional dialect. 3. Singing. 4. Speaking in chorus. 5. Shadowing. 6. Speaking to an animal. 7. Speaking alone. 8. Speaking alone with cards. 9. Speaking while being relaxed. 10. Response contingent. 11. Slowing down of speech. 12. Masking noise. 13. Speaking while swinging the arm; 14. Syllable-timed speech. 15. Prolonged speech during DAF (delayed auditory feedback). They found that all of these conditions reduced stuttering, but to varying degrees. Overall, a 70 percent decrease in stuttering was found, all conditions collapsed into one category. Moreover, all subjects reduced their stuttering under all conditions with the exception of speaking while writing, speaking while relaxed and speaking alone with cards.

2.2.3.2 Reduced reading rates

Adams, Lewis & Besozzi (1973) found an increase in fluency in stutterers when reading rate was slowed down. The observation that slowing down speech production has a beneficial effect on stuttering has been included as part of the explanation of why a variety of other phenomena lead to reduced stuttering rates (see below).

2.2.3.3 Pitch changes

Ramig & Adams (1980) found that both stutterers and nonstutterers spoke more fluently when instructed to speak at higher and lower pitches than normally. Most subjects lengthened both vowel and pause durations at the new pitches, which, they argue, could account for the decrease in disfluency rate.

2.2.3.4 Choral reading

It has been known that reading in unison reduces disfluency among stutterers at least since Bloodstein (1950). Adams & Ramig (1980) examined normal speakers and stutterers in a choral reading condition, and found that stutterers reduced their disfluency to the level of normal speakers under choral readings. Stutterers also evinced other changes under choral reading, e.g. that their vowel durations were shorter under choral readings, as opposed to longer, which is normally associated with a number of fluency-enhancing conditions (*vid.* e.g. Brayton & Conture, 1978). However, the vowel durations of the stutterers were still longer than the durations of the normal speakers, while nonstutterers had higher sound pressure levels than stutterers in all conditions. Since vowel durations were different between stutterers and nonstutterers, Adams & Ramig (1980) concluded that "when stutterers are presumably speaking in their habitual manner, their *fluency* is different from normal speakers" (Adams & Ramig, 1980, p. 468, my italics).

2.2.3.5 Masking noise

Shane (1955) was the first to report that masking noise reduced stuttering. Silverman & Goodban (1972) found that non-stutterers also became more fluent under masking noise, and concluded that the masking noise condition is not an acid test to distinguish stutterers from nonstutterers. Wingate (1970) observed that deaf and hard-of-hearing are underrepresented in stuttering, and reviewed masking noise and delayed auditory feedback and concluded that all fluency-enhancing conditions entail some kind of change in vocalization in the speaker. Garber & Martin (1978) opposed Wingate's view, and concluded that the effect of masking noise is due to reduced auditory feedback rather than change of vocalization, since changes in voice level alone did not have any effect. Brayton & Conture (1978) observed significant reduction in stuttering frequency during noise and rhythmic stimulation, but ascribed the effect to changes in temporal patterning, since vowel durations were increased in both conditions.

2.2.3.6 Delayed auditory feedback

Lee (1950, 1951) employed the term *artificial stutter* for speech under **delayed auditory feedback** (DAF), since it was found that while decreasing stuttering in the speech of stutterers, DAF *induced* stuttering in normal speakers. Neelley (1961) and Garber & Martin (1978) found that changes in vocalization could not explain reduced stuttering, and concluded that decrease in auditory feedback must be the more important factor (cf. the previous paragraph).

2.2.3.7 Adaptation and consistency

Starbuck & Steer (1953) examined fluency in stutterers and nonstutterers during successive oral readings of the same material, and found that both groups increased their fluency in readings of the same passage, but also that this phenomenon, **adaptation**, was "not the same" in stutterers and nonstutterers (Starbuck & Steer, 1953, p. 255). Neelley & Timmons (1967) studied adaptation in stuttering and nonstuttering children (five to eight years of age), and observed that both groups exhibited both adaptation and **consistency** (the phenomenon that stuttering tends to occur on the same words in repeated readings), but that the patterns were odd and hard to interpret at a general level. They concluded that neither adaptation nor consistency can be used as a diagnostic. Williams, Silverman & Kools (1968) examined

adaptation and consistency in stuttering and nonstuttering elementary school children and observed adaptation in both groups, leading to the conclusion that "adaptation is not uniquely a characteristic of the disfluency behavior of stutterers. It is a characteristic also of the disfluency behavior of normal speakers." (Williams, Silverman & Kools, 1968, p. 628). Horii & Ramig (1987), too, examined repeated oral readings in stutterers and nonstutterers, and found that both made fewer reading errors after repeated readings.

2.2.3.8 Self-pacing

Brown et al. (1990) examined how stutterers and nonstutterers performed under self-paced conditions. The subjects were asked to perform three simultaneous tasks: tapping with their finger and open and close the jaw while saying *ah*. All tasks were carried out in three conditions, comfortable, slow and fast. They found that stutterers were slower under all conditions and for all tasks, and also less varied than nonstutterers. They suggested that the movement system of stutterers might be less flexible than the movement systems of nonstutterers.

2.2.3.9 Singing

It has been known for a long time that singing reduces stuttering. Healey, Mallard & Adams (1976) set out to find out what is more important in singing: change of vocalization (Wingate, 1969) or familiarity with the melody/lyrics, the co-called **repeated readings effect**. They concluded that both contribute in and by themselves. Colcord & Adams (1979) also found that the reduction of disfluency during singing was attended by an altered pattern of vocalization that included an increase in voicing duration.

2.2.3.10 Whispering and silent articulation

Perkins et al. (1976) studied adult stutterers in three different speaking conditions: voiced, whispered and articulated without phonation. Stuttering was reduced considerably during whispering, and eliminated during silent articulation. They suggested that the additional problem of coordinating phonatory movement associated with articulated speech, lacking in whispering and silent articulation, might be the cause of stuttering.

2.2.3.11 Metronome pacing

In 1830, Marc Colombat[1] suggested that stutterers should speak in synchrony to a "ticking machine"—originally named the *isochrone*, later the metronome—to reduce stuttering levels. Several later studies have confirmed the validity of this claim. Barber (1940) examined stutterers under a host of different rhythmic conditions, including walking, foot tapping, arm swinging, speaking to a metronome and so on. She found that all rhythmic tasks improved fluency in the speakers. Fransella & Beech (1965) examined speech synchronized with a metronome in order to see whether disfluency reduction could be explained in terms of distraction, i.e. achieving its effect by attracting attention of the speaker to something outside the speaker's own speech, or the alternative hypothesis that the metronome produces the effect by controlling the rhythm of the stutterer's speed of speech. They tested their subjects under different settings: rhythmic metronome, arrhythmic metronome and no metronome. The

[1] Colombat de L'Isère, Marc. 1830 (second edition 1831; third edition 1840). *Du Bégaiement et de Tous les Autres Vices de la Parole Traites par de Nouvelles Méthodes, précédées d'une théorie nouvelle sur la formation de la voix.* Paris: Mansut.

metronome tasks were subdivided into "slow" and "usual" speeds to test whether a speed factor was at play. If the distraction hypothesis is correct, then the arrhythmic metronome should produce the effect as well as the rhythmic metronome. They found that there were significant differences in error rate between the rhythmic metronome and the arrhythmic metronome, but not between the arrhythmic metronome and the control conditions, and consequently ruled out the distraction hypothesis. They also noted a speed effect, but this was seemingly independent of the metronome effect. Therefore they concluded that the rhythmic metronome produced the effect by some other means than slowing down the speech of the speaker. Brady (1969) found that neither slowing of speech, distraction, mode of stimuli (auditory, visual or tactile) or rhythmicity could explain the metronome effect. Hanna & Morris (1977) studied the fluent speech of six stutterers under three metronome-paced conditions, slow, normal and fast, and concluded that the fluency-enhancing effect of the metronome was independent of speech rate, thus rejecting hypotheses claiming that the metronome effect was due to slowing down of speech, confirming both Fransella & Beech (1965) and Brady (1969). Christenfeld (1996) studied the effect of a metronome on fluent speakers and observed "a dramatic effect [i.e., decrease] on the production of filled pauses" (Christenfeld, 1996, p. 1232).

2.2.3.12 Protensity estimation

Stuttering can also be viewed as a disruption in speech timing, which has led some researchers to hypothesize an underlying, more basic, temporal disorder in stutterers. One way to approach this is to have subjects estimate protensity and to distinguish relative duration of tones. Ringel & Minifie (1966) let stutterers and nonstutterers push a button when they thought ten seconds had passed. They found that the stutterers overestimated the duration of ten seconds under all test conditions, and that mild stutterers were closed to nonstutterers than moderate or severe stutterers. They concluded that stutterers are less able than normal speakers to monitor the passing of time, and that this disorder is correlated with the degree of stuttering. More recently, Barasch et al. (2000) examined whether there is a correlation between degree of disfluency and the ability to estimate protensity in twenty stuttering and twenty nonstuttering subjects. They found a positive correlation between disfluency and length of protensity estimates, but also "that whether a person stutters or not is less important as a determining factor in DPS [Duration Pattern Sequence] scores or protensity estimates than whether he or she is more or less fluent" (Barasch et al., 2000, p. 1435).

2.2.4 Disfluency-enhancing conditions

Hand in hand with research on fluency-enhancing conditions, it has also been noted that some conditions create increased *dis*fluency. For example, as was mentioned above, while being beneficial to the fluency of stutterers, delayed auditory feedback increases disfluency in nonstutterers, as noted by Lee (1951). The same goes for **shadowing** (e.g. Cherry, 1953; Cherry, Sayers & Marland, 1955), which increases fluency in stutterers, and decreases fluency in nonstutterers. Consequently, these observations further point to a qualitative difference between stutterers and nonstutterers.

2.2.5 Voice level, the Lombard effect

It has been shown that speakers raise their voices in the presence of background noise, the so-called **Lombard effect** (Lombard, 1911; Lane & Tranel, 1971), or modify their voices in general to accommodate to other, external, sound sources. Howell (1990) examined how

stutterers and nonstutterers adjusted their voices to a set of different conditions, including white noise, delayed auditory feedback, frequency-shifted speech, and speech modified by the *Edinburgh Masker* (Dewar et al., 1979), a throat-activated microphone device which triggers a buzz that is fed back to the speaker by dint of a set of headphones. Howell found that stutterers and nonstutterers responded in similar ways to all stimuli, and discussed the results in the light of auditory servocontrol mechanisms (e.g. Black, 1951), that emphasize the role of auditory perception during speaking. Howell rejected most varieties of auditory servocontrol mechanisms as an explanation for both fluent speakers and stutterers.

2.2.6 Differences between stutterers and nonstutterers

So, as should be obvious by now, many, if not most, of the studies carried out on stuttered speech have also studied normal disfluencies in the form of control groups. This means that there is a huge body of comparative observations, both linguistic and extra-linguistic, and I will just briefly summarize some of the alleged differences noted in the literature. It should be pointed out already here that most of these differences are equivocal in that there are almost always studies that failed to replicate the observations made. That these studies still are of interest is based on the assumption that if there were *no* differences at all, then the research should either exhibit no differences between stutterers and nonstutterers, or "go fifty–fifty". The way it seems, however, is that there are consistent *tendencies*, always in the same direction, even if not *all* studies confirm the said tendencies.

Discussing the alleged differences, Cordes & Ingham (1995) commented that:

> A growing practice divides stuttered disfluencies from normal disfluencies by defining the former as "within-word" and the latter as "between'-word." /…/ a strong form of this definition (that no between-word disfluencies are stuttering and that all within-word disfluencies are stuttering) cannot currently be supported. A weaker form of this definition might prove useful for the definition and measurement of stuttering, but only if such a definition can be both internally consistent and consistent with available clinical and empirical information. (Cordes & Ingham, 1995, p. 382.)

In the following sections, I will briefly list some of the studies that have pointed to differences between stutterers and nonstutterers. Of interest is that some of these studies do not exclusively address speech, but other tasks, such as manual reaction times and so on.

2.2.6.1 Respiratory function

It has been proposed that stutterers suffer from a deficient respiratory function. While Adams, Runyan & Mallard (1974), performing a respirometric study of six normals and six controls, found no significant differences between the two groups as to airflow proper, others have pointed to differences between stutterers and normal speakers regarding other parameters.

Zocchi et al. (1990) found that while normal speakers maintained a constant subglottic pressure during speech production, stutterers were unable to control subglottic pressure, which varied "chaotically from too high to too low" (Zocchi et al., 1990, p. 1510). During fluent periods, the subglottic pressure of stutterers was better controlled.

Johnston, Watkin & Macklem (1993) found that stutterers—during relative fluent stretches of speech—spoke at either higher or lower lung volumes than did normal speakers, and that the former confined their speech to the inspiratory or expiratory reserve volume. While normal

subjects had a gaussian distribution of breath sizes, stutterers exhibited a log-normal distributed distribution. They concluded that stutterers sustain fluency by speaking at abnormally high or low lung volumes, which also shows up as different muscle pattern in stutterers (fluent) speech, as compared to normal speakers.

2.2.6.2 Reaction times differences

It has been shown that stutterers perform worse than nonstutterers on a variety of different tasks—verbal and nonverbal—in a variety of different ways.

Dinnan, McGuiness & Perrin (1970) had 15 stutterers and 15 nonstutterers react to pure tones at different frequencies and amplitude. By measuring galvanic skin response, they found that the stutterers were almost a second slower to react than the nonstutterers.

Cross, Shadden & Luper (1979) examined vocal reaction times (VRT) in stutterers and nonstutterers by asking their subjects to initiate a vowel-sound [ʌ] in response to a tone in either the left or right ear. While no ear-preferences were shown, stutterers were overall significantly slower than nonstutterers, suggesting a general problem in initiation of phonation in stutterers.

Venkatagiri (1981) studied reaction times in the production of voiced and whispered /a/ in stutterers and nonstutterers, and summarizes that:

> The stutterers were approximately 23 msec slower in producing voiced /a/ than were nonstutterers. In contrast, the stutterers were about 11 msec faster in producing whispered /a/ than were the nonstutterers. The stutterers took about 33 msec longer to initiate the voiced /a/, as compared with whispered /a/. The nonstutterers, however, took about the same amount of time to initiate both voiced and whispered /a/. (Venkagatiri, 1981, p. 268.)

Cross & Luper (1979) examined voice reaction times in 5-year-old, 9-year-old and adult stutterers and nonstutterers. Once again, the subjects were asked to initiate the vowel [ʌ]. For both stutterers and nonstutterers, VRT decreased with age, but stutterers were significantly slower than nonstutterers at all ages levels, lending further support to the notion that some kind of laryngeal disorder might be part of stuttering.

Cross & Luper (1983) studied finger tapping and voice reaction times in 5-year-old, 9-year-old and 18-year-old stutterers and nonstutterers. Finger reaction times were consistently slower for stutterers as a group, implying that a more general motor execution problem might be at play in stuttering.

Reich, Till & Goldsmith (1981) measured manual and vocal reaction times of stuttering and nonstuttering adults in different settings. In a manual setting, the subjects were asked to press a button with the right and left forefingers upon hearing a tone. In two nonspeech vocal tasks, subjects were asked to react with either inspiratory phonation or expiratory throat clearing. In two speech-mode settings, subjects were asked to vocalize either a vowel sound or a VCV word. They found significant differences only in the speech-tasks, which strengthens the notion that there are laryngeally related problems in the speech apparatus of stutterers.

Other studies that have observed slower reaction times in stutterers than in nonstutterers include e.g. Hayden, Adams & Jordahl (1982), who studied speech initiation times under

pacing and masking conditions; Adams & Hayden (1976), who examined the initiation and termination of phonation; Adams (1987), voice onset times (VOTs) and durations; Starkweather, Hirschman & Tannenbaum (1976), voice onset times; Starkweather, Franklin & Smigo (1984), vocal and manual reaction times; Dembowski & Watson (1991), laryngeal reaction times; Prosek et al. (1979), manual, acoustic and laryngeal reaction times and Bakker & Brutten (1990), laryngeal reaction times.

Studies which failed to replicate reaction time differences between stutterers and nonstutterers include e.g. Murphy & Baumgartner (1981), voice initiation/termination times; Cullinan & Springer (1980), voice initiation times; Watson & Alfonso (1982), laryngeal and voice onset times, and Long & Pindzola (1985), manual reaction times.

Finally, McFarlane & Shipley (1981) found that there were sometimes differences between stutterers' and nonstutterers' reaction times, depending on the task. However, when there were differences, stutterers were always the slower group.

2.2.6.3 Fundamental frequency

Since stuttering is associated with tension, it has been hypothesized that stutterers should exhibit higher fundamental frequencies than nonstutterers. Healey & Bernstein (1991) did not observe any F_0 differences between stuttering and nonstuttering preschool children. Schäfersküpper & Simon (1983) studied F_0 in stuttering and nonstuttering children (ages ten to twelve), and while there were no differences in fundamental frequency in read speech, stutterers had significantly higher F_0 in spontaneous speech.

2.2.6.4 Neurological differences

Orton (1927) and Travis (1931) were early studies to suggest that stuttering might have neurological causes, and later it has been proposed that stutterers differ from normal speakers in that the speech function is less lateralized in stutterers, which leads to abnormally programmed speech output. Several studies also seem to point to such a difference, some of which will be described here.

Moore (1976) investigated stutterers and nonstutterers for visual half-field (hemi-field) preferences,[1] and found that nonstutterers had a right visual half-field preference for linguistic stimuli, which has been explained in the literature as a function of the more direct visual pathways between the right visual half-field and the language centers in the left hemisphere. Moore did not find right visual half-field preferences for stutterers, while finding a significantly larger proportion of subjects with a left visual half-field preference in the stuttering group than in the nonstuttering group, which lends support to the hypothesis that stutterers are less lateralized.

Kent (1983) concluded that "[a]lthough the evidence is inconclusive regarding anomalous hemispheric asymmetry for speech production in stutterers, evidence favors the proposition that stutterers differ from nonstutterers on tests of central auditory function" (Kent, 1983, p. 250).

[1] In human vision, the left hemisphere processes information in the right hemifield of both eyes, while visual information that appears in the left hemifield of both eyes is processed by the right hemisphere of the brain.

Hand & Haynes (1983) performed a lexical decision task where nonword and real words were presented tachistoscopically to the right and left visual hemifields. Vocal and manual reaction times were measured. Stutterers exhibited a left visual hemifield preference (indicating a right hemispheric preference) and were slower in both vocal and manual reaction times.

Brutten & Trotter (1986) tested nonstutterers and stutterers, matched for gender, handedness and age in a set of dual-task experiments. Subjects were required to tap with their left and right hand fingers as fast as possible, both in a single-task setting (tapping only), in a dual-task setting where the subjects were speaking at the same time, and in a third setting where the subjects were tapping while sounding like a siren. They found that stutterers were slower than nonstutterers across all experiments, but that the decrease in performance between the two groups was similar, which should not have been the case if stutterers were less lateralized, in which case effects in the speech task should have been more apparent in the nonstuttering group. However, it has been shown that weak lateralization per se is often associated with poor motor performance. They concluded that lateralization remains an open question, but that their results imply a poorer neuromotor system in general in stutterers.

Greiner, Fitzgerald & Cooke (1986) examined hemispheric functioning in an interference study, i.e. how carrying out different activities simultaneously led to performance deterioration. Subjects (stutterers and nonstutterers) performed four experimental tasks: tapping, tapping–spontaneous speech, tapping–reading and tapping–singing. Stutterers showed more interference than did nonstutterers, and the tapping–spontaneous speech condition resulted in the greatest amount of interference. They concluded that stutterers' speech production is influenced both by intrahemispheric competition and interhemispheric integration processes.

Moore & Boberg (1987) reviewed the literature on neurological differences between stutterers and normal speakers, and concluded that:

> [T]here appears to be compelling evidence from many studies, using a wide variety of investigative techniques, that there are differences in CNS [central nervous system] functioning amongst stutterers. Data from well controlled dichotic, EEG, blood flow, tachistoscopic, sequential finger tapping and Wada technique[1] studies show that stutterers typically do not use primarily left hemisphere strategies to process language as do normal speakers. Rather, most stutterers use primarily right hemispheric, or greater bilateral strategies in processing language. (Moore & Boberg, 1987, p. 31; my footnote.)

Rosenfield & Nudelman (1987) evaluated neurological models of speech dysfluency. They pointed out that stutterers and normal subjects alike "do not speak with their mouths; they speak with their brains" (Rosenfield & Nudelman, 1987, p. 5) and that stutterers have abnormal speech-motor output. Consequently, they continue, "[a]ny model that purports to explain this phenomenon, regardless of its orientation, must address how it is that the brain produces these dysfluencies" (Rosenfield & Nudelman, 1987, p. 5). However, unlike Moore & Boberg (1987), they are more critical of the results found in the literature, and their conclusions are more cautious as to whether or not any hemispheric differences exist between stutterers and non-stutterers.

[1] The **Wada technique** (Wada & Rasmussen, 1960) is a method where a short-acting barbiturate is administered in one of the hemispheres of the brain. Patients thus become hemiplegic in the contralateral side of the body, relative to the hemisphere where the injection is done. While normal speakers lose their ability to speak as a result of a left-hemispheric sedation (but not right), it has been shown in some—but not all—studies that stutterers lose their ability to speak irrespective of which side is sedated.

Another, specific, neurological difference is that of **alpha rhythm activity**.[1] At rest, the brain exhibits waves referred to as the alpha rhythm. Its frequency is normally between 8 and 13 Hz, and is strong in amplitude, but diminishes in areas known to be associated with speech production just before a person speaks (Starkweather, 1987, p. 8, referring to Linebaugh, 1975.[2]). However, as Starkweather points out:

> This rhythmic [alpha wave] activity is typically present when the brain is relatively inactive and disappears when it is engaged. Just before a person begins to speak, the alpha wave disappears, indicating presumably that the brain is formulating language. The inference, although logical, is rather far from the observation. The EEG observation only tells us that the brain is relatively busy, not what it is doing. (Starkweather, 1987, p. 219.)

As is the case concerning most other parameters, there are studies both indicating that there are alpha rhythm differences between stutterers and nonstutterers, and that there are no such differences to be found (*vid.* e.g. Van Riper, 1971/1982, pp. 347–348, pp. 377–378 and p. 417), but there are enough studies suggesting a difference to assume that alpha rhythmic patterns differ between stutterers and nonstutterers (Moore & Boberg, 1987, p. 22 and p. 29).

Stromsta (1964) studied bilateral EEG potentials in the alpha range in 15 stutterers and 15 nonstutterers, and found significant differences between the groups:

> This would indicate that the total power of the frequencies common to the bilateral brain potentials of the stutterers did not differ significantly from the total power of the frequencies common to the bilateral brain potentials of the nonstutterers. However, there was a significant difference in the distribution of power as a function of frequency for the two groups. The latter point was evidenced by a concentration of power at 10 cps (sharp tuning) for the stutterers as compared to a lack of such concentration of power (broad tuning) for the nonstutterers. (Stromsta, 1964, p. 419.)

Moore & Lang (1977) found a reduction of alpha activity in the left hemisphere of nonstutterers prior to speaking, while most stutterers showed a reduction of alpha in the right hemisphere, which "suggest[s] right hemispheric processing for the stuttering group" (Moore & Lang, 1977, p. 223). Moore & Haynes (1980) measured alpha activity in male stuttering and nonstuttering subjects and found that the stutterers processed both speech and tones in their right hemisphere, whereas the nonstutterers showed equal hemispheric activity.

To conclude this section, Travis (1978) summarized five decades of cerebral dominance theory, and observes that several new research methods have been developed since 1931 that have enabled neurological studies that support the theory. More recently, Fox et al. (2000) performing a PET study of stutterers and nonstutterers, conclude that their findings "support long-held theories that the brain correlates of stuttering are located in speech-motor regions /…/ especially of the non-dominant (right) cerebral hemisphere (Travis, 1978), and extend this theory to include the non-dominant (left) cerebellar hemisphere. The present findings also indicate a specific role of the cerebellum in the fluent utterances of persons who stutter." (Fox et al., 2000. p. 1992.)

---

[1] See Zeman (2001, p. 1267) for a description of alpha, beta, theta and delta waves in the brain.

[2] C. Linebaugh. 1975. *Interhemispheric asymmetries in the contingent negative variation and cerebral dominance for speech production*. PhD thesis. Temple University, Philadelphia. Early work showing hat EEG/alpha activity is indicative of left hemispheric processing of linguistic material—as opposed to spatial or musical stimuli—was also shown in Galin & Ornstein (1972), McKee, Humphrey & McAdam (1973), Callaway & Harris (1974), Dumas & Morgan (1975) and Galin & Ellis (1975), among others.

Which brings us into the—allegedly—fluent speech of stutterers.

2.2.7 Fluent speech in stutterers?

Several studies have shown that ostensibly fluent speech produced by stutterers still exhibits patterns that are not found in normal speech, and that the differences are often enough to enable listeners to differentiate between stutterers and nonstutterers.

Healey & Gutkin (1984) studied voice onset time (VOT) for voiced stops and fundamental frequency changes for voiceless stops in fluent speech of stutterers and nonstutterers, and found that there were significant between-group differences in that stutterers have slower VOTs and wider fundamental frequency ranges.

Love & Jeffress (1971) studied fluent speech of stutterers and nonstutterers and found that fluent speech of stutterers contained significantly more brief pauses (150–250 ms) than did fluent speech of nonstutterers. These pauses are often imperceptible to the human ear, but indicate that there are differences between perceptually fluent speech of stutterers and fluent speech of nonstutterers. Love & Jeffress suggested that these brief pauses may partly explain why stutterers exhibit a higher tendency to judge their own speech as disfluent than other listeners, in that they alone are aware of the high incidence of brief pauses that are imperceptible to everyone one else.

Adams & Runyan (1981) compared fluent speech of stutterers with fluent speech of nonstutterers, and found that stutterers had longer vowels and more variable fundamental frequency. Physiologically, they were slower in starting phonation in the transition from voiceless to voiced speech sounds, and they spoke with an excess of air pressure above the glottis. They also reviewed the literature on whether or not listeners are able to differentiate between fluent speech of stutterers and nonstutterers, and report equivocal results. It seems, then, that fluent speech of stutterers can be imperceptibly different from fluent speech by nonstutterers, but that sometimes listeners are still able to tell that the speaker is a stutterer. Adams & Runyan (1981) performed their own experiment, and tentatively concluded that fluent speech of stutterers is perceptibly different from that of nonstutterers. They pointed to phenomena such as the difficulty stutterers exhibit in starting phonation, lapses that are not so big as to disturb the normal flow of speech, but big enough to affect voicing where voicing is due, resulting in voiceless versions of phonologically voiced sounds—or vice versa.

Zebrowski, Conture & Cudahy (1985) compared temporal parameters for word-initial /p/ and /b/ in fluent speech of stutterers and nonstutterers, and found that the former exhibited an inverse relation between stop-gap and aspiration duration that did not appear in the nonstutterers' speech. They attribute these finding to difficulties affecting the relations between laryngeal and supralaryngeal behaviors in the stutterers.

Peters, Hulstijn & Starkweather (1989) examined acoustic and physiological reaction times in stutterers' fluent speech, compared to nonstutterers' fluent speech, and found that stutterers performed slower, overall. Since this effect was located at the beginning of utterances, and especially prominent for longer utterance, they suggested that stutterers may have problems in the motor programming of speech.

As is most often the case, there are also studies that fail to replicate any differences between stutterers and nonstutterers. Conture, Colton & Gleason (1988) examined onsets, offsets and

durations of respiratory, articulatory and laryngeal behaviors of fluent speech of stutterers and nonstutterers, and found no significant differences. Likewise, Borden, Baer & Kenney (1985) found that voice onset times of stutterers' fluent speech were "well within normal limits" (Borden, Baer & Kenney, 1985, p. 371).

However, Bloodstein (1969/1987) summarized his review of fluent speech in stutterers thus:

> [T]he weight of the evidence strongly suggests that what observers consider to be the fluent speech of stutterers frequently reveals features on careful study that are not to be found, at least not in the same degree, in the community of nonstutterers. Although the precise extent of these differences is not yet fully clear, most of them appear to entail aspects of slowness or limitation of movement, lateness of response, or incoordination of the vocal apparatus. Many of the abnormal features of stutterers' "fluency" appear to bear a broad resemblance to those of overt stuttering. (Bloodstein, 1969/1987, p. 31)

In summary, as was previously mentioned, it could be argued that if there were no differences between the fluent speech of stutterers and nonstutterers, then the studies that report that listeners are indeed able to tell apart stutterers' (fluent speech) from nonstutterers, always in the same direction, would be hard to explain.

2.2.8 Developmental factors

Given that stuttering most often appears during childhood,[1] it comes as no surprise that the studies published by Johnson et al. (1955) include research devoted to developmental factors, e.g. Branscom, Hughes & Oxtoby (1955) and Egland (1955). Most of children who stutter exhibit spontaneous recovery,[2] and cease to stutter when they grow up. Consequently, it is of course also of interest to try to find out whether there are signs that would predict whether recovery will occur or not, especially since there is general agreement within the stuttering community is that there is no *cure* for adult stuttering, although different therapies can improve stuttering to varying degrees.

To summarize the huge body of research on developmental disfluency would be an overpowering task, so I will just briefly mention a few studies in this section to provide the reader with a feel for what kind of studies can be found. Basically, studies have either focused on speech development of children in general, i.e. without any diagnosis of stuttering, or have focused on comparison of children who are diagnosed as stutterers and children of the same age and/or gender who do no stutter (although there are studies entirely devoted to stuttering children exclusively, of course).

Yeni-Komshian, Chase & Mobley (1968) examined delayed auditory feedback (DAF) in children between two and three years of age. They concluded that auditory feedback monitoring system is operative at this age, but also marked stronger DAF effects in the older children.

---

[1] Adams (1982), to take just one example, states that approximately 75% of all stuttering develops between two and seven years of age.

[2] Bloodstein (1969/1987, p. 96) stated that between 36 and 79 percent of those who at any time begin to stutter recover spontaneously. Bloodstein also pointed out that many of those who recover "may bear a certain risk of developing stuttering again in later life" (op. cit., p. 99). Franklin Silverman (1992) gives an example of a stutterer who relapsed into stuttering after 28 years of speaking fluently (Franklin Silverman, 1992, p. 108).

Around the second birthday, the speech of the typical child exhibits around 5% disfluency. However, as Adams (1982) pointed out, there are major differences between the adult speaker and the child, e.g. that hemispheric lateralization is normally not completed until ten years of age, neurons in the cortex have not yet attained their full size, that dendritic growth has only begun which leads to a limited number of functional interconnections, that the central alpha rhythm is of relatively small amplitude in the child and so on. Thus, from a neurological point of view, the child is in many ways a very different speaker than the adult.

So, what should one look for? What is the definition of childhood stuttering? Beginning with fairly recent work (as opposed to e.g. the pioneering studies by Johnson et al.), Conture (1990) defined stuttering as "any within-word speech disfluency, for example, sound/syllable repetitions, sound prolongations, broken words, and so forth" (Conture, 1990, p. 2), but also mentioned that "there is considerable overlap in the number of between- as well as within-word disfluencies of children considered to be normally fluent and those considered to be stutterers, especially during early childhood" (ibid., loc. cit.). Conture further pointed out that:

> [T]here are *no known objective, listener-independent* criteria for identifying instances of stuttering or classifying children as stutterers versus normally fluent speakers /…/ there *is no consensus* among experienced clinicians and researchers regarding behavioral definitions of stuttering in childhood or classification of children as stutterers. (Conture, 1990, p. 3; italics in original.)

Conture (1990) concurred with previous proposals that an overall frequency of 10% disfluency or more should be regarded as a sign of children at risk for stuttering, but also proposed that 3% or more of *within-word* disfluencies constitutes a useful metric (normally fluent speech, according to Conture, contains 1%, or less, within-word disfluency). Consequently, mere rates are not enough, but rates of *specific types* of disfluency should be looked for.

Onslow et al. (1992), using Johnson's eight categories, presented speech samples of stuttering and nonstuttering children aged 2–4 years to clinicians and laymen listeners, and found that high rates of agreement concerning the classification of who were stutterers did not coincide with any of the categories employed. They suggested that, in order to describe early disfluency, single categories should be replaced by multiple categories, that more categories should be used, and that nonverbal speech events should be included in the description of the data language.

2.2.8.1 Children who do no stutter

So, what kinds of studies have been done? Starting with studies on children with no previous diagnosis of stuttering (or any other speech disruption), Yairi (1981) found no sex differences in the speech of two-year-olds. He also found significant individual differences, and even children who were disfluent only infrequently. The most common type of disfluency was the repetition of short segments, one syllable or less.

Ellen-Marie Silverman (1973a, 1973b) and Colburn (1985) both pointed out that disfluencies of children tend to appear in *clusters*, i.e. more than one disfluency per instance. Colburn (1985) concluded that clustering of disfluency is normal in speech of children from the time they begin to talk in sentences.

Schuckers & Lefkov (1979) investigated whether normal-speaking children (with a mean age of 7 years and 9 months) could perceive misarticulations in contextual speech. The first task was to recognize sentences that contained misarticulated words. The second task was to identify a specific misarticulated word within a sentence. The children were able to successfully identify misarticulations in both tasks, but were less successful in identifying [θ]/[s] substitutions than [w]/[l] or [t]/[k] substitutions. They were also significantly better at identifying misarticulated consonants when they appeared as singles than when they appeared in clusters.

DeJoy & Gregory (1985) compared a group of 3.5-year-old and a group of 5-year-old nonstutterers and found that the younger group evinced significantly more repetitions (part-word, word, phrase), incomplete phrases and dysrhythmic phonations, while the older group exhibited more grammatical pauses. The groups did not differ as to ungrammatical pauses or interjections (filled pauses).

Cecconi, Hood & Tucker (1977) investigated disfluency in the oral readings of children from grades 3 through 6. They found that disfluency rates went up as a function of the difficulty of the reading material, and that stuttering disfluencies (part-word repetitions, dysrhythmic phonations and tense pauses) were more prone to increase than normal disfluencies. They also found that fourth-graders were the most disfluent, something they attributed to the fact that this is the age where reading is not activity by itself, but becomes a tool for learning in general, and thus introduces the notion of content, which makes it a more complex activity.

Wexler & Mysak (1982) compared 2-, 4- and 6-year-old males and found only minor age-related differences. Incomplete phrases were the most common type in all age-groups, while part-word repetitions were the least occurring in the 2- and 4-year-old groups, and dysrhythmic phonations was the least frequent in the 6-year-old group. Wexler (1982) compared 2-, 4- and 6-year-olds in a neutral and a stress-situation. Again, 6-year-olds exhibited fewer disrhythmic phonations that the other two groups, but the only significant difference was that the 2-year-olds had more word and phrase repetitions than the other groups in the neutral situation.

Kools & Berryman (1971) compared male and female first-graders and found no significant overall differences, although males produced more incomplete phrases than did females.

Gordon & Luper (1989) compared 3-, 5- and 7-year-olds, and found that disfluency rates went down as a function of age, in that 3-year-olds were significantly more disfluent than the 5-year-olds, who in turn were more disfluent than the 7-year-olds.

Finally, Wijnen (1991) studied disfluency in two 2-year-olds, one of whom was excessively disfluent, the other only mildly disfluent. Wijnen concluded that the excessively disfluent child had problems with the phonological encoding, while the mildly disfluent child had problems with sentence planning.

2.2.8.2 Comparisons between stuttering and nonstuttering children

Williams, Silverman & Kools (1968) studied the adaptation effect in children with an age span ranging from kindergarten through the sixth grade. They found that adaptation occurred in both groups, to approximately the same degree, and concluded that adaptation is not unique to stutterers. Williams, Silverman & Kools (1969a) compared the consistency effect in

stuttering and nonstuttering children aged from kindergarten through sixth grade, and observed an effect in both groups, although the effect was slightly stronger in the stuttering group.

Westby (1974) compared normally disfluent children, highly disfluent children (not diagnosed as stutterers) and stuttering children as to semantic and syntactic language performance. She found that the highly disfluent and stuttering children obtained significantly lower vocabulary scores, made more grammatical errors and obtained significantly lower scores on the semantic tasks, indicating that there might be a language problem underlying disfluency in children. Neither the highly disfluent nor the stuttering group exhibited "deviant" language.

Wall (1980) found that stutterers used less complex and mature language (e.g. less varied use of conjunctions, higher number of coordinate clauses that did not begin with a coordinate word, paucity of complete sentences and syntactic complexity and so on) than did nonstutterers. That language skill might play a role was further indicated by Ryan (1992), who studied articulation, language, fluency and speech rate in stuttering and nonstuttering children. He reported that stutterers performed worse than nonstutterers on seven out of eight language measures. Stutterers also obtained lower scores than average scores for their age groups. There were differences concerning articulation proficiency (although several of the stuttering boys later required treatment). Girls demonstrated higher language scores and faster articulation rates than boys.

That sentence complexity might play a role was also demonstrated by Gaines, Runyan & Meyers (1991) who found that sentences that contained a stuttering event within the first three words were significantly longer than sentences that did not. However, Ratner & Sih (1987) found no significant differences between stutterers and nonstutterers as a function of syntactic complexity, although increases in syntactic complexity correlated with fluency breakdown in both groups. Karniol (1995) provided a thorough review of the literature, and reached the conclusion that some kind of language problem is at play in stuttering, remarking that "[d]evelopmentally, then, stuttering is related to producing sentences rather than to producing speech per se" (Karniol, 1995, p. 105).

Blood, Blood & Hood (1987) studied lateralization in young stutterers and nonstutterers, and although both groups showed a significant right-ear advantage, this tendency was significantly smaller for stutterers, thus lending some support to the notion that stutterers are less lateralized than nonstutterers. It should be pointed out that the results in this study might be confounded by several factors. For instance, four of the stutterers recovered from stuttering during the study.

Meyers & Freeman (1985) studied the interrupting behavior of stuttering and nonstuttering children and their mothers. They found that the mothers of nonstuttering children interrupted disfluent speech of their children significantly more often that did mothers of stuttering children. However, all mothers interrupted disfluent speech more often than they interrupted fluent speech. All children tended to be disfluent when they interrupted their mothers. Meyers (1986) studied nonstuttering and stuttering children in dyadic conversation with either their own mother, an unfamiliar mother of a nonstutterer or an unfamiliar mother of a stutterer. She observed a remarkable consistency of disfluency rates over the three sessions, but also concluded that stutterers and nonstutterers can be differentiated both qualitatively and quantitatively in that the former produced far more disfluencies overall, and also more

part-word repetitions, prolongations and tense pauses. Waspwocz, Yairi & Gregory (1985) observed that "sophisticated listeners could not identify the stutterer" (Waspwocz, Yairi & Gregory, 1985, p. 186) in a group of three stuttering and seven normal preschool children.

Hubbard & Yairi (1988) studied clustering (cf. Ellen-Marie Silverman, 1973a, 1973b; Colburn, 1985) in stuttering and nonstuttering preschool children and found that disfluencies occurred more often than chance in clusters for both groups. They also noted that stutterers showed both higher proportions of clusters, and also size of clusters, than did nonstutterers.

Howell, Kadi-Hanifi & Young (1991) studied phrase revisions in stutterers and nonstutterers aged between three and 11 years. They found significant differences both concerning syntax and prosody between the two groups, e.g. that stutterers made far fewer prosodic changes in their revisions. They suggest that e.g. prosodic analysis could be of help in early identification of stuttering.

Kelly & Conture (1992) studied speaking rates, interrupting behaviors and response time latencies in stuttering and nonstuttering children and their mothers. They observed no differences except that the mothers of the nonstuttering children have significantly faster speaking rates than both groups of children. They interpret that as support for a demands–capacities model of conversational interaction in which mothers adjust their speech to the demonstrated capabilities of their children.

Zebrowski (1994) studied school-aged children who stuttered and observed that the average duration of stuttering was around 750 ms, and was not correlated with either age or general speech disfluency. She suggested that there might be a relationship between duration of stutter and the amount of prolongations produces, as well as to general articulatory rates.

Razzak & Ratner (1999) found that stuttering children prolonged their utterances significantly more under delayed auditory feedback than did nonstuttering children, and suggested that stuttering children monitor their speech far more intently than do nonstuttering children.

Concerning the question when children become aware of the notion of stuttering, Ezrati-Vinacour, Platzky & Yairi (2001), used puppets one of which spoke fluently and one who was disfluent, and asked children of varying ages to identify what puppet spoke "like them". They concluded that awareness of disfluency begins at age 3, and that most children reach full awareness of disfluency at age 5. They also noted that by age 4, disfluent speech was considered as "not good" and that fluent friends were preferred to disfluent friends.

De Nil & Brutten (1991) studied attitudes vis-à-vis disfluent speech in stuttering and nonstuttering children, and found that stuttering children were far more negative than nonstuttering children, already from age 7 (which was the youngest group in the study). Moreover, the negative attitudes evinced by the stutterers increased with age. However, in the nonstuttering group, negative attitudes decreased as a function of age after age 9. De Nil & Brutten pointed out that attitudes must be addressed in any therapy for youngsters.

2.2.9 Listener judgments: stutterer or nonstutterer?

One final issue with regard to stuttering is whether speaker-listeners are able to classify speech as stuttered or normal. Moreover, when it comes to judgments, are there any differences between speaker-listeners who stutter themselves and speaker-listeners who do

not stutter? Also, are there any differences between professionally trained clinicians or speech pathologists and laymen? This question has been investigated several times over the years. Thus, Tuthill (1940) found that stutterers were harsher judges than were nonstutterers when labeling speech samples as being stuttered, i.e., stutterers were significantly more inclined to judge speech as stuttered than were nonstutterers. Tuthill (1946) found that clinicians were no more consistent than were laymen when judging speech as being stuttered, and that stutterers exhibited even less agreement as a group when classifying speech as stuttered. However, no group—speech pathologist, laymen, stutterers—performed better than any other group. It is interesting to note that normal speakers were far less inclined to judge speech samples as instances of stuttering than were clinicians or stutterers. The observation that professional speech pathologists are more inclined to label speech as stuttered was replicated by Boehmler (1958), who also found that certain types of disfluency was more likely to be regarded as stuttering, i.e. sound or syllable repetitions were more often regarded as signs of stuttering than other types of disfluency, while interjections (filled pauses) were not considered sign of stuttering.

Tuthill (1946) found that judges were not affected by whether or not they were watching a film or just listening to tapes in their classification of speech as stuttered or not. Luper (1956) exposed judges to silent film of stuttering, and compared that to audible samples without film, and found that the (silent) films yielded slightly more instances of stuttering labeling than did the auditory samples, showing that visual cues are also at play. In order to investigate the role of visual information, Williams, Wark & Minifie (1963) used three kinds of material to investigate assessment of stuttering severity: audio-only, visual-only and full audio-visual samples. Their judges were nonprofessional nonstutterers. Overall, the results replicated Tuthill's (1946) observation that no major influence could be detected. However, some speakers were judged to stutter more often under visual-only observation than under audio-only or audio-visual observation, indicating that there are individual differences between different speakers (stutterers). Hartsuiker et al. (2003) replicated the observation that stutterers are harder judges than non-stutterers when labeling speech as fluent or not.

Curlee (1981) played videotapes of stutterers to 23 college students and asked them to identify normal disfluency and stuttering, with or without having stuttering defined to them. He concluded that neither disfluency nor stuttering are reliable response classes, and that considerable overlap occurs. However, certain types of disfluency triggered stuttering-classification more often than other types of disfluency, notably repetition and prolongation. Moreover, not only was interjudge agreement poor, even *intra*judge agreement was unsatisfactory with regard to stuttering sites.

More recently, Cordes (2000) examined the reliability with which judges identified individual disfluency types. Thirty judges were asked to identify all perceived disfluencies in a five-second sample on videotape, either alone or in pairs (of judges). While intrapair and interpair agreement was higher than interjudge or intrajudge agreement, consensus averaged less than 50%, which led Cordes to caution against disfluency-type based definitions of stuttering.

In conclusion, in assessing whether speech is stuttered or not, very little agreement is at hand among listeners, be they nonstutterers, stutterers or professional speech pathologists. Stutterers seem to be more inclined to judge disfluency as instances of stuttering—perhaps only showing that they are more aware of the phenomena—and certain types of disfluency are more likely to be classified as stuttering (e.g. sound repetition). This only strengthens the

notion that a categorical difference between stuttered and nonstuttered speech is hard to define, even if one assumes that such a distinction is valid.

2.2.10 Different views on stuttering

To summarize the different views on the causes of stuttering is simply an overwhelming task, especially given the limited scope provided here. So many factors and variables are at play, and so much research seems to prove that a particular phenomenon seems to be a determinant, only to be contradicted by other studies. There seemingly is no consensus within the stuttering community concerning either the causes or the treatment, and some tend to stress social or psychological factors, while others consider physiological causes to be more important. During the 1940s, Johnson and colleagues introduced the so-called **diagnosogenic** (or **semantogenic**) theory of stuttering. Briefly, it regarded stuttering as something that was created by drawing attention to the normal hesitations, or repetitions, of a child. Stuttering, it was meant, did not begin in the mouth of the child, but rather in the ear of the parent. Consequently, those who are labeled, or diagnosed as, stutterers will become stutterers.[1] (It could be pointed out here that this theory of course sat well during the heyday of behaviorism.) This theory has lost some of its impact, but still has its advocates. In a similar vein, Bloodstein, Alper & Zisk (1965), argued that stuttering was an outgrowth of normal disfluency, a view that later was dubbed the **continuity hypothesis**, given the overlap between what is considered stuttered speech and normal disfluency (both with regard to categories and frequency). Sheehan (1958) proposed that stuttering was the result of an **approach–avoidance conflict**, i.e. the speech disruption is caused by the conflicting goals of both wanting to speak and fearing to speak. Flanagan, Golddiamond & Azrin (1958, 1959) argued that stuttering was an **operant behavior**, and consequently could be brought under conscious control.[2]

An *enormous* amount of additional observations adds to the complexity of the issue, such as the low incidence of stuttering in diabetics (Van Riper, 1971/1982, p. 48), the observation that stutterers seem to be more external in their locus of control than nonstutterers (McDonough & Quesal, 1988), score lower on IQ and language tests (Andrews et al., 1983), exhibit more right-hemispheric alpha wave suppression during speech tasks than nonstutterers. Research has been done on the blood, urine and saliva of stutterers (Van Riper, 1971/1982, p. 350). Also, as we have seen, stuttering is reduced during different artificial conditions such as delayed auditory feedback, masking noise, shadowing tasks, speaking while performing other motor tasks, reaction time tests (both phonatory and manual). Moreover, stuttering is more prevalent in males than in females—from five to ten times as many, according to Andrews et al. (1983); a ratio of three-to-one, according to Bloodstein (1969/1987); between three-to-one and five-to-one, according to Franklin Silverman (1992)—but Yairi (1981) observed that this difference was not be found in very young children, the incidence of stuttering is higher among the retarded (Starkweather, 1987, p. 158). If one twin is a stutterer, it is more likely that a monozygotic sibling is also a stutterer than a dizygotic sibling (Starkweather, 1987, p. 160).

---

[1] A sequitur of the diagnosogenic theory of stuttering is that one should be able to turn normally-speaking people (at least children) into stutterers by diagnosing them as stutterers, and draw their attention to their hesitation. This is indeed what may have happened in one experiment that resulted in an M.A. thesis under Johnson's supervision, the so-called "Monster Study" (Tudor, M. 1939. *An experimental study of the effect of evaluative labeling on speech fluency*. Master's Degree thesis, University of Iowa). For an account of the (possibly) tragical story, see Franklin H. Silverman (1988).

[2] See also Shames & Sherrick (1963) for an early discussion of non-fluency and stuttering as operant behavior.

Stuttering seems to be more prevalent among the left-handed (Starkweather, 1987, p. 214), is only rarely found among the congenitally deaf, and often disappears with the onset of deafness (Starkweather, 1987, p. 243).

Starkweather (1987) pointed out that stuttering as a phenomenon seems to be a uniquely *human* behavior:

> [N]othing that even resembles stuttering has been found in the communicative behaviors of other species, but of course, ours is the only species that uses language and speech to communicate with. Still, if stuttering were a simple mechanical problem, one might expect to find it in the complex sequences of birdsong or whale songs. Interestingly, one of the early attempts to teach a chimpanzee human language was the Kelloggs' raising of Vicki as if she were a human baby. Vicki learned to say two or three words—*papa* and *cup* were demonstrated—but she produced these words with great difficulty and strain, and many a speech clinician has thought, on seeing the film of the Kelloggs' work, that Vicki's speech resembled that of a human stutterer. (Starkweather, 1987, p. 155; italics in original.[1])

Given all the observations listed above, and many more, and given that they all seem to rest on some empirical evidence, one is tempted to assume a **multi-causality view** on stuttering, that there must be some truth to them all (repeating Bloodstein, 1969/1987, p. 81). Bloodstein (1969/1987), however, takes a critical stand on this position:

> Those who find this [the multi-causality view] an easy solution to the problem must be prepared to answer the objection that we as yet have essentially no conclusive evidence to show that *any* of the current theories of stuttering is wholly or partially valid, let alone to support the somewhat improbable conclusion that they all are. (Bloodstein, 1969/1987, p. 81; italics in original.)

Despite the qualitative and quantitative overlap of many of the phenomena presented above, there seemingly are some unequivocal differences between stutterers and nonstutterers, after all. Other such differences could be the presence of **tense pauses**, i.e. silent pause accompanied by audible and/or visual sounds of struggle (Franklin Silverman, 1974, p. 33; Silverman, 1992, p. 6, p. 21 and p. 40; Onslow, 1995, p. 587; Adams, Sears & Ramig, 1982, p. 24; Adams & Ramig, 1980, p. 460). An example of a phenomenon that "seems to be the one that does differentiate stuttering from normal speech disfluency" (Silverman, 1992, p. 55) comes from observations concerning the **adaptation effect**, previously described. If stutterers and nonstutterers read the same passage several times, both groups exhibit increased fluency. However, if the speakers pause some time after such a session, and then begin anew, nonstutterers will maintain their obtained fluency, whereas stutterers are likely to be more

[1] Starkweather seemingly mixes things up here (which does not annihilate his general point). The Kellogs did indeed raise a chimpanzee, but the name of that chimp was **Gua**, as described in e.g. Kellog & Kellog (1933) or Kellog (1968). *Viki* was indeed a talking chimp, but was raised by Keith and Cathy Hayes (Hayes & Hayes, 1951, 1952; Hayes, 1951; see also Aitchison, 1976/1993 or Deacon, 1997, p. 355). At least two films were made of Viki: *Vocalization and Speech in Chimpanzees*, Psychological Cinema Register film no. PCR–2032, Pennsylvania State College (referred to in Hayes & Hayes, 1951) and *Mechanical interest and ability in a home-raised chimpanzee*, Psychological Cinema Register, State College, Pennsylvania, USA—referred to in e.g. Greenfield (1991) without number. Incidentally, other chimps who learned to use sign language (ASL) or other symbolic languages (e.g. **Washoe** or **Nim Chimpsky**), have been noted to make excessive use of repetition in their signing (e.g. Aitchison, 1976/1993, pp. 33–47), which probably should not be viewed as disfluency in the human sense.

disfluent on the first run of the second session than they were on the last run of the first session, thus exhibiting **spontaneous recovery** of disfluency.[1]

For those who wish to dive deeper into this issue, good reviews of different theories and findings are found in Andrews et al. (1983) and Perkins (1990), who both summarize stutterer–nonstutterer differences, describe different treatments and cover major theories of stuttering.

2.2.11 Summary

As we have seen, despite decades—or even centuries—of research on stuttering, there is still no clear definition of what it is or is not, there is no acid test diagnostic to differentiate between stuttering and normal disfluency, and there is little agreement within the community as to the underlying causes. Qualitatively, almost all phenomena exhibited by stutterers are also found in nonstutterers. Moreover, there is also a considerable quantitative overlap in most respects between stutterers and nonstutterers that further complicate the picture.

A couple of quotes from the stuttering literature illustrate the problem:

> There is no test within science which can determine once and for all whether a fluency departure is a stuttering instance or a nonstuttering disfluency. (Young, 1985, p. 13.)

> There is a consensus among speech-language pathologists that the cause of stuttering is unknown. This lack of understanding is not due to the lack of research effort, but it may be due to asking an unanswerable question. (Boehmler & Boehmler, 1989, p. 447.)

Wingate (1984c), debating whether or not stuttering is different from normal disfluency, in a response to Perkins (1983), points out that:

> The fact that some stutters and normal disfluencies may be difficult to differentiate seems to me to be a relatively minor issue; certainly it does not provide any substantial ground for contending that one cannot tell the two kinds of speech apart. Even more certainly it cannot be taken as evidence that the two kinds of speech or disfluencies are essentially the same. Research which finds evidence of problems in disfluency differentiation (particularly in light of the extensive evidence to the contrary) indicates only that there was a problem. Such findings do not provide any answers; at best, they simply raise a question. (Wingate, 1984, p. 430.)

Despite the overlap between diagnosed stuttering and normal disfluency, both qualitatively and quantitatively, and despite the fact that professionals and laymen alike find it hard to diagnose stuttering when exposed to speech samples, I will for the rest of this thesis assume that there *is* a difference, along the lines of Wingate's (1984c) argumentation above, however cumbersome it might be to pin-point exactly what that difference would be (but there I am in good company).

Stuttering research constitutes a massive body of research on disfluency and ensuing findings of utmost interest to anyone interested in disfluency, be it stuttered or "normal". However, although there is must controversy with regard to the "different or same" issue concerning stuttered and normal disfluency, I will for the rest of this thesis side with Wingate in assuming

[1] Not to be confused with the "spontaneous recovery" used to refer to stutterers who lose their stuttering without any treatment whatsoever.

that there (most likely) *is* a genuine difference between the two. It should be remembered, however, that this view is not self-evident.

2.3 Psychotherapy and psychology

Alongside disfluency research within the field of stuttering, speech disturbances were also studied from a psychological perspective, with a starting point within psychotherapy. What was new here was that while the linguistic *content* of what the patient said in the psychiatric interview was, by definition, of utter interest, the notion that the linguistic *form* in and by itself could be used to reveal the mental status of the patient. A fairly detailed description of this research will be given in the following section.

2.3.1 Speech disturbances in psychotherapy

In the 1950s, alongside research on stuttering, hesitation in speech was also studied within psychotherapy. Mahl (1956, 1958), Kasl & Mahl (1958, 1965), Lerea (1956), Dibner (1956, 1958), Meisels (1967) and Zimbardo, Mahl & Barnard (1963) thus pioneered disfluency studies from the perspective that anxiety in patients could be gauged by counting the number of disfluencies during given stretches of speech in a therapy situation.[1] The (then) novel approach was to study not only the semantic *content* of what the patient uttered, but the linguistic *form*, the speech disturbances produced by the patient, in order to appreciate what topics raised the anxiety levels in a patient during the interviews. Or, as Mahl put it:

> The basic working hypothesis underlying the present use of recordings for *this* purpose has been that the most valid linguistic measures of anxiety will be those based on the behavioral or "expressive" aspects of the speech rather than those based on manifest verbal content analysis." (Mahl, 1956, p. 1, italics in original.)

Mahl continues:

> Empirically, two of the many behavioral attributes of speech in the interview that are useful to the therapist in assessing anxiety in the patient are *(a)* disturbances in speech called "jumbled," "confused," or "flustered" speech, and *(b)* hesitancies and longer silences by the patient when he is free and motivated to talk. /…/ Speech disturbances and short hesitancies may also be conceived as predominantly indirect consequences of anxiety that do not have the instrumental function of reducing anxiety." (Mahl, 1956, pp. 1–2.)

Mahl listed what counts as disturbances (Mahl, 1956, p. 2):

1. *Ah*, to be distinguished from *er* and *um* and so on. (It is not entirely clear how Mahl treated those in this work. In a later listing (Mahl, 1987b, p. 218), *eh*, *uh* and *uhm* are included as "less frequent variants" of *ah*.)

2. Sentence correction. Basically any correction on content or form that is perceived by the listener as a correction.

3. Sentence incompletion. Any expression that is interrupted and not repaired.

[1] Already in 1949, Verzeano & Finesinger (1949) presented an automatic (sic!) analyzer for free association speech during interviews that included analysis of silent pausing.

4. Repetition of one or more words.

5. Stutter.

6. Intruding incoherent sounds. Any incomprehensible sound made by the speaker that cannot be categorized as a stutter, omission or slip-of-tongue.

7. Tongue-slip. Erroneous words, neologisms, transpositions of words from their correct serial order, and so on.

8. Omission. Truncations of words. Normal contractions are not included.

This list appears later in Kasl & Mahl (1965, p. 426; Mahl (1987b, p. 218), with only minor differences. As can be seen, this list corresponds well with the categories employed within stuttering research, although the mapping is not completely one-to-one.

Mahl then measures the anxiety level of the patient by using two measures:

$$\textit{Speech–Disturbance Ratio} = \frac{N \text{ Speech Disturbances}}{M \text{ “Words” Spoken by Patient}}$$

… and:

$$\textit{Patient–Silence Quotient} = \frac{N \text{ Seconds of Silence}}{M \text{ Seconds Available to Patient to Talk}}$$

Mahl (1956) interviewed twelve patients, rated their anxiety and concluded that “the Speech–Disturbance Ratio and The Silence Quotient are reliable and discriminating measures” (Mahl, 1956, p. 11). Kasl & Mahl (1958) interviewed 25 experimental and 10 control subjects in a neutral and a stress situation (where anxiety was elicited, and thus a controlled variable), and found “a very significant increase in speech disturbances” under the anxiety condition (Kasl & Mahl, 1958, p. 349). Interestingly, however, Mahl (1958) found that the filler word *ah* did not vary with anxiety, thus differentiating it from other speech disturbances. Thus, in Mahl’s words “speech disturbance and silence seem to be expressive attributes that are useful as anxiety indices.” (Mahl, 1956, p. 13.) Boomer & Goodrich (1961) studied speech disturbance in two patients and found support for Mahl’s hypothesis in one patient, but not in the other. This observation was replicated by Christenfeld & Creager (1996), who, when reviewing the literature, found five studies that reported an increase in filled pause production as a function of anxiety,[2] ten studies where no difference was found,[3] and one study where a decrease was found.[4] Christenfeld & Creager (1996) also extend the Levelt’s (1989) suggestion that filled pauses are a sign of error detection to the notion that “anything that makes a speaker stop the automatic production of speech can lead to an *um*, whether or not there is an error involved” (Christenfeld & Creager, 1996, p. 452; italics in original), thus making it adhere to Baumeister’s (1984) notion of “choking under pressure” (*vid.* 2.3.3).

---

[1] In the original sources, both the denominator and numerator are indicated with *N* for all measures. Since *N* divided by *N* always give a ratio of 1 (one), I have changed the numerator to *M* in all the measures, which is obviously what the authors intended. (Thanks to Martin Eineborg for pointing this obvious fact out to me.) However—which Joakim Nivre pointed out to me—if *N* is interpreted as “number of”, the original version makes sense.

[2] Boomer (1963), Jurich & Polson (1985), Koomen & Dijkstra (1975), Lalljee & Cook (1973) and Panek & Martin (1959).

[3] Cook (1969b), Feldstein (1962), Kasl & Mahl (1956), Mahl (1956, 1987), Meisels (1967), Paivio (1965), Pope et al. (1970) and Siegman & Pope (1965a, 1965b).

[4] Blass & Siegman (1975).

Mahl (1987a, 1987b) summarized three decades of research on speech disturbances and anxiety conducted by himself and colleagues (e.g. Lassen, 1987; Wolf, 1987; Kasl & Mahl, 1987; Mahl & Bender, 1987: Schulze, 1987; Mahl, Schulze & Murray, 1987; Zimbardo, Mahl & Barnard, 1987) and concluded that the best indicator of anxiety levels is the *Non-ah Ratio*, defined thus (Mahl, 1987b, p. 219):

$$\textit{Non-ah Ratio} = \frac{N \text{ Non-ah Disturbances}}{M \text{ “Words” in sample}}$$

To fully account for all the findings from a psychiatric perspective, as presented in the works by Mahl, is not possible here, but let me just mention an observation of interest.

Mahl (1987b) discussed individual differences between speakers, and made a distinction between *ah-ers* and *sentence-changers* (Mahl, 1987b, p. 231). Mahl mentioned that “‘Ah-ers’ report having had strict parents and being ruminative in thinking, while ‘Sentence-changers’ have difficulty in concentrating and in speaking publicly” (Mahl, 1987b, p. 264).

The main finding of Mahl and colleagues, however, would probably be that *ah*, i.e. what is called a filled pause here (and *interjection* and so on in the literature), has a distinct function that separates it from all other kinds of disfluency.

2.3.2 Disfluency as a function of anxiety, intimacy and sex

Anxiety levels can be raised for a number of reasons. Jurich & Polson (1985) videotaped interviews with female college students while they were asked four questions about premarital sex. The rationale for doing this was that:

> When discussing personal matters, there seems to be a difference between sexual content and any other type of content area wherein sexual content evokes an approach-avoidance reaction. Although such conflict may be less present in other societies /…/, anxiety very often accompanies sexual content presented to a subject from the American culture” (Jurich & Polson, 1985, p. 1247.)

The questions presented to the subjects, in an order of supposed increasingly intimate and anxiety-raising nature, were about kissing, petting, sexual intercourse and oral-genital sex. They found an increase in editorial errors and filled pauses between all categories with the largest increase between kissing/petting and intercourse/oral sex.

That sex (or sexual inferences) in itself does not suffice to elicit speech disturbances was shown by Schulze (1987). Schulze asked subjects to describe high-stress and low-stress films. The low-stress film segment showed Arunta tribesmen engaged in hunting (and similar), while the high-stress film segments showed “at close range, subincision of the penis of adolescent Arunta tribesmen” (Schulze, 1987, p. 236). While subjects rated themselves as more anxious during the circumcision film segment, this did not manifest itself in their verbal descriptions of the film. Mahl (1987b, p. 247 and p. 251) argued that the explanation possibly is that description of *external* events, like, a film segment, does not lead to an increase in speech disturbances, even if this is accompanied by an increase in anxiety, while verbal production of high-anxiety *internal and personal* matters is characterized by an increase in speech disturbances.

Pope & Siegman (1962) found that patients in psychotherapy became less disfluent as the therapist became more specific in his comments. Or, as they say: "it would appear that as therapist specificity increases patient clause units (productivity) and speech disturbance ("anxiety") decrease" (Pope & Siegman, 1962, p. 489).

Panek & Martin (1959) investigated whether **galvanic skin response** (GSR) was related to speech disturbances (using Mahl's scheme), and found that this was the case. This lead to the proposal that a combination of GSR and speech disturbance scores would constitute a reliable indicator of anxiety levels in psychotherapy interviews.

2.3.3 "Choking under pressure"

Fenigstein, Scheier & Buss (1975) and Carver & Scheier (1978) discussed different kinds of self-consciousness, and their respective roles in performance. They differentiated between **private self-consciousness**, defined as attending to one's inner thoughts and feelings, **public self-consciousness**, defined as one's awareness of the self as a social product, i.e., how one appears to other people, and **social anxiety**, which amounts to discomfort in the presence of others, exemplified by e.g. fear of giving public speeches. They reported that they had not found any gender differences in connection with this typology.

The roles of both self-awareness and personality type have been studied in connection with performance, both for speech production and other activities. The phrase "choking under pressure" was used by Baumeister (1984) to define "performance decrements under circumstances that increase the importance of good or improved performance" (Baumeister, 1984, p. 610). Baumeister showed that increased self-attention resulted in decreased performance capabilities. He also showed that people who were low in self-consciousness (as a general personality trait) performed better in control conditions, while these people to a higher degree choked under pressure, thus showing that both pressure levels and general personality traits play a role. Baumeister (1984) concluded that both pressure proper and self-consciousness harm performance. Baumeister, however, took no stand concerning the suggestion that self-consciousness and attention to something else (but oneself) are mutually exclusive, as has been suggested by Duval & Wicklund (1972).

It has been shown that any kind of motor action—which speech is an example of—is subject to deterioration under stress. In recent work, Beilock & Carr (2001) compared golf putting between novice and expert golfers under a variety of different conditions. They found that choking occurred in putting (a sensorimotor function), but not in the control task alphabetic arithmetic. According to Beilock & Carr (2001), there are two main competing theories to explain "choking". The first is the **Distraction Theory**, which proposes that pressure creates a distracting element that shifts attentional focus from the relevant task. The second is the **Explicit Monitoring Theory** that suggests that pressure raises self-monitoring, or self-awareness, and is thought to disrupt (automatized) skilled performance. Their study mainly supports the latter theory.

So, does pressure act decrementally on the highly automatized, skilled, performance that is speech? Several studies have delved into the field, and I will briefly summarize but a few of them in the following.

2.3.4 Disfluency under manipulation

It has been shown that disfluencies can be manipulated in a number of ways that both imply that they are under conscious control and that they reflect subconscious processes in human behavior. Some of these studies will be described below.

2.3.4.1 Disfluency and instruction

The easiest way to reduce disfluency seems to have been observed by Boomer & Dittman (1964) and Siegel & Martin (1966) who found that subjects became less disfluent simply as a result of being *instructed* to be less disfluent. This is a clear indication that disfluency production is under speaker control.

2.3.4.2 Disfluency and verbal punishment

Stassi (1961) had 24 subjects read out aloud nonsense words from a set of cards. After each word, the response "right" (pronunciation) or "wrong" was given. Stassi used four reinforcements schedules:

1. 100% reward—0% punishment,

2. 66% reward—33% punishment,

3. 33% reward—66% punishment, and

4. 0% reward—100% punishment.

Stassi reported that speakers became more disfluent when their verbalizations were punished. Moreover, he observed than men were more disfluent than women in the 100% punishment condition.

In two studies carried out by Siegel & Martin (1965b, 1967), the word "wrong" was made contingent upon disfluencies in the spontaneous speech of subjects. A significant decrease of disfluencies was noted, which was not the case when "wrong" was presented in a random way.

2.3.4.3 Disfluency and electric shocks

It hardly comes as a surprise that stuttering research was first out. Hill (1954) investigated whether stutterers and non-stutterers exhibited more disfluent speech under the threat of electric shocks. Speech disfluencies, as well as EMG recordings of hand movements, were studied. It was concluded that threat of penalty resulted in very significant disorganization of speech, i.e., speech became more disfluent.

Siegel & Martin (1965a) found that when electric shocks were given contingent on speech disfluency, it resulted in a decrease in disfluency production, making speech more fluent. They concluded that speech disfluency is a manipulable phenomenon, and consequently under (at least some) control of the speaker.

In a similar vein, Flanagan, Golddiamond & Azrin (1959) instated "stuttering" (i.e. disfluency) in normally fluent subjects by submitting them to electric shocks while reading passages of text.

2.3.4.4 Making people pay for their disfluency

In order to investigate further ways to decrease disfluency in the speech of normal-speaking subjects, Siegel, Lenske & Broen (1969) studied five students in a number of differently designed sessions. During one of the sessions, the students had to pay (literally) a penny per disfluency produced, which led to a significant decrease in disfluency production, "in all but one subject, the disfluencies were reduced to near-zero levels during spontaneous speech" (Siegel, Lenske & Broen, 1969, p. 275). This lends further support to the hypothesis that disfluency production is partly under the control of the speaker.

2.3.5 Disfluency in different speaker settings

That social support has positive effect on subjects submitted to stressful situations has been shown by e.g. Sarason (1981). It comes as no surprise, then, that studies have been devoted to disfluency production under different speaker settings, with familiar or unfamiliar interlocutors, and with different imagined audiences.

Ellen-Marie Silverman (1971) studied the speech of preschoolers in three different settings: during free play, in a testing room and in their homes talking with family members. She observed that the frequency of disfluency varied systematically across the different situations. Ellen-Marie Silverman (1972) studied ten 4-year boys in two settings, in the classroom, and in a structured interview, and found considerably more disfluencies in the structured interview, which prompted her to warn researchers against regarding structured interviews as representative data.

Broen & Siegel (1972) asked 49 subjects to speak in four different situations: 1) In an "alone" situation, where they talked spontaneously about whatever they wanted while being alone in a room. 2) In a second setting where they talked to a TV camera and studio lights, and were told that the session was being videotaped. 3) The subjects were instructed to talk about anything but *as if* to a live audience. 4) A final setting, where the subjects conversed casually with the experimenter. It was found that the subjects were most disfluent in the casual conversation setting, and less disfluent in the audience or TV settings, which were comparable to the alone setting. The explanation given is that the "less important" the subjects regarded the task, the more disfluent they were.[1] Lay & Paivio (1969) found that unfilled pauses were positively correlated with audience sensitivity, while filled pauses were not.

Wexler (1982), while studying developmental disfluency in 2-, 4- and 6-year old (nonstuttering) boys in neutral and stress situations found only one statistically significant difference, where the 2-year old group had a significantly higher frequency of word and phrase repetitions in the neutral (sic!) situation. This confirms the findings of Martin, Haroldson & Kuhl (1972a, 1972b) who found no differences in disfluency rates in nonstuttering children when they compared conversations where the child spoke with another

[1] However, Hulit & Haasler (1989) had sixty normal-speaking subjects read three common tongue-twisters under six conditions, where suggestions of degree of difficulty were given, ranging from "easy" to "extremely difficult". They found that the instructions did not affect rates of nonfluency. They simply, and somewhat laconically, conclude that: "Some normal speakers are very comfortable speaking to large groups of people. Others are terrified." (Hulit & Haasler, 1989, p. 367.)

child and conversation with their mothers (Martin, Haroldson & Kuhl, 1972a) or situations where the child spoke with a talking puppet versus their mother (Martin, Haroldson & Kuhl, 1972b).

2.3.6 The alcohol effect

In order to verify their hypothesis that filled pause rates are due more to self-attention than to increased anxiety levels, Christenfeld & Creager (1996) studied people in a bar. As is pointed out above, filled pauses are more frequent when speakers are self-conscious (other disfluencies not affected), but are not affected as a function of increased anxiety (which affects all other types of disfluency).

Christenfeld & Creager pointed out that:

> Drinking alcohol interferes with just about every task that intoxicated people attempt. /…/ Alcohol /…/ reliably decreases people's self-awareness /…/ decreases people's ability to attend to more than one thing at a time. Thus, it should be hard for drinkers to produce speech and also attend to what they are saying. /…/ One advantage of examining alcohol's effects is that the challenges of the speech task and the level of self-consciousness are oppositely affected. (Christenfeld & Creager, 1996, pp. 456–457.)

Consequently, a field study was carried out in eight local bars, where 108 subjects were interviewed by one experimenter, while a second experimenter counted the number of filled pauses produced by the interviewee. The subjects were asked how many drinks they had consumed, and how much they weighed. The results clearly indicated that the number of drinks consumed and filled pause rates were significantly associated. "[t]he more intoxicated the speaker, the rarer the *um*s" (Christenfeld & Creager, 1996, p. 457). However, Christenfeld & Creager included the following passage for readers to heed:

> Before suggesting intoxication as a strategy to concerned public speakers, it should be noted that, to eliminate the average speaker's *um*s, about 19 drinks in the course of an evening are required (assuming a linear alcohol-*um* relationship). (Christenfeld & Creager, 1996, p. 457.)

Thus, Christenfeld & Creager (1996) provided further evidence that:

1. Filled pause production is dissociated from all other disfluency production.

2. Filled pause production is less related to anxiety or complexity levels, and more related to self-awareness/-monitoring.

Christenfeld & Creager (1996) concluded that filled pauses probably provide information about moments in the flow of speech where speech is not produced automatically, but instead is attended to by the speaker, and that filled pauses are not only produced when the speaker detects an error, but rather when the speaker attends to his/her own speech for whatever reason there might be.

2.3.7 Depression

Szabadi, Bradshaw & Besson (1976) studied pause time in a number of healthy volunteers. The task was to count from 1 to 10, in what they called an "automatic speech" task. They

found that pause times were significantly longer during a period of moderate depression in four of the patients, as compared to pause times after recovery.

More recently, Friedman (1991a, 1991b) pointed out that "speech hesitation pauses" are a diagnostic of depression, and also that "pauses are more accurate than fundamental frequency in monitoring dysprosody in Broca's aphasia" (Friedman, 1991a, p. 140). Friedman also mentioned that "[d]iurnal variation in depression, which may be manifested by slowing of speech and response /…/ can be monitored by speech pauses analyzed on a time base" (Friedman, 1991b, p. 181).

2.4 Physiological factors

As possible underlying causes of stuttering, purely physiological reasons have been considered, e.g. less lateralization of the brain, as was previously mentioned. From there, it is not a long stretch to ask whether all disfluency might be the result of physiological phenomena, and some such studies have been carried out.

It must be pointed out already here that there is an obvious overlap between what should be considered psychological and physiological, and that any division between the two must be taken cum grano salis, e.g. whether or not the alcohol effect is psychological or physiological. Given that the assumed *reason* for its effect, less self-monitoring, it was put under the psychological header, however. In this section, I will briefly just mention a couple of more distinctive attempts to relate disfluency rates to underlying physiological causes.

2.4.1 Gender differences

It has been shown that there are gender differences in disfluency production. Stassi (1961) submitted subjects to four different reinforcement schedules, varying the degree of reward and punishment, and found that while all speakers became more disfluent when punished, men became more disfluent than women in the 100% punishment condition.

Feldstein, Brenner & Jaffee (1963) interviewed men and women in two topic settings, problem and non-problem. While they found no differences in non-ah production (Mahl, 1956, 1987a, 1987b), they found that the production of filled pauses (i.e., *ahs*) was higher for men than women, and that it was also positively related to educational level (but not to verbal intelligence).

Lickley (1994) found males more disfluent than women, and Shriberg (1994) reported higher filled pause rates in men than in women. Bortfeld et al. (1999) that men were more disfluent than women overall. Branigan, Lickley & McKelvie (1999) found that men were significantly more disfluent than women in an eye-contact setting, but not in an no-eye-contact setting, which they interpreted as evidence that women are better than men in picking up visual cues from the interlocutors.

On the other hand, Christenfeld (1995) found no gender differences in his study. Likewise, Bell, Eklund & Gustafson (2000) found no gender differences for Swedish subjects. Edelsky (1981) found significant gender differences as to floor (or turn) holding, but also that these differences were dependent on what kind of discourse venture was being carried out.

2.4.2 Disfluencies during the menstrual cycle

One factor that could possible affect disfluency rates is affective state from a more biological point of view. It has been reported that women are significantly more disfluent during premenstruation than during ovulation, both for stutterers (Silverman, Zimmer & Silverman, 1974) and nonstutterers (Silverman & Zimmer, 1975). In the first study (Silverman, Zimmer & Silverman, 1974), on stutterers, four stutterers provided four three-minute speech samples over the phone, one at ovulation and one at premenstruation for two consecutive cycles. The data were then transcribed and analyzed with regard to the following disfluency categories: interjection of sound or syllable, part-word repetition, whole-word repetition, phrase repetition, revision-incomplete phrase, dysrhytmic phonation [i.e. truncation or prolongation] and tense pause.[1] All four women produced more disfluencies at premenstruation, both for total frequency and for individual type of disfluency. A post-experiment conversation revealed that no subject had guessed the objective of the study.

In the second study (Silverman & Zimmer, 1975), on nonstutterers, twelve Caucasian university students aged 17 to 22 were recorded as they spoke on four different topics presented to them on cards. Each topic was given three minutes. Two recording sessions were carried out, one at premenstruation, one at ovulation. The data were then transcribed and analyzed with regard to the same disfluency categories as the previous study. Nine of the women produced more disfluencies at premenstruation than at ovulation. Of the three women who were more disfluent during ovulation, two had very minor differences as to rates. A closer look at the data, however, revealed that of the different disfluency types examined, only repetition-incomplete phrases were significant at the 0.05 level, the others having "approximately the same frequencies of occurrence at premenstruation as at ovulation" (Silverman & Zimmer, 1975, p. 205). A post-experiment interview revealed that two of the subjects had guessed the goal of the study. Both these subjects produced more disfluencies premenstrually.

Thus, it seems that if biological cycles, or states, such as the menstrual cycle do affect speech behavior in general, and disfluency production in particular, this difference seems to be minor indeed. Giles & Giles (1976), also pointed out some methodological weaknesses of the two aforementioned studies, as well as other reasons to be very careful about such studies that attempt to link biological cycles with behavior, for political reasons. Giles & Giles (1976) concluded that:

> Admittedly, there is a body of evidence to suggest such a psychogenic link between biological changes in the menstrual cycle and certain cognitive functionings. /…/ Yet in any case, their [Silverman et al.] finding that women's speech at premenstruation is more disfluent than at ovulation seems to be methodologically biased, statistically dubious, and perhaps socially meaningless. (Giles & Giles, 1976, p. 188.)

Be that as it may, biological approaches of this, and other, kind are way beyond the scope of the present work. Suffice it to say that the study of disfluency phenomena obviously knows few, if any, borders.

[1] The first five categories were adopted from Johnson and Associates (1959), and the two other from Williams, Silverman & Kools (1968).

2.4.3 Hesitation vowels as a phonomotoric subroutine

Schönle & Conrad (1985) studied the hesitation vowel *ah* and *mh*, i.e. filled pauses, in German in connection with respiration. They started out by observing that linguists and psychologists have taken pausal phenomena as indicants of speech planning. Schönle & Conrad (1985), however, suggested that there is an alternative explanation, *viz.*, that filled pause incidence can be explained from a speech motor physiology perspective.

Schönle & Conrad asked sixteen subjects to speak spontaneously for five minutes on a topic of their own choice, while they measured respiration using a chest pneumograph. Five of the subjects did not produce any filled pauses, but for those subjects who did, 62.5% of the filled pauses fell within the first segment of respiration.

Schönle & Conrad:

> The preponderance of hesitation vowels early during expiration can be interpreted in physiological terms. During speech the air volume available in the lungs is at its maximum level early during expiration and drops continuously toward the end of expiration. As speech respiration is reset to vegetative breathing whenever the stream of speech ceases /…/, such resetting at early points in exhalation would lead to a sudden loss of large air volumes and dramatically increase breathing frequency. Hesitation vowels therefore are used by the speech production system as phonomotoric subroutines to compensate for the missing speech material and to prevent the respiratory system from uneconomic air loss. (Schönle & Conrad, 1985, pp. 295–296.)

Thus, they concluded that "a straightforward account for the existence and distribution of hesitation vowels can be given on speech motor physiological ground without the need for psycholinguistic interpretations." (Schönle & Conrad, 1985, p. 296.)

Certainly a controversial claim within the linguistics community.

2.4.4 Disfluency in space: pilot studies

As we have seen in the previous paragraphs, phenomena such as task pressure and speaker situation (context) might seriously affect speech (disfluency) production. While this most often is of slight importance, there are occasions where speech production is crucial. The introduction of automatic speech recognition (ASR) systems in several settings have made disfluency studies important from a technical (rather than linguistic or psychological) point of view, and it is not surprising to find industry-driven studies of speech under stress (e.g. Steeneken & Hansen, 1999; Murray, Baber & South, 1996; Brenner, Doherty & Shipp, 1994; see also Cairns & Hansen, 1994).

Speech stress can, as we have seen, have many underlying reasons. Thus, when cosmonauts on the Russian space station MIR communicated with the ground crew, psychologists were monitoring their speech for signs of stress (Berthold & Jameson, 1999). It goes without saying that the introduction of ASR system on the International Space Station (Rayner et al., 2003) ultimately needs to handle both acoustic deterioration in speech as a function of psychological or physiological stress, as well as speech disfluency, for the same reasons.

Another obvious example of the importance of well-functioning automatic systems is that ASR systems are being proposed for inclusion in the cockpits of military aircraft, where pilots

are subject both to severe physiological stress (like *g*-forces) and psychological stress (e.g. being shot at), while at the same being exposed to high amounts of vital information which need to be reacted to instantly. Thus, not only does physiological and psychological stress affect the behavior of pilots, they are already under heavy cognitive load. Baber et al. (1996) studied the effect of high workload on the performance of subjects, and found effects at both syntactic, lexical and phonemic levels.

Berthold & Jameson (1999) reviewed the literature[1] concerning how high workload affects speech performance, and summarized the findings. They found that high cognitive load leads to an increase in a number of different disfluency categories, such as sentence fragments, false starts, self-repairs, silent and filled pauses and repetitions.

2.5 General linguistics

So, given that we are dealing with a linguistic phenomenon, a natural question is what research was carried out within linguistics. To Chomsky, disfluencies were just evidence of the difference between the *performance* capabilities of a speaker, and the underlying linguistic *competence* which reflected the grammar proper of the language, or in his words: "[a] record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on" (Chomsky, 1965, p. 4). This view, as we know, has not been unchallenged. Fillmore (1979), on discussing (different kinds of) fluency, points out that:

> [T]he distinction between competence and performance may not be as important for a larger understanding of language behavior as some scholars have considered it to be. It is a distinction which is most helpful when talking about a world in which language is produced solely for the sake of producing language. In a situation in which language use plays an essential role in a speaker's engagement in a matrix of human actions, however, the distinction seems not to be particularly helpful. (Fillmore, 1979, p. 91.)

So, linguistics proper traditionally made a difference between language *competence*, reflecting "true" language, and language *performance*, the way language actually occurs given the error-prone behavior typical of non-perfect human behavior. This view resulted in (almost exclusively) language descriptions and grammars of idealized versions of language, where assumed underlying rules were the focus, and phenomena outside these descriptions were often simply discarded, much the way newspapers and magazines clean up quotes by interviewees before printing. In the 1950s, however, studies of *speech* phenomena, rather than idealized language, took off, and what had previously been considered performance aspects typical of speech, was beginning to be seen as objects of study in their own right. It may be pointed out already here, however, that while extensive typologies of disfluencies were created within stuttering research and psychotherapy—including cut-offs or truncations, prolongations, repetitions, omissions, intruding sounds, changes, dysrhythmic phonations and so on—most of the early work within linguistics focused mainly on (hesitation) pausing and/or slips-of-the-tongue.

[1] They summarized a number of previous studies, but do not provide the original references. These, they say, are given in: Berthold, André. 1998. *Repräsentation und Verarbeitung sprachlicher Indikatoren für Kognitive Ressourcenbeschränkungen*. Master's Degree thesis, Department of Computer Science, University of Saarbrücken, Germany. I have not been able to obtain a copy of this work

2.5.1 Hesitation and pausing

Early disfluency research within general linguistics focused to a large extent on pausing. Early work was carried out by Bloch (1946) who discussed facultative pauses in Japanese. Cowan & Bloch (1948) studied the relation between perceptual judgments of silent pauses and acoustically present silences in the speech signal, and correlated those with the syntactic structure of the utterances. Other pioneers were Hegedüs (1953) on Hungarian and Lounsbury (1954).

Goldman-Eisler (1954a, 1954b, 1955, 1957, 1958a, 1958b, 1958c, 1961, 1968,[1] 1972) was the first to do extensive studies of hesitation phenomena in spontaneous English speech.[2] Her findings included such observations as that variability in total speech rate seemed to be dependent on time spent pausing, rather than time spent articulating, and that hesitation pauses tended to precede informationally heavy items (i.e. lexical items), occur where speech planning becomes more complex, and that the "distribution of pause lengths is determined by the type of situation in which speech is uttered" (Goldman-Eisler, 1961, p. 233). She also observed that pauses could be very long indeed, up to 30 seconds of duration (ibid., p. 234). Hawkins (1971) observed that "two-thirds of all pauses, and three-quarters of all pause time, are located at clause boundaries /…/ we are justified in concluding that the clause-boundary is the place where much of the speech-planning occurs" (Hawkins, 1971, p. 285).[3]

The previously mentioned work by Bloch (1946), the early work by Goldman-Eisler and the studies by Mahl and colleagues, served as a starting point for Maclay & Osgood (1959) to study hesitation phenomena in English spontaneous speech. Having read Mahl's work, they conceived a large number of disfluency categories, including *ahs* (filled pauses), sentence incompletions and corrections, word repetitions, stutters, intruding sounds, omissions of words or parts of words (truncations) and so on. However, their study included only four of these, namely *repeats*, *filled pauses*, *false starts* and *unfilled pauses*. They observed consistent differences between speakers, both as to total frequency of all disfluencies, but also concerning speakers' preference for the different types of hesitation phenomena. They conclude that:

> Hesitations are not pre-linguistic in this sense; they function as auxiliary events which help to identify and circumscribe linguistic units, rather than as part of the raw data for which a structural statement must account. The fact that they serve this function shows a recognition of their non-random relation to linguistic form. (Maclay & Osgood, 1959, p. 39.)

Maclay & Osgood were also early in pointing out that a difference between the filled pause and the unfilled pause is that the former serves as a means for the speaker to "keep control of the conversational 'ball'" (Maclay & Osgood, 1959, p. 41), i.e. serving a floor-holding

[1] The 1968 reference is a collection of previous studies. Boomer (1970) provides a rather critical review of Goldman-Eisler's (1968) work, partly since she provides "virtually no references to contemporary psycholinguistic research outside the author's laboratory, although a good deal of work has been done elsewhere on each of the problems she has set herself" (Boomer, 1970, p. 162).

[2] Interestingly, Goldman-Eisler (1954a, 1954b, 1955), while not sharing data with Mahl et entourage, also described speech in psychiatric interviews.

[3] Ford & Holmes (1978) argued that the deep structure clause is the major unit of speech planning, which contrasts "the generally accepted view that the surface or phonemic clause is primary" (Ford & Holmes, 1978, p. 46). Cook, Smith & Lalljee (1974) argued that filled pauses "reflect processes of syntactic organization at the clause level, rather than at the sentence level" (Cook, Smith & Lalljee, 1974, p. 11), while Butterworth (1975) argued that "the speaker tends to plan ahead in terms of well-understood linguistic units–namely clauses and sentences" (Butterworth, 1975, p. 84).

function. The filled pause is also said to occur more often before lexical words than function words. The view that hesitation phenomena may serve a signaling function in English was also forwarded by Blankenship & Kay (1964). However, Cook (1971) did not find that filled pauses occurred more often before lexical words than at other locations.

Livant (1963) picked up the floor-holding thread proposed by Maclay & Osgood (1959) and argued that filled pauses serve "antagonistic functions", in that they do increase speakers' control of the conversation, but at the same time decrease the quality of their production. He based his conclusion on an experiment where subjects were told to solve mathematical calculations under two conditions, one silent and one when they vocalized (a filled pause), and found that calculations took longer time to solve during the vocalized condition. Livant thus concluded that while filled pauses arguably "jam" other speakers, they also "jam the speaker himself" (Livant, 1963, p. 4). Lallgee[1] & Cook (1969) argued against the floor-holding hypothesis, having found no support for the notion that filled pause rates should go up when pressure to speak was higher. They pointed out that the Maclay & Osgood (1959) study was based on *mono*logues, while their study focused on *dia*logues, and suggested that the floor-holding theory may apply only to monologues (sic!), whereas floor-holding in dialogues are achieved by other means, such as raising one's voice.

Siegman & Pope (1966), however, found that subjects in a monologue setting exhibited slower reaction times, more silence, slower articulation rates and fewer filled pauses than in a dialogue setting, and concluded that:

> The presence of another communicator in the dialogue situation compels one to respond promptly and not to be silent for long periods of time. As a result, potential silences are likely to be filled in by "ah's" and allied hesitation phenomena. (Siegman & Pope, 1966, p. 244.)

Boomer & Dittman (1963) made a distinction between "juncture pauses" that occur after a phonemic clause, and all other pauses, referred to as "hesitation pauses". They observed that hesitation pauses are discriminated better than juncture pauses at three different durations. Given a threshold of 75% correct discriminations, "the thresholds would be about 200 msec. for hesitation pauses and somewhere between 500 and 1,000 msec. for juncture pauses" (Boomer & Dittman, 1963, p. 217). Boomer (1965) found that the most frequent pause location was *after* the first word of a clause, rather than before it. Cook (1971) also made the observation that filled pauses tended to occur "at the beginning of a clause, either before the first word or before the second or third word" (Cook, 1971, p. 138).

Rochester (1973) found the time ripe to review two decades of studies of filled and silent pauses. She started out by pointing at the extent to which previous research had failed to be influenced by findings from related research:[2]

> Although both linguists and psychologists became interested in pausal phenomena at about the same time, they were rarely influenced by each other's perspective. Thus, psychological studies of pause location have tended either to ignore linguistic analyses or to use only weak approximations to them. At the same time, linguistic theories have not focused on speaker performance and consequently have not provided models of production. (Rochester, 1973, pp. 52–53. [3])

[1] Obviously the same person as Lalljee of e.g. Cook, Smith & Lalljee (1974).
[2] See also Boomer's (1970) review of Goldman-Eisler (1968).
[3] Fromkin (1971/1973) is commonly considered the first stab at a production model, and will be described later.

Rochester went on to describe how hesitation phenomena distribution was studied using two different methods of measuring transition probabilities, either a "guessing game" based on Shannon (1951), or the so-called **Cloze technique** (Taylor, 1953).

A good summary of the two techniques is found in Beattie & Butterworth (1979):

> In the Shannon guessing technique employed by Goldman-Eisler, judges guess each successive word in a sentence, they therefore have only the preceding linguistic context available to them when guessing. In the Cloze procedure, every nth word is deleted and replaced by a blank which judges attempt to guess. Therefore, both the preceding and the following linguistic contexts are available. The Cloze procedure thus yields a measure of contextual probability rather than strict transitional probability. (Beattie & Butterworth, 1979, p. 202, footnote.)

Cook (1969a), using the Cloze method, found that words following filled pauses had a lower transition probability than other words, with the exception of pronouns. Beattie & Butterworth (1979), also using the Cloze procedure, observed that words of low contextual probability in spontaneous speech were more likely to be hesitant, which was also the case for words of low frequency, and that when contextual probability was held constant, there was no difference in the word frequency between fluent and disfluent lexical items. They concluded that:

> [T]he contextual probability of lexical items in a continuous sample of spontaneous speech, as measured by the predictability of these words in context, is related to word frequency. Unpredictable, high-information, lexical items are significantly more infrequent that [sic!] are the more predictable lexical items. These results make it difficult to interpret the earlier studies. (Beattie & Butterworth, 1979, p. 208.)

Holmes (1988) found that hesitations occurred less often before embedded clauses than before other clause types. Moreover, silent pauses tended to occur before finite clauses, rather than non-finite clauses. Holmes concluded that:

> [D]eep structure clauses within surface structure clauses function as speech planning units /… / it is primarily pauses occurring before finite combined clauses in spontaneous speech that have a listener and/or breathing function. (Holmes, 1988, p. 323.)

As is seen, although a great deal of work is done on the distributional aspects of pausing (be it hesitational or junctural), the interpretations are not all that obvious. Various other methods have been employed to gauge the role of different parameters in the location of pauses in spontaneous speech,[1] but I won't go into them here.

Turning to the issue of different kinds of speech, Duez (1982) compared three different speech styles, political interviews, casual interviews and political speeches, the latter being carefully prepared. She found that silent pauses were 50% more common in the political speeches, while non-silent (filled) pauses were virtually non-existant.

In the *Discover* (magazine) report of Clark & Fox Tree (2002), it is pointed out that "[n]ot a single *uh* or *um* appears in the recorded inaugural speeches of American presidents between 1940 and 1996" (Glausiusz, 2002, p. 13). But, as unimpressed reader Bill Schmeer pointed

[1] For example, Butterworth (1980) employs a Stimulus–Response paradigm to measure the number of cognitive operations carried out by the speaker, and the delays (pauses) that certain stimuli provoke. Choice reaction time is measured using **Hick's Law**, since "choice response time is directly proportional to the number of alternatives when this is expressed as units of $\log_2$ [Hick's Law]" (Butterworth, 1980, p. 156).

out in a later issue of *Discover* (Schmeer, 2003): "presidential inaugurals and most, if not all, political speeches are scripted; and /…/ experienced public speakers prepare to deliver their speech" (Schmeer, 2003). This also agrees with the aforementioned observations that disfluency can be brought under control (see **2.3.4**), which of course is what professional (thespian) actors do for a living.

Grosjean & Collins (1979) studied breathing in a reading study, and found that both breathing and non-breathing pauses were dependent on both the rate of speaking and the syntactic structure of the pause location, but that non-breathing pauses were always shorter and tended to occur at minor constituent breaks. At fast rates, the differences disappeared, and the sole determinant became the physiological need to breathe. Grosjean, Grosjean & Lane (1979) and Grosjean (1980a) point out that the linguistic surface structure of a sentence was a good predictor of pause durations, and that speakers tended to place pauses between segments of equal length.

More recent studies have shown that pausing is a marker of discourse structure in other languages, such as Dutch (van Donzel & Koopmans-van Beinum, 1998; Swerts, Wichmann & Beun, 1996) and German (Serzisko, 1992). Watanabe & Ishi (2001) studied five different fillers in Japanese, and found that their distribution was different. While the fillers *e*, *eto* and *ma* tended to occur at major syntactic boundaries, others, like *sono* never occurred at sentence boundaries. They concluded that different kinds of fillers might reflect different speech production processes.

2.5.2 Disfluency in different social groups

From a sociolinguistic perspective, Bernstein (1962) observed that disfluency production was class-dependent, and that working-class subjects spent less time pausing, and exhibited longer phrase lengths, shorter mean pause durations and considerably shorter word-lengths than did middle-class subjects. The same pattern was also found for hesitation phenomena. Bernstein viewed the results as a function of different linguistic codes, elaborated and restricted, and concluded that "[m]iddle-class and working-class subjects /…/ are orientated to different levels of verbal planning which control the speech process. These planning orientations are independent of intelligence as measured by two reliable group tests and by word length. They are thus independent of psychological factors and inherent in the linguistic codes which are available to normal individuals." (Bernstein, 1962, p. 44).

2.5.3 Slips-of-the-tongue and spoonerisms

Besides focusing on pauses, early disfluency work within linguistics was carried out on slips-*of-the-tongue* (SOT). Much of this work served as the basis for linguistically motivated models for speech production, something that will be discussed in detail later on.[1]

Although everyone is familiar with the expression "Freudian slip", as observed and studied by Freud (1901/1973), the first study of slips from a linguistic point of view is normally attributed to Wells (1951/1973). A large number of studies followed, many based on elicited, induced, slips. An early example is Veness (1962) who used time pressure in a word association task to induce slips in her subjects. She found great individual differences, and although she did not perform precise personality measures, she attributed much of these differences to personality traits. Other early studies with their starting point in slips or

[1] A good collection on paper on slips is Cutler (1980).

spoonerisms were e.g. MacKay (1970, 1971), Fromkin (1971/1973) and Shattuck-Hufnagel (1979), to mention but a few.

Nooteboom (1969) found that words that were erroneously selected always belong to the same word class as the intended words, thus hinting both at underlying production processes and constraints.

Extensive work on laboratory-induced spoonerisms has been carried out by Bernard J. Baars and colleagues (e.g. Baars, Motley & MacKay, 1975; Motley & Baars, 1975; Motley, Camden & Baars, 1982; see also: Motley, 1980), and Baars also edited a collection of papers on "experimental slips", by himself and several other researchers (Baars, 1992a). Baars, Motley & MacKay (1975) started out by pointing out that slips, or spoonerisms, like *bad good–gad boof* or *darn bore–dart board*, can be elicited by priming subjects with bias items. They argue that the notion of a "slip" presupposes the existence of a rule-governed speech plan that occasionally fails to be executed correctly. They demonstrate that regardless of whether the priming is lexical or nonsense, lexical outcomes are significantly more frequent, and take their study as the first direct evidence of editing processes in speech production. Further support for prearticulatory editing of covert editing planning is given in Motley, Camden & Baars (1982), by studying potential anomalous (taboo) outputs

Much of the data studied by Baars and colleagues—and indeed data used in speech production in general—are *elicited* (slip) data. Ferber (1995) questioned the validity of such data. She agreed that spontaneous slips of the tongue have yielded valuable insights into speech production, but cautioned that such data are not easily verified. She made a distinction between *on-line* data, which means "jotting down" slips when they are heard, and *off-line* data (which refers to tape-recorded data), which is then later transcribed (which is what she used herself in her own studies).[1]

Others who commented on the reliability are Stemberger (1992), who remarked that it simply is not feasible to collect naturalistic errors (slips), since they occur with such low frequency, and Baars (1992d), who reviewed different methods for inducing speech errors. However, Levitt & Healy (1985) argued "that the experimental elicitation of errors provides critical tests of hypotheses generated by the analysis of naturally occurring speech errors" (Levitt & Healy, 1985, p. 717).

Besides questioning the quality of much of the data upon which many of the studies within the field have been based, Ferber (1995) also pointed out that there still is no satisfactory definition of the term slip-of-the-tongue.

However, slips are still the focus of linguistic studies, as is evidenced in e.g. Hokkanen (2001) who studied slips in Finnish, and Frisch & Wright (2002) who performed an acoustic analysis of slips in English. A good collection of papers covering the early period of slip research is found in Fromkin (1980).

[1] An important online collection of slip data is the London-Lund corpus (Garnham et al., 1982).

2.5.4 Tip-of-the-tongue

Another, related, area is that of *tip-of-the-tongue*, or TOT.[1] In a seminal article, Brown & McNeill (1966) defined TOT as "a state in which one cannot quite recall a familiar word but can recall words of similar form and meaning" (Brown & McNeill, 1966, p. 325). Brown & McNeill asked subjects to read definitions of English low-frequency words and asked them to recall the intended words. In that way, they collected several hundred of TOT states, and found that subjects had better-than-chance access to "letters"[2] of the intended words, number of syllables, and location of stress, before complete recall occurred. They concluded that words are stored in a mental dictionary, but that this dictionary has the form of an associative network, where various parts can be retrieved before full retrieval occurs, as mentioned above. They also pointed out that they do not regard TOT as something that only occurs on low-frequency words (Brown & McNeill, 1966, p.337).

Yarmey (1973) concluded that TOT states are retrieved from semantic and episodic memory systems, based on verbal and imaginary encodings, and pointed out that several retrieval systems must be at play in lexical access, in this case, name retrieval.

Browman (1978) compared TOT with slips of the *ear* (i.e., perceptual errors), and proposed a mechanism common to lexical and perceptual errors. This mechanism focuses on the beginning and ending of words, as well as to the initial portions of stressed syllables.

In later work, Caramazza & Miozzo (1997) found that Italian speakers had access not only to phonological information (e.g. the initial phoneme), but also to the word's gender, i.e. purely (morpho-)syntactical information.

So, what has TOT to do with SOT? Tweney, Tkacz & Zaruba (1975) pointed out that "the SOT phenomenon bears striking resemblance to another type of performance failure, the 'tip-of-the-tongue' (TOT) phenomenon" (Tweney, Tkazc & Zaruba, 1975, p. 388). They concluded that slips possess some of the same properties of TOTs, e.g. that the number of syllables and stress assignment in slips are almost always the same as the number of syllables and stress position of the intended word, which lends further evidence to the notion that lexical retrieval works on different aspects of the word to be retrieved, and that certain features are more likely to be retrieved than other aspects, as is obvious both in SOTs and TOTs.

Fay & Cutler (1977) studied malapropisms, i.e. the replacement of an intended word with another, erroneous word. Malapropisms are not really disfluencies in that they are real words. Rather they are related to slips in that there is a misexecution of the intended plan. Malapropisms are typically unrelated in meaning, but bear phonetic resemblance to the intended word. Like slips, they are almost always the same word class as the intended word, as well as the same number of syllables and stress pattern (Fay & Cutler, 1977, pp. 507–508). It goes without saying that these observations have potential consequences for speech production models.

---

[1] As is often the case, William James (1890) was the first to draw attention to a phenomenon. TOT is no exception, as acknowledged by both Brown & McNeill (1966) and later researchers. NB! Brown & McNeill give the reference as 1893, which is then repeated by Yarmey (1973). All other information I have come across, including my own facsimile edition, states 1890.

[2] They probably mean phone(me)s.

Finally, rounding off this section, slips have also been used as a tool in other areas of linguistics. Davidsen-Nielsen (1971) studied slips in his phonological analysis of the English consonant clusters *sp*, *st* and *sk* to establish whether a monosegmental or bisegmental interpretation of the said clusters were more correct. He concluded that the speech error evidence supported a bisegmental interpretation since all three clusters were frequently broken down into their constituent parts in slips.

2.5.5 Prosody

An aspect that has been lacking in the previous discussion of speech is that of **intonation**, the way speech is executed as to melody, duration, pitch, fundamental frequency and other such phenomena. There is no doubt that intonation, or prosody, constitutes an important part of language/speech, although it is employed in different ways in different languages. There is also compelling evidence that prosodic patterns are processed in other ways than e.g. lexical retrieval in processing (in some speech errors the wrong lexical form is uttered, with the stress pattern of the intended word, *vid.* Boomer & Laver, 1968/1973, p. 129; Garrett, 1975, p. 147) or in perception or comprehension. This phenomenon also occurs at higher level, like Cutler's observation that "primary sentence stress often does not shift when the element that would carry it in the target utterance shifts" (Cutler, 1980,[1] p. 75; see also Fromkin, 1971/1973, pp. 42–43). Also, some pauses are clearly linguistic means with a structuring function, and are both produced and perceived as such.

So, where does intonation fit into our conception of language? Bolinger (1983) asked the same, rhetorical, question, and provided an attempt to an answer:

> [I]ntonation belongs wherever people have a use for it.[footnote removed here.] It belongs in syntax, because it helps to mark the start and finish of stretches of speech such as clauses and sentences. It belongs in pragmatics because it is the best audible cue we have as to what a speaker is doing with his utterance. It belongs in psychology because it gives a running account of emotion and counts among the symptoms of certain brain disorders. (Bolinger, 1983, p. 101.)

Given that few would argue with Bolinger as to the role intonation plays, one could safely assume that prosody also plays a part in the fluency of speech, with the ensuing consequence that it plays a part in the disfluency of speech. However, it has also been shown that prosody varies across different kinds of task of speech modes and settings. Shriberg et al. (2000) observed that while pause and pitch information were of importance in the segmentation of broadcast news speech, duration and word-based cues were more important for natural conversations. A number of studies have tried to establish the kind of relationship prosody has with the different categories of disfluency we have discussed so far. Lickley (1994) observed that prosodic information was used by listeners to distinguish between fluent and disfluent utterances, and that a combination of acoustic and prosodic cues was used. Moreover, Lickley (1996) argued that juncture cues typical of fluent speech are absent in disfluent speech at the interruption point.[2] The absence of this "fluency linking" helps listener detect disfluency.

---

[1] Cutler (1980) also pointed out that: "on closer inspection it turns out that the stress pattern is preserved only when both the words involved in the shift are open class items. When closed class words shift or exchange, the stress moves with its bearer" (Cutler, 1980, p. 76).

[2] Speech repairs are commonly divided into two parts, the *reparandum* (the erroneous item(s) being repaired and the *reparans*, or *repair*, the item(s) that replace the erroneous material. These are separated by the inferred *interruption point*.

Hieke (1981) classified repetitions into two distinct types, *prospective*, used to hold the floor, and thus similar to filled pauses as to communicative function, and *retrospective*, with a bridging function between the continuation (reparans) and the preceding material. Shriberg (1995) found that durational and fundamental frequency properties support Hieke's categorization of repetitions. In later work, Plauché & Shriberg (1999), replicated Shriberg's (1995) support of Hieke's two categories, but also found support for a third type, which they call *covert self-repairs*, exhibiting a distinctive prosodic pattern.

Levelt & Cutler (1983), studying speech repairs, found a relation between prosodic marking and semantic factors, while they did not observe any relation between prosody and syntactic structure.

In a study on clause-internal filled pauses, Shriberg & Lickley (1992a, 1992b, 1993) found that the fundamental frequency of filled pauses is related to prior prosodic context, thus lending support to a *relative hypothesis*, indicating a systematic relationship between a preceding peak F_0 value and the F_0 value of the filled pause.[1]

From an application-based point of view, prosodic and acoustic information has so far and most often been neglected in automatic speech recognizers. However, several studies have shown that the inclusion of prosodic information enhances the performance of such systems considerably (e.g. Shriberg & Stolcke, 1996, 2004; Shriberg, Bear & Dowding, 1992; Baron, Shriberg & Stolcke, 2002; O'Shaughnessy, 1992a; Shriberg, Stolcke & Baron, 2001 and Nakatani & Hirschberg, 1993, 1994; see also Ostendorf, Price & Shattuck-Hufnagel, 1997 for an overview) or are of help in the detection of disfluency (e.g. Shriberg, Bates & Stolcke, 1996, 1997).

Wightman et al. (1992) studied the relationship of segmental lengthening and prosodic phrase boundaries, and found that segmental lengthening was correlated to the rhyme of the syllable preceding the (prosodic) boundary. Although "lengthening" here is not equal to prolongation (as a disfluent category), it must be borne in mind that prolongation can be the result of such phenomena as *prosodic phrases*, *prosodic words* or *phrasal accents*, and that a disfluency interpretation of extra-long segments should, if possible, be pitted against other possible explanations, such as a structure-giving device in the prosodic realization of phrases.[2] Finally, Stirling et al. (2001) report ongoing work on the interaction of prosody and discourse structure (using silent pause location, among other phenomena), and also discuss methodological issues associated with such analyses of speech corpora (in this case the Australian map task corpus).

That prosodic information can improve automatic speech recognition in language other than English is shown in e.g. Lee & Chen (1997) for Chinese, or Tseng (1999) for German.

2.5.6 Disfluency as a conversational tool

While some of the approaches above make the tacit assumption that disfluencies are "detriments" in the speech signal, or evidence of problems in the production of linguistic

[1] The alternatives were either an absolute hypothesis, where filled pauses occurred at constant, speaker-dependent F_0 values, or a random hypothesis, where filled pauses occurred in ways unrelated to any prior prosodic context.

[2] Intonation phonology is beyond the scope of this work, but Wightman et al. (1992) provide a succinct introduction to the field, including some of the more influential references.

messages, an alternative way to view the occurrence of hesitation phenomena is to regard them as a linguistic *means* that contain meaningful information in its own right. Rather than focusing on linguistic competence, from an idealized linguistic perspective, human communication is viewed as an interactive activity, where **speech acts** (Austin, 1962/1975; Searle, 1969) are carried out, with the main objective of achieving goals common to both speaker and listener. Thus viewed, disfluencies are not necessarily detriments or flaws in the communication, but could instead be viewed as part of the communicative game, and provide valuable information. In this section, some of the research adhering to this stance will be reviewed.

2.5.6.1 The role of *um*, *uh* and (silent) pauses

As was mentioned in the previous paragraph, disfluency might be regarded as a phenomenon with communicative *function*. Clark (1996) listed a number of "suspension devices" such as pauses, word cut-offs, elongation and fillers, and remarked that "[s]uspension devices aren't produced accidentally. The are the result of the speaker's own actions – they are *self*-suspensions – and are signs in Peirce's sense" (Clark, 1996, p. 261; italics in original).[1] Clark & Wasow (1998) pointed out that there are two complementary ways one could view disfluency: the first treating disfluencies as the outcome of *processes* that once initiated cannot be controlled by the speaker, thus eschewing any notion of intention or purpose on behalf of the speaker, the second regarding disfluencies as the result of speaker *strategies* under speaker control. Thus, speakers have different options at hand when speech production turns problematic. As an example, speakers tend to use filled pauses when they expect a long delay, and unfilled pauses when there is only a brief interruption in speech production. Along the same lines, Clark & Fox Tree (2002) found that speakers make a difference between *uh* and *um*, the former being used to signal shorter breaks, the latter for longer breaks, supporting the "filler-as-word hypothesis" according to which *uh* and *um* are English interjections.

Fox Tree (2001) found that *uh* had a beneficial effect on listeners' ability to understand ensuing words in upcoming speech, while *um* did not produce any such effect, either beneficial or detrimental. Thus, disfluency might in fact help listeners understand spoken utterances, something that clearly support the notion of disfluency as something with a communicative function.[2]

Brennan (2000) also found that listeners were quicker to recognize speaker intentions relative to target words in disfluent utterances than in fluent utterances. Brennan also concluded that listeners used latencies, especially those that included filled pauses to signal their *degree* of uncertainty, their *Feeling-of-Knowing*, in her words.

Schachter et al. (1991) tested the hypothesis that filled pauses (*er*, *uh*, *um*) indicate that the speaker is facing several options as to how to proceed speaking. The corollary hypothesis would be that speech with more inherent options should exhibit more filled pauses. To test this hypothesis, Schachter et al. studied lectures within three disciplines with varying degrees of inherent optionality: Natural science, with very few options (there are very few options of describing the orbit of a planet or the outcome of a chemical reaction), social science (with an intermediate degree of available options) and humanities (with an infinite number of ways to

[1] Clark is referring to Charles Sanders Peirce's theory of signs or semiotics. The reader is referred to pp. 156–161 in Clark (1996).

[2] In a previous study, Fox Tree (1995) found that false starts were detrimental to speech understanding, while repetitions were not.

describe, for example, what Shakespeare meant in a certain passage). The hypothesis was confirmed. Lecturers within the humanities used more filled pauses than lecturers within social sciences, who, in turn, used more filled pauses than did lecturers within natural sciences. To rule out individual differences, the same set of lecturers also gave talks on a common subject, in which case they all produced an equal number of filled pauses.

As has previously been mentioned, that filled pauses function differently from other disfluencies have been shown over and over again, e.g. by (starting with Mahl, 1956, 1958), who developed his *non-ah ratio* to count disfluencies since filled pauses did not vary as a function of anxiety the way other disfluencies did.

Brennan & Schober (2001) exposed subjects to fluent and disfluent instructions to select an object on a graphical display. They found that instructions containing interruptions with filled pauses resulted in the fastest responses, and faster than either completely fluent instructions or interrupted instruction with silent pauses. They concluded that filled pauses helped listeners comprehend the instructions given to them.

2.5.6.2 Speech Management

Instead of regarding language as a formal system where an idealized, perfect, "competence" should be regarded as the basic phenomenon, with imperfect surface performance, evident from phenomena like disfluencies, human language can also be seen as an interactive game, where speech acts are carried out to convey not only meaning in a purely semantic sense, but also extralinguistic information about mental states, wishes, desires and so on constitute important units. This is the basic unit in Allwood's model of human interaction. The concept of **speech acts** was introduced in Austin (1962/1975) and was then further developed by Searle (1969). Although Allwood does buy Austin's and Searle's general views on communication (some critical points are found in Allwood, 1977), Allwood's model is above all rooted in the view that human communication is *interactive*.

The concept of **Speech Management** (SM) was introduced in Allwood, Nivre & Ahlsén (1990), where it was argued that editing and self-repair belong to a systematic linguistic system whereby speakers manage their linguistic contributions, and are related both to intraindividual phenomena such as memory and planning, but also to interindividual phenomena such as turntaking (in dialogue) and feedback (Allwood, Nivre & Ahlsén, 1992). While typical speech management phenomena include repairs, hesitation, corrections, repetitions, reformulations and so on, speech errors without signs of external management, e.g. slips of the tongue are not part of the system. Allwood et al. pointed out that studies of speech management have either been oriented towards psycholinguistics (e.g. Levelt, 1989), or towards sociolinguistics (e.g. Schegloff, 1979; Sacks, Schegloff & Jefferson, 1974; Schegloff, Jefferson & Sacks, 1977; Schegloff & Sacks, 1973). Allwood et al. argued that speech management must include these different disciplines simultaneously, and they present a taxonomy with that as the objective. The main point is that disfluencies in reality should be seen as a normal, informative, natural phenomenon of spontaneous human speech, and thus part of what contributes to fluency in speech. Human communication is characterized by *turns* —with a slightly different definition than given in Sacks, Schegloff & Jefferson, 1974, as pointed out in Allwood (1988a)—where phenomena like disfluencies are communicative *tools*, rather than detriments, that help getting the message through.

There are obvious similarities between Allwood's program and Clark's.[1] In a way, one could claim that both models ultimately are rooted in interactive views on human communication, as forwarded in e.g. Grice (1975, 1989/1997) or Schegloff & Sacks (1973). However, Allwood (1997a) prefers to view human dialogue as *cooperative*, as opposed (albeit similar) to Clark's term *collaborative*. Indeed, to further stress the view of dialogue as interactive, Allwood goes so far as to label human communication as "collective thinking" (Allwood, 1997b). The view that utterances should not be studied, or considered, by themselves was also forwarded by e.g. Goffman (1978), who pointed out that, according to the interactionist view, "every utterance is a statement establishing the next speaker's words as a reply, or a reply to what the prior speaker has just established, or a mixture of both." (Goffman, 1978, p. 787).

In the same vein as Clark, Allwood argued that self-corrections, hesitation, feedback and so on primarily exist as a management tool in dialogue (Allwood, 1994a). Allwood, like Clark, views human communication as fundamentally interactive, and that management of a speaker's own contributions in a dialogue, referred to as **Own Communication Management** (OCM) can be distinguished from **Interactive Communication Management** (ICM) (Allwood, 1995). The basic categories of communication provided by Allwood and colleagues are summarized below (as presented in e.g. Allwood, 1988a).

ERM: Explicit Referential Message The linguistic meaning carried in a sequence of words in an utterance, or *turn*. This is what is normally studied in linguistics, i.e., corresponds to the default notion of language.

ICM: Interactive Communication Management A term to describe procedures and mechanisms whereby interlocutors manage their communication. This includes both linguistic means of giving feedback, such as the use of words like *yeah*, but also extralinguistic phenomena like head nods, gazes and so on.

OCM: Own Communication Management Consists of procedures and mechanisms with which speakers manage their own communication, i.e., means signaling that the speaker needs time to plan or choose how to proceed speaking, or change things already said. These include hesitation sounds such as *eh*, or explicit editing terms and so on. OCM includes most of the phenomena elsewhere regarded as disfluency, but are here seen as tool to enhance comprehension from the listener's point of view.

B: Background Information Basically contextual information—physical and linguistic—needed to interpret any turn, or utterance. Words like *you* require that there is a physical person who could be described by that word, while words like *that* need a linguistic referent to carry meaning.

2.5.6.3 "Conversational grunts"

That disfluencies have a communicative role in language is also forwarded by Ward (2000). He supports the notion that "non-lexical speech sounds" such as *ah*, *uh*, *u-huh*, *yeah*, *okay* and so on are important in conversation, both for control and for conveying attitudes. The term Ward uses for the items he discusses is *conversational grunts*, which includes some (but evidently not all) of the disfluencies discussed in this thesis. An example would be the filled pause *eh*, which Ward also refers to as a *disfluency* or *filler*. While including items that are

[1] For instance, Clark & Fox Tree (2002) acknowledge that their filler-as-word hypothesis "owes much to Allwood" (Clark & Fox Tree, 2002, p. 79).

obviously lexical, like *okay*, his list encompasses many more items, such as "unusual voicing", "superimposition of phonetic components", spectral stability", "sound symbolism" and so on and so forth.

Ward summarizes his view on "Fillers and Disfluency Markers" (from an application-point of view) thus:

> Different users have different information-uptake capabilities. Inserting fillers and disfluencies in system output may be a relatively easy way to reduce the information transmission rate. Appropriate fillers and disfluency markers may also assist the listener by signaling what sort of information is coming up, how long it will be, and so on, so that he can deploy his attention appropriately. (Ward, 2000, p. 573.)

This obviously strikes the same cord as do Clark, Allwood and colleagues.

2.5.6.4 Support from the stuttering community

The view that disfluencies really are a kind of fluency is voiced even in the stuttering community. Starkweather (1987), in his chapter on the development of fluency in children, discusses the etiology of disfluencies[1] thus:

> What kind of behavior are these discontinuities? Are they stumbles in the forward flow of speech, errors of speech production, or do they result from a more purposeful intention? The parenthetical remark—"well", "you know," "what I mean to say"—and the filled pause or interjection—"uh"—are surely more than stumbles. As behaviors, they result in our being able to stall for time so as to keep talking while we think or plan or edit. They may not have very much meaning—indeed, their purpose is to fill up time at a point when the speaker has nothing meaningful ready to produce—but the parenthetical remark is a coordinated and studied use of language. It appears to be a more elaborate version of the filled pause, a speaker's way of keeping the floor while gaining a little more time to produce next utterance. So the filled pause and the parenthetical remark do not seem to be errors. (Starkweather, 1987, p. 86.)

Having thus stated that some forms of language disfluency might in fact be conversational tricks, rather than errors, he goes on to discuss the status of other forms of errors:

> If we can accept the idea that the parenthetical remarks serve a correcting function by providing the speaker with time to revise or better plan utterance, it should be difficult not to see the false start, revision, and incomplete phrase also as corrections, essentially the same kind of corrections as the parenthetical remark, except that the error isn't quite detected until after the utterance has begun. /…/ The tense pause and disrhythmic phonations are briefer and harder to ascertain. They may be very small stalls or they may be stumbles. In any event, they are a less frequent category. Repetitions, the form of discontinuity found in the youngest children, may be errors. But even in the case of repetitions, it is not clear that they are errors. /…/ It seems then that all of the discontinuities that are vocalized, with the possible and important, exception of the part-word repetition, represent corrections, or a correcting function, rather than stumbles, slips, or errors in the production of speech. It should be noted that these discontinuities also help the listener in a number of different ways. They are conversational devices. /…/ The odd thing about these discontinuities is the persistent belief that they are errors of speech. (Starkweather, 1987, pp. 86–87.)

[1] Starkweather prefers the term *discontinuities*.

As is evident from this survey, Allwood and colleagues look at language from a very wide perspective, where the focus is human communication in general. The main issues are questions like how do human beings communicate, how do they contribute to carrying messages, both linguistic and non-linguistic through, how do they manage both interactive processes (the conversation as a whole), their own contributions, and how do they incorporate and adapt to contextual, world-knowledge, phenomena "present" in a physical, linguistic or even philosophical sense?

It goes without saying that this is indeed an ambitious program, and definitely helps to explain why disfluencies are still around in language, despite thousands of years of practicing. It is also evident that much of what is incorporated in this model is presently beyond reach from an application-perspective, since it probably requires that the AI problem be solved—at least to some degree—in order to even begin including much of what is viewed as central in the Speech Management model. For readers wanting to delve deeper into Allwood's (et al.) model, Allwood, Nivre & Ahlsén (1990) provides the most exhaustive presentation.

2.5.7 Summary

As is shown in this section, early disfluency work within general linguistics focused on two phenomena, **hesitation** (which is by far the most common), and **slips-of-the-tongue** (which is very rare, in fact so rare that many of the studies were carried out on elicited data). The works referred to above consequently had a less fine-grained typology of disfluencies than was the case within stuttering research or psychotherapy, with a couple of exceptions (such as Blankenship & Kay (1964). The outcome of these two focus points could be seen as having resulted in two contingent fields:

Studies on hesitation laid the ground for the view that **pausing** might indeed be a linguistic means rather than a detriment, which is, basically, the position held by e.g. Clark and Allwood, where the focus has been to view language/speech as it exists in the society of human beings who interact. This incorporates such views as forwarded by philosophers such as Austin and Searle, and the notion of **speech acts**, that verbal messages are interactive (collaborative, cooperative) actions that take place, in the context of a culture, between human beings with specific goals.

Studies on slips-of-the-tongue were the basis for speech production models, where the focus turned inwards, to what happens inside the brain when we speak, where other philosophical issues are raised as to more ephemeral notions such as consciousness and free will.

In the next section, we will look at speech production.

2.6 Speech production

Speech—and disfluency, the way we are discussing these phenomena here—is, of course, *produced* by someone, the speaker. *How* speech and disfluency are produced has been the object of study in much research, and within a variety of different fields, ranging from general linguistics, philosophy all the way to pure neuroscience.

2.6.1 Introduction

A very fundamental question throughout the history of humankind has been what human **consciousness** really is.[1] What does it mean to be conscious? Are there different "consciousnesses"? Is there a way to disinter this, perhaps the innermost, secret of what it means to be human. The issue of consciousness and volition has been addressed from a wide variety of disciplines such as **philosophy** (e.g. Dennett, 1991; Churchland, 2002; Flanagan, 1992; Devitt & Sterelny, 1987), **psychiatry** (Jaynes, 1976/2000, 1980, 1986, 1990), **neurology and neurobiology** (Damasio, 1999; Crick & Koch, 1990; Koch & Crick, 1991; Gazzaniga, 1992; Deacon, 1997, Sperry, 1976, 1980; Ingvar, 1999; Spence & Frith, 1999; Schultz, 1999), **anaesthesia** (Hameroff, 1998a, 1998b; see also: Dixon, 1989), **biology** (Edelman, 1992; Edelman & Tononi, 2000), **physics** (Penrose, 1989, 1990; Davies, 1987, 1995; Stapp, 1999, 2001; Mohrhoff, 1999; Wilson, 1999; Hodgson, 1999), **linguistics** (Chafe, 1994; Jackendoff, 1995), **computer science** (Copeland, 1993) and so on and so forth. This question proper clearly goes beyond the scope of the present work, and I will not delve further into this field in general. However, *one* way of trying to obtain at least partial answers to these questions, whether or not this is explicitly mentioned, has been the study of disfluencies within the field of speech production, to be described in the following.

It is often stated that speech/language is what most separates us from the other animals on this planet, i.e., the most human of all our traits. Granted, many, if not most, animals do communicate in one way or another, employing various kinds of signal systems, with more or less specified meanings, but human language stands out among all these signal systems, especially the recursiveness human language exhibits, which Hauser, Chomsky & Fitch (2002) claim is the only truly *unique* component of human language as compared to signal systems employed by other animals.

Although it has always been a matter of debate to what extent language is "pure thought", merely *reflects* inner cognitive functions, or is a *prerequisite* for inner cognitive processes, the problem remains constant: the study of language is most often stuck with *outdata*, nothing else. The only thing attainable to study is what we say, and there is no way we can alter or manipulate either the "black box" or whatever is fed into it. Consequently, all assumptions concerning language production need be based on the study of the "final product", as it were.

Daniel Dennett (1991) pointed out that this very problem is the reason that linguistics is replete with studies on language *perception*, as opposed to production, or as he puts it:

> Utterances are readily found objects with which to begin a process. It is really quite clear what the raw material or input to the perception and comprehension systems is: wave forms of certain sorts in the air, or strings of marks on various plane surfaces. And although there is considerable fog obscuring the controversies about just what the end product of the comprehension process is, at least this deep disagreement comes at the end of the process, not the beginning. (Dennett, 1991, p. 231.)

A basic question is whether language perception is some kind of *decoding* or *translation*. Do we understand the world around us by dint of language "as we know it", or is our world knowledge represented, deep inside our brains, in some kind of "mentalese", in the form of semantic deep structures, or similar?

[1] For a recent, synoptic and succinct introduction to consciousness research and associated different stances and schools therein, see Zeman (2001).

Perception studies can be carried out by the study of observable phenomena, production studies cannot so easily be carried out this way. However, one way of indirect insight into speech production is to look at speech errors, i.e. disfluencies. By looking at how things "went wrong", one can gain insight into the inner structure of speech production. So it comes as no surprise that most, if not all, speech production models have focused on speech errors of different kinds.

Or, as Chafe (1994) puts it:

> Finally, is consciousness just a matter of talking to oneself, or are people conscious of more than language? The fact that consciousness consists *in part* of inner speech cannot be in doubt, but it is obvious at least from my own introspection that not all of what passes through my consciousness is language. /…/ It may seem paradoxical that language itself provides evidence that consciousness contains more than language, but in fact there are linguistic reasons to believe that the content of consciousness at any moment cannot be equated with any particular linguistic manifestation of it. /…/ One kind of evidence is the presence of disfluencies. People often have trouble "putting thought into words" and may believe that they have not adequately stated what they "had in mind". If people were conscious of nothing more than words to begin with, the task of overt verbalizing should be effortless, simply a matter of vocalizing what was already present subvocally. But almost any observation of natural speech shows that talking is not that easy. Disfluencies are evidence for a nonconformity between what one is conscious of and what one says. (Chafe, 1994. p. 34.)

In the following sections I will summarize some speech production models that appear in the literature.

2.6.2 Early models of speech production

Whereas slips-of-the-tongue have been known for a long time—just consider "Freudian slips"—the first stab at describing them from a scientific/linguistic point of view may well have been Wells (1951/1973), who also started out by referring to Freud. Other early work was also carried out by Lashley (1951), Morton (1964) and Cohen (1968/1973). However, the beginning of speech production models, influenced by slips, began later.[1] Hockett (1967/1973) introduced the notions of "editing", which could be either "overt", errors that are uttered, and then corrected, or "covert", errors that are corrected before being uttered, implying "inner speech" which can be attended to.[2]

Laver (1969/1973, 1970, also 1980a, 1980b) presented a neurolinguistic model of speech production. In the 1969/1973 work, Laver identified four[3] functions: **ideation**[4] (the "idea"), **neurolinguistic program-planning** (the Planner; lexical, grammatical and phonological information), **myodynamic execution** (muscle execution) and **monitoring** (the Monitor; detection and correction of errors). The Monitor includes both auditory and tactile

[1] For a review of early speech production models, see Butterworth (1981).

[2] Faaborg-Anderson & Edfeldt (1958) showed that silent speech is accompanied by electrical activity in the intrinsic laryngeal muscles. Paulesu, Frith & Frackowiak (1993) examined the neural correlates of working memory during inner speech, and point out that "[a]lthough it cannot be ruled out that some unconscious oro-pharyngeal activity was occurring during the experimental tasks, 'inner speech' is probably not dependent on such movements." (Paulesu, Frith & Frackowiak, 1993, p. 344).

[3] A fifth, separate, function was added in Laver (1970): that of long-term storage of linguistic information. In Laver (1980b), it is back to four functions.

[4] Laver refers to James (1890) in that "ideation is to be distinguished from the linguistic resources that are exploited to construct expository linguistic programs" (Laver, 1980b, p. 292).

exteroceptive reports, as well as positional and kinesthetic proprioceptive reports. Given how few slips we produce, Laver assumes that monitor surveillance must be nearly constant. Moreover, since speakers often correct mistakes without even being aware of doing so, monitoring is assumed to be automatic, and be able to operate without conscious awareness. Laver equates his Planner with Hockett's covert editing, and his Monitor with Hockett's overt editing.

Fromkin (1971/1973) presented an "utterance generator", sometimes considered the first production model that aimed at incorporating all steps from ideation to spoken-out speech. Or as Butterworth (1981) puts it:

> Fromkin (1971) was the first to make the much bolder step of trying to relate errors in a systematic way to an integrated linguistic theory (generative grammar, with emendations) ranging from syntax and lexical selection to phonetic features, and to sketch a performance model — 'utterance generator' — to collate the linguistic levels into a single, psychologically plausible system. (Butterworth, 1981, p. 634. [1])

Like most early disfluency studies within linguistics, it focused on slips and spoonerisms. Her proposed model includes five stages:

Stage 1: A meaning is generated.

Stage 2. The meaning is structured syntactically and semantically.

Stage 3. The output of stage 2 is given an intonational contour and stress pattern.

Stage 4. Lexical lookup occurs.

Stage 5. Phonetic and phonological rules convert the sequence into neuromotor commands.

As we shall see, this basic scheme later reappears in a variety of guises.

Garrett (1975, 1980a, 1980b) picked up from Fromkin (1971/1973), and proposed a speech production model based on speech errors, or slips. Garrett (1975) held the view that sentence production must be viewed as a translation process between a message and instructions to the articulators, by way of some kind of psychological representation. A principal feature of Garrett's proposal is the separation of meaning-related and form-related processes.

Other models, with varying degrees of specificity, have also been proposed, e.g. Shattuck-Hufnagel (1979). Garnsey & Dell (1984) argued that a complete model of speech production must include a prearticulatory editing component, with the specific function of monitoring inner speech. Garnsey & Dell also pointed out that speech production models must be able not only to provide explanations for normal speakers, and their respective speech errors, but also to cover speakers with pathological symptoms, like aphasia. Harley (1984) argued against top-down models of speech production, based on findings that imply that phonological similarities between the intended target and the intrusion is a major determinant in error occurrence.

[1] Butterworth characterized Fromkin (1971/1973) as "Linguistics meets errors: Fromkin" (Butterworth, 1981, p. 633), and Garrett (1975, 1980a, 1980b) as "Psychology meets errors: Garrett" (ibid., p. 640).

Bock (1982; see also Bock, 1987) presented an ambitious speech production program that, although its focus was on syntax, included a discussion on various phenomena, such as semantics, phonology, phonetics, motor assembly, tip-of-the-tongue phenomena and other speech errors, and so on. She presented a "General Framework for Utterance Formulation" that includes all of the above in separate modules. Garrett's "Message" is replaced with a "Referential Arena" which employs two routes, a syntactic and a semantic-phonological (with unidirected information being passed from the semantic-phonological modules to the syntactic module), that both connect with a "phonetic coding" module, which in turn feeds into the motor program. The Referential Arena, and the Phonetic Coding module both have access to a "Working Memory" module, in a bidirected way. Bock's proposal takes into consideration a wide variety of communicative aspects, and tries to encompass not only speech errors and slips, but also phenomena like focus of attention, givenness (of information), discourse functions and so on, however with a main focus on syntax.

De Smedt & Kempen (1987) proposed a "global" speech production model with four main modules: a conceptual module, a lexico-syntactic module, a morpho-phonological module and an articulatory module. However, they favored the view that these modules can work on different parts of the utterance simultaneously, in a more parallel, way, and suggested the term "incremental production", of "streams", to describe such a production model.

2.6.3 Levelt's model of speech production

Despite the previous attempts mentioned above, the first full-fledged stab at a speech production model is normally attributed to Levelt (1983a, 1983b, 1989). While Bock based her model mainly on higher-level linguistic phenomena such as syntax, semantic frames and so on, the proposal made by Levelt (1983a, 1983b, 1989) is to a large degree focused on speech errors. Levelt's model of speech production consists of a set of modules, shown in **Figure 2.1**.

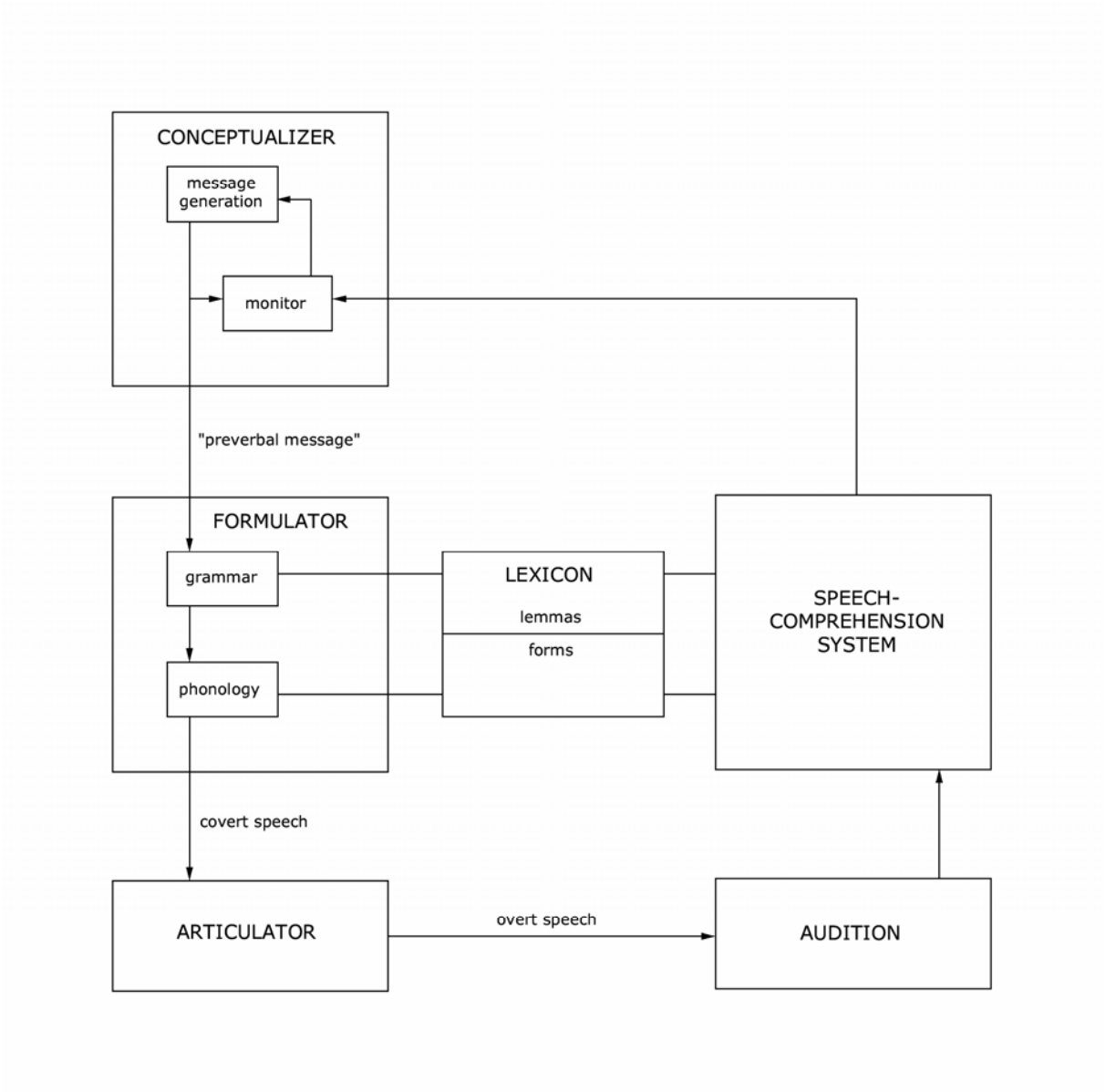


Figure 2.1: Levelt's model. Simplified version of Levelt's "blueprint for the speaker" (Levelt, 1989, p. 9).

First, a **Conceptualizer** creates a message. The Conceptualizer has access to some kind discourse model, a situation model and so on. This resulting preverbal, or prelinguistic, message is then fed into the next module, the **Formulator**, which collects the suitable linguistic units (words, sounds, intonation etc.), i.e., provides the message with the required linguistic form. The Formulator contains a submodule for grammatical encoding and another submodule for phonological encoding, in order to yield a linguistic surface structure of the message to be uttered. This linguistic form is then fed into the next module, which is the **Articulator**, which governs the suitable set of articulators (muscles) to produce the message.
This is simple enough. What is more crucial, and brings disfluencies into the picture, is that the model also includes a **Monitor**, which serves as a control function. The Monitor can be regarded as a process that supervises the speech production process, and detects and repairs errors. In Levelt's model, the Monitor has two important traits:

- It resides inside the Conceptualizer. This means that generation and control takes place in the same module.

- It makes use of the speech understanding system, i.e., the system we are using to understand others is also used to interpret our own speech.

The Monitor does not require acoustic input, but can work at an earlier stage. This means that it works by dint of two different loops, an *outer loop*, which makes use of the acoustic signal, and an *inner loop*, which keeps track of the message throughout the process. The inner loop can even detect errors before anything has been passed on to the Formulator, i.e., it can change messages already at the Conceptualizer level.

2.6.3.1 Comments on Levelt's model

Several people have proposed alternative monitoring systems, some in response to Levelt, some earlier. One such alternative would be that monitoring feedback occurs in a proprioceptive or tactile way, i.e., that the articulatory muscles per se send back information to the Monitor, which then can detect errors. This was proposed by Borden (1979) and Lackner & Tuller (1979). Postma & Kolk (1993) suggested that Levelt might have missed the possibility of proprioceptive feedback. Kimble & Perlmuter (1970),[1] on discussing volition in general, conclude that:

> From Taub's[2] work we must conclude that proprioceptive feedback is *not* essential to the control of voluntary behavior. There is considerable evidence, on the other hand, that such feedback normally plays an important role. (Kimble & Perlmuter, 1970, p. 370; italics in original.)

The important point here is of course, that the fact that something is not necessary does not entail that it is not used.[3]

Another alternative, proposed by e.g. Laver (1969/1973, 1970, 1980a, 1980b), Schlenk, Huber & Willmes (1987) and Van Wijk & Kempen (1987), is that the formulation process, i.e., the collection of lexical material, and the organization of syntactic units and so on, might be subject to monitoring. This alternative has been rejected by Levelt (1989), since this alternative requires that a **production model** exists, which in turn means that cognitive processes proper would be accessible for attention or supervision, something Levelt does not believe is the case. Another problem production models entail is that information needs to be "doubled", i.e., the information that is used inside a module must also exist outside that module for comparison between the initial program (as conceived by the Conceptualizer) and the program that is now being executed. Also, such a system would possibly be slowed-down considerably, as compared to Levelt's "flow-through" monitor, where monitoring occurs in parallel with production.

---

[1] Note: Virtually all sources cite this work as "Perlmutter" with two <t>:s. On the paper the second author's name is given as "Lawrence C. Perlmuter", with only one <t>.

[2] For example: Taub, E. & A. J. Berman. 1963. Avoidance conditioning in the absence of relevant proprioceptive and exteroceptive feedback. *Journal of Comparative and Psychological Psychology*, vol. 56, pp. 1012–1016.

[3] Recent evidence for somatosensory feedback in speech production comes from work on adults who have become deaf as adults, but continue to produce intelligible speech for several years (Tremblay, Shiller & Ostry, 2003).

Monitoring has been widely discussed in the literature, and quite often poses considerable problems of both philosophical and neurological nature. It has been suggested that monitoring probably is not continuous, but instead intermittent (Borden, 1979; Neilson & Neilson, 1987), which would explain why it is not foolproof, but misses errors occasionally. Monitoring also seems to depend on motivation and attention (Laver, 1969/1973), and that it is both context- and type-driven (Baars, Motley & MacKay, 1975).

I will refrain from covering the entire monitoring discourse, here. However, I will describe alternative models in some detail, beginning with Postma & Kolk's *Covert Repair Hypothesis*, which is up next.

2.6.4 Postma & Kolk: the Covert Repair Model

Postma & Kolk (1990, 1991, 1992, 1993; Postma, Kolk & Povel, 1990; see also Postma, 2000) assume that some kind of monitoring process is part of human speech production, based on the fact that self-repairs exist in human speech. They also agree with Levelt in that error detection can occur both auditively, through an outer loop, as well as on-the-fly through an inner loop. The latter, of course, has as a consequence that speech errors can be detected and repaired before they have been realized at an articulatory level, i.e. been spoken out.

The central part of Postma & Kolk's model is that disfluencies are side effects of covert repair processes, i.e., errors that were not repaired. Monitoring as such is noticed at the surface through disfluencies, which might be regarded as some kind of disturbance. One of the advantages of Postma & Kolk's model is that it is applicable both to stuttered speech and normal speech.

Postma & Kolk divide the repair process into three separate steps:

1. Error detection.

2. Interruption.

3. Correction.

These steps have been discussed extensively in the literature by different researchers, and will be described in some detail in the following paragraphs.

2.6.4.1 Error detection

The basic question here is exactly *how* error detection occurs. One suggestion has been that outdata is compared to some kind of "norm". Discrepancies between this norm and outdata results in error signaling. Donald G. MacKay (1987) has objected that if there exists an immaculate representation within the system, then the question is why this representation is not used directly, instead of using it for comparison with an error-prone representation.

Another suggestion is that the system includes a set of rules, against which outdata are compared (Baars, Motley & MacKay, 1975). Candidate utterances are controlled according to syntactic, morphological, phonological (and so on) rules, and these context-sensitive rules signal aberrations, whereupon a certain candidate utterance can be prevented from traveling further.

As was mentioned above, Levelt (1983a, 1983b, 1989) believed that monitoring occurs with the speech understanding module, or system. A sequitur here is of course that all errors one can detect in one's own speech also can be detected in the speech of others. A problem with this model is that it can be hard to explain some very fast repairs that occasionally occur.

Norman (1981a, 1981b) proposed that the level at which monitoring occurs cannot be too far away from the level where the message resides. For instance, it would be hard to detect an over-all meaning error (which is often corrected) by monitoring phonological features such as [±voice] or similar.[1] Norman did not believe monitoring is carried out by a general module or process (i.e. Levelt's opinion), but rather with a multitude of different, specialized monitors, a view that is supported by modern evolutionary psychology (e.g. Cosmides et al., in press; Cosmides, Tooby & Barkow, 1992, Fodor, 1983).[2] Norman's model is also very similar to connectionist models of monitoring, like Dell's (1986) **spreading activation theory**.

Lackner & Tuller (1979) observed that subjects are able to detect self-produced speech errors much faster than errors in speech they simply listened to, and concluded that they must use an inner monitor, either a proprioceptive (feedback from muscles) or efference-copy monitor (i.e., monitoring the commands sent out to the muscles). Furthermore, since reaction times occasionally were as short as 0–100 ms, they concluded that an **efference-copy** monitor must be used, since proprioceptive feedback cannot be that fast.

2.6.4.2 Lexical retrieval

Another way to infer how the speech production chain works is to study the process of lexical retrieval. Kohn et al. (1987) gave subjects word definitions and asked them to say aloud all the words that popped into mind as they searched for the target word, as defined to them. It proved that phonologically related items or fragments were much better predictors of correct target word retrieval than were semantically related items. They conclude that successful retrieval is dependent on whether or not some lexical root information is available when the subjects initiate the search.

2.6.4.3 Interruption upon detection

So what happens when the monitor detects an error? According to Levelt, production is immediately stopped, something that has been criticized by e.g. Blackmer & Mitton (1991) and Nooteboom (1980), who claimed that there seemingly are tendencies to postpone the stop, at least to some extent. For instance, production stops seem to respect certain linguistic borders such as the integrity of constituents.

2.6.4.4 Repair

The third step is the repair of the error. Kolk (1991) suggested that some kind of trial-and-error strategy is employed, i.e. to run the program again and see whether it works better this time around. This model is supported by connectionist models such as Dell (1986) and MacKay (1987) insofar as the right nodes are activated which make them good candidates for recall.

---

[1] This possibly occurs, if marginally so, *vid.* Fromkin (1971/1973) and Shattuck-Hufnagel & Klatt (1980).
[2] A good primer to evolutionary psychology is available at **http://www.psych.ucsb.edu/research/cep/primer.html**

Others, e.g. Levelt (1983a, 1983b, 1989) and Van Wijk & Kempen (1987) believe that the speaker contributes more actively, based on the observation that repairs are almost without exception well-formed. Repairs include backtracking to a previous and appropriate point from where a new attempt is made, and subsequently and successfully executed.

2.6.5 Spreading-activation theory

An alternative view of speech production, also rooted in what speech errors look like, is **spreading-activation theory** (Dell, 1986). Dell (1984) found that speech errors such as phoneme exchanges where the misplaced phonemes are not adjacent—such as *heft lemisphere* instead of the intended *left hemisphere*—are difficult to incorporate in serial-order models of speech production. Instead, a hierarchical (connectionist) network model, as proposed by e.g. Dell & Reich (1975, 1980), where different nodes represent semantic features, words, morphemes, syllables, rhyme phonemes and so on, with two-way connections between all nodes, and where processing occurs by spreading activation, is consistent with observed speech errors.

2.6.6 Rapp & Goldrick: an evaluation of speech production models

One requirement that a given speech production model must be able to meet is, of course, that it should be able to explain documented speech errors. Another requirement is that a model must be able to *generate* speech errors typical of spontaneous speech, while at the same time not generate speech errors that do not occur in spontaneous speech.

Far from all models have been formal enough to allow testing, but several such models exist, and an evaluation of five such models are found in Rapp & Goldrick (2000). The models under scrutiny vary on a scale ranging from a high degree of discreteness to a high degree of interactivity, the latter term referring to multiple processes and their ability to influence one another during the execution of the process. Interactivity comes in different flavors:

Forward-backward interactivity means that later processes receive data from earlier processes, while being able to feedback information backwards in the chain. Forward-backward interactivity has been used to explain phenomena like the "word superiority effect", i.e., that people recognize letters that are part of a word much faster than they recognize single letters, when stimuli are flashed to the subjects.

Lateral interactivity refers to a process where different stages are not ordered in relation to one another.

Integration refers to a model where parallel processed that are not ordered relative to each other merge into a single process at a later stage, which occurs later than the previous, parallel, processes on a time-line.

2.6.7 Dennett: the "Pandemonium" or "Multiple Drafts" Model

So, what do all these models tell us about human speech production and the potential role disfluencies play in revealing anything about deeper linguistic processes? The first thing to be pointed out is that most models trying to explain speech production "from the beginning" *actually* seem to start from step *two* in the process, rather than from the very beginning.

Dennett (1991), in criticizing Levelt's model, focused on the description and role of Levelt's Conceptualizer. The Conceptualizer decides what it wants to say, whereupon this message is forwarded to the Formulator, which in turn initiates the entire modular process that eventually leads the speech sounds being produced. The problem with this model, as Dennett views it, is that nothing is said about what kind of representation the Conceptualizer "speaks". If it speaks English (or Swedish, or any other human language), then all the work is already done, and the rest will only be decoration or ornamentation at a rather detailed level. If, on the other hand, the Conceptualizer employs some kind of "mentalese", or other "language" designed for speech acts (but probably not for other motor acts), then the Conceptualizer will first have to translate it into English (or other human language), which obviously makes the work harder on the Formulator, but still has not explained anything about the beginning of the process. It is still *translation*, and where or how the entire thing begins remains unanswered. How does the Conceptualizer find what "words" it should send to the Formulator? Is it not the case that there must exist a Levelt-like model *inside* the Conceptualizer, too? Instead of putting the problem of how speech is done inside the brain, the question is how speech is done inside the Conceptualizer, leading to a prototypical infinite regress problem. Granted, Dennett points out that Levelt himself acknowledges that the Conceptualizer "is a reification in need of further explanation" (Dennett, 1991, p. 233, citing Levelt, 1989, p. 9).

Another critical point Dennett raises is that Levelt borrows too much from von Neumann machines, which according to Dennett is not supported psychologically. Human consciousness does not, in most essential ways, function like a serial von Neumann machine. According to Dennett, the basic similarity between a general von Neumann machine and Levelt's is, broadly speaking, that there at all stages of the process exists a specifically coded sequence whose accumulated content is passed on to the next module for processing, that is to say (1) all processes work on already established contents, and (2) the "bureaucracy" proper must be carefully designed, all decision-making must be specified in excruciating detail, and all agents must be aware of exactly what tasks they are allowed to take care of. While Dennett buys premise number (1), that there somewhere is some kind of thought waiting to be dressed in words, he protests against premise number (2), that a hierarchical structure slavishly will dress exactly *that* thought in words, according to a von Neumann machine-like architecture. What is lacking in Levelt's model, according to Dennett, is a clarification of what the creative and judging role of the Conceptualizer is. Either everything resides inside the Conceptualizer, which just sends out an order to the Formulator, or the Formulator does all the work, basically[1].

So then, what would an alternative model look like? While Dennett explicitly states that his own proposed model is something of a caricature, he proposes something he labels "a pandemonium of demons", clearly a reaction to the highly modular stance taken by Levelt. So, Dennett asks us to imagine an argument taking place between two people, where our imagined speaker has just been insulted and wants to fire away a good retort. How does he do that? Instead of a finished thought or concept inside a Conceptualizer, Dennett suggests that, for no specific reason, the horn just goes off, thus:[2]

Beeeeeeeeeeeeeeeeeeeeeeeeeeep…

[1] For further discussion on this problem, see Fodor, Bever & Garret (1974, pp. 373–384).
[2] The following paragraphs are derived from Dennett (1991), pp. 235–240.

The reason this occurs is simply because there is no reason to prevent the horn from going off. The horn excites a number of "demons", that start working on the signal to make it more structured, resulting in a slightly more structured signal:

Yabba-dabba-doo-fiddledy-dee-tiddly-pom-fi-fi-fum…

Of course, this is nonsense, but nonsense in English. This, in turn, is further processed by another host of demons, resulting in:

And so, how about that? Baseball, don't you know, in point of fact, strawberries, happenstance? That's the ticket. Well, then…

Demons then work on that, creating a set of **multiple drafts** of the utterance, so that some demons come up with a draft like:

You big meany

… while other demons have created:

Read any good books lately?

… but the "winning" candidate is:

Your feet are too big!

Granted, this is perhaps not the snappiest thing our speaker could have said, and he will surely chew on all the other possible, much more snappy things he *could* have said, had he been in more control of the process. What is obvious from Dennett's model is its focus on parallelism, rather than the highly modular, feed-forward-only, characteristics of Levelt.[1]

The **Multiple Drafts Model** is further developed (in a less caricature-like way) in a target article by Dennett & Kinsbourne (1992a, 1992b). While speech production is explicitly discussed in Dennett (1991), as an alternative to Levelt's model), the emphasis in Dennett & Kinsbourne (1992a/1992b) lies more on presenting the Multiple Drafts model as an alternative to the prevailing view that there is a "Cartesian Theater", where all sensory input comes together. Although Dennett & Kinsbourne (1992a/1992b) do include speech in their presentation/discussion, both from a perception point of view (p. 188) and a production point of view (p. 190), and although a couple of the critics include speech in their discussion (Block, 1992; Warren, 1992), this fuller presentation of the Multiple Drafts model does not shed much more light on speech production proper. However, Young (1992) is of the opinion that the so-called **McGurk effect**[2] (McGurk & MacDonald, 1976), lends some support to the Multiple Drafts model, as opposed to more modular and serial models.

[1] For a recent review of feed-forward models and their viability, see Miall & Wolpert (1996).

[2] The McGurk effect occurs when subjects are presented with mismatching auditory and visual phoneme information, which results in a blending effect. Thus, when watching a video of a person mouthing [ga], while at the same time listening to a person saying [ba], most people *hear* the fusioned sound sequence [da], despite the fact that [da] does not exist in either the visual or auditory channel. This occurs even when the subjects *know* what the film and soundtracks consist of (McGurk & MacDonald, 1976). Young (1992) pointed out that this effect is entirely consistent with a Multiple Drafts model.

So, what do we say about this model? First of all, the obvious difference between Levelt's and Dennett's models, respectively, is that while the former is highly modular, the latter is non-modular to the extreme. In Dennett's model a huge number of demons work in parallel on a signal, resulting in an extremely large number of drafts, where most never surface, as something that gets spoken out, even as alternatives the speaker ever becomes aware of. Most drafts simply die before they get a chance of reaching any conscious level in the speaker's mind. This, Dennett points out, is psychologically grounded, in that we, as speakers, quite often do not "write" things in our brain, edit them and then speak them out, or in Dennett's own words:

> In the normal case, the speaker gets no preview; he and his audience learn what the speaker's utterance is at the same time. (Dennett, 1991, p. 238.)

Turning on his own model, Dennett then points out that one needs to explain how this tournament is staged, how discrimination occurs between the huge numbers of alternative utterances produced so that the final outcome is something that reflects the communicative intent of the speaker, even if there is no Central Meaner. Dennett's model needs restrictions.

So, if the preverbal message is conceived in some kind of mentalese, then most of the work, if not all, is done *before* Levelt's model even comes into play, while Dennett's demons need some kind of instructions, which Dennett acknowledges is lacking from his proposal. Perhaps, Dennett suggests, does intent emerge from a number of intelligent questions. Speech production would then be some kind of quasi-evolutionary process where word-demons pose questions like "can we say this?", that are then answered by a set of content-demons. Communicative intentions would then emerge from a speech-act-like that runs both in parallel and serial order, exploiting a huge number of subsystems that are more or less capable of dressing the desired speech act into words.

So, is this at all possible? Levelt based his model on actual speech data, and tried to conjecture what deeper structure might have caused the observable speech. Is there any such evidence in favor of Dennett's model? Dennett thinks so. He points out that a large number of *constraint satisfaction* models have been proposed that work in favor of a Pandemonium model (whose greatest problem is the limited power of the demons), e.g. several connectionist models, like e.g. Rumelhart & McClelland (1986), and more abstract alternatives, such as Hofstadter (1983). Dennett also thinks that Minsky's (1985) description of agents that create a "society of mind" is on the same track. What these models have in common is that they are simulations of (human) behavior, and Dennett confesses that all these models need to be more elaborate to conclusively confirm a Pandemonium model of speech production.

To account for such a different model as the Pandemonium model in only a couple of pages, as well as discuss its empirical foundation or possible consequences is clearly not possible, but it is still interesting to point out that modular approaches—so often taken for granted—are in no way the final word. Dennett's proposal might have some trait of caricature, as he points out himself, but it is easy to be misled by work that tries to pinpoint exactly between which modules a certain disfluency occurs so that one thinks that these postulated modules are proven to exist. This, of course, is not the case, which Dennett illustrates. Any successful model of speech production, Dennett points out, must describe some kind of evolutionary process of message generation, lest we get caught up in the kind of infinite regression that would be the case with a Conceptualizer inside the Conceptualizer (and so on).

2.6.8 Consciousness, brain potentials, free will

A major question when discussing repairs, monitoring, backtracking and so on, is of course whether or not the processes discussed are *conscious decisions* made by the speaker or whether they are *automatic*. It comes as no surprise that Levelt (1989), who placed the monitor inside the Conceptualizer, considers the repair processes more or less conscious. Since the Conceptualizer controls different stages in the production process, error detection is to a certain degree conscious. One of Levelt's proposed functions is the "main interruption rule", which states that speech (inner and outer) is constantly monitored, and as soon as an error is detected, speech is halted. The reason for assuming such a rule, according to Levelt, is that cut-offs are not linguistically motivated, i.e., words or even syllables can be interrupted. This rule has received criticism in the literature. Berg (1986b, see also Berg, 1986a, 1992) is of the opinion that there can be no such thing as the main interruption rule. Also, Laver (1969/1973) and Nooteboom (1980) believe that self-repairs are more or less automatic, subconscious processes. As should be obvious from the discussion above, this is where it gets complicated, since the **time factor** now begins to play an important role.

What the previous models—and indeed all similar models—have in common is that they presume that some kind of **motor action** is the final stage of the processes they aim to describe. Motor actions are initiated in the brain, executed by motor processed and are being monitored—using inner and outer loops—and reacted to upon detection of error. Haggard (2001) summarizes the issue of "the psychology of action" thus:

> Actions are part of the way that the mind controls the body. Two fundamental psychological questions about actions are 'Where do they come from?' an 'How does the mind produce them?' These may be called the 'internal generation problem' and the information expansion problem, respectively. (Haggard, 2001, p. 113.)

This goes for all kinds of motor activity, where speech, from that point of view, is but one example. Initiation, execution, monitoring and correction are all processes that take time, so there is an inherent time factor to consider here, especially when (re)action times are small. In speech production—a motor activity—detection-and-repair processes (just to take an example) sometimes occur at phrase level, but often at much lower levels, which implies that actions such as monitoring and execution as discussed earlier must also occur very close to real-time. So, there are two issues that need to be addressed here:

The first issue that is most often *not* included in the speech production models described above is a general perspective of human reaction times. How fast can we react to stimuli, and initiate some kind of motor response?[1] Also, given that some (but not all) of the models assume conscious monitoring, how fast we can react *consciously* to stimuli, be they external (auditory feedback loop), or internal (brain-internal monitoring)?

The second issue, however, is of extreme interest. Quite often when the terms *monitoring*, *detection* and so on are used, it is not always explicitly mentioned whether this occurs *consciously*, *subconsciously*, or *preconsciously*. And this is where it gets difficult to draw too far-reaching conclusions as to the inner workings of the speech production process, since

[1] It has been shown that human subjects react faster to acoustic or tactile stimuli of moderate intensity faster than they do to visual stimuli, on average 140 ms versus 180 ms, respectively (Elliot, 1968). Also, imitation reaction times are faster than simple visual reaction times, consistent with a direct matching circuit in the CNS (Tessari, Rumiati & Haggard, 2002). Seminal work on voluntary movement and reaction times was done by Woodworth (1900).

conscious will and motor action have been shown to interrelate in bizarre ways, which will be devoted some space on the following pages.

To sum up, when discussing motor action, reaction time, monitoring, inner and outer loops, different kinds of "awarenesses" of internal of external stimuli, I would like to agree with Lord Brain (1963; quote in "Tuning in…) that there is no way to waive the physiology of how the brain works. Or, as Brain (1963) puts it:

> [W]e must now examine what is implied by the statement, freely used by psychophysiologists, that the pattern of nerve impulses is conveying information, and that this information is conveyed in the form of a code. (Brain, 1963, p. 389.)

Also, before Brain, Young (1962) mentioned speaking and writing as examples of the code that needs to be stored in memory for interpretation. Consequently, the following sections will delve into the electro-physical world of the brain in general, and its relation to speech production in particular.

2.6.8.1 Endogenous action: readiness potentials ("Bereitschaftspotential")

In a now classic paper, Kornhuber & Deecke (1965) showed that willed, spontaneous, action—in this case the bending of a finger—was preceded, in the order of several hundred milliseconds, by a **readiness potential** (RP, or "**Bereitschaftspotential**", BP) in the brain in the order of several hundred milliseconds. Or, succinctly put, several hundred milliseconds *before* the conscious decision to bend the finger was made, the brain started to prepare the finger muscles, and only *after* this brain activity took place was the conscious decision made to tap the finger.[1] Thus, the brain seems to "know" beforehand what the conscious agent was intending to do, or think. The Kornhuber & Deecke results have since been repeated by other researchers—mainly by Benjamin Libet and colleagues (Libet et al, 1983; Libet, 1985a/1985b, 1987, 1990, 1991a, 1991b, 1992a, 1993, 1999, 2002), but also by Lüder Deecke and colleagues (Deecke, 1987a, 1987b; Deecke, Weinberg & Brickett, 1982; Deecke, Sheid & Kornhuber, 1969; Deecke et al., 1983, 1984; Deecke, Grözinger & Kornhuber, 1976), as well as others (see e.g. Vaughan, Costa & Ritter, 1968; Fourneret & Jeannerod, 1998; Frith, 2002; Frith, Blakemore & Wolpert, 2000; Haggard, 2001; Haggard & Eimer, 1999; Keller & Heckhausen, 1990; Haggard & Eimer, 1999; Johnson & Haggard, 1999, Blakemore & Frith, 2003; to mention but a few). Libet et al. (1983) wanted to distinguish between the physical time of the first noticeable electrical activity in the brain, the subject's reported time of awareness of the intention to move, and the first recorded electrical activity in the muscle (EMG). This was achieved by letting the subjects monitor their own actions on a clock—a cathode ray oscilloscope (CRO), see Libet et al. (1983) and Libet (1983, 1993 and 1999, the latter with a figure showing the clock)—and report the hand position of the clock when they became aware of the intention to act. They granted that subjective reports are not-as-reliable-as-one-would-wish measures, so they also used a skin stimulus as a control factor. I will—*in medias res* fashion—jump straight to a summary of the findings as reported by Libet and others (given where appropriate), given in **Figure 2.2**.

[1] Although finger movements have been most widely studied (with corroborating results), other movements, such as hip, knee, leg and toe movements have also been studied, with slightly different figures, and also with different lateralization of recorded brain activity. See Brunia (1980), Boschert & Deecke (1986), Deecke et al. (1983), Boschert, Hink & Deecke (1983).

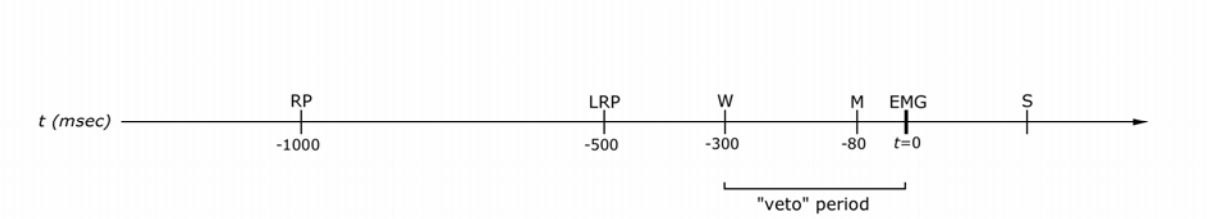


Figure 2.2. The "timeline of the brain".[1] **RP** (Readiness Potential, "Bereitschaftspotential"/BP), i.e., the first observable activity in the brain. **LRP**(Lateralized Readiness Potential), i.e., the electric signal in the hemisphere opposite of the intended movement (i.e., for a left finger movement the recorded activity in the right hemisphere). **W** The willed/conscious decision to move. **M** The reported time of awareness of muscle movement. **EMG** Actual body movement, electric activity in the muscles. **S** Contingent stimulus, e.g. a click.

RP precedes the conscious decision **W** to bend the finger with between 500 and 1000 ms, that is to say that the *unconscious* brain activity (the readiness potential) precedes *conscious* brain activity by up to a second (Kornhuber & Deecke, 1965; Libet, 1983, 1985a/1985b; Deecke, Scheid & Kornhuber, 1969; Deecke, Grözinger & Kornhuber, 1976; Libet et al. 1983; Deecke, 1987b; Haggard, Newman & Magno, 1999; Keller & Heckhausen, 1990). RP can further be divided into **RP1**, and **RP2**, depending on whether or not some pre-planning of the movement took place. RP1, involving some pre-planning, typically appears at 800 ms before **W** (Libet et al., 1983; Libet, 1985a/1985b), while **RP2**, being completely spontaneous, typically appears at around 500 ms prior to **W** (Libet et al., 1983, Libet, 1985a/1985b). Moreover, Deecke, Grözinger & Kornhuber (1976) also noted that "[p]receding speech production the BP [Bereitschaftspotential, i.e. RP] shows early side differences between the hemispheres" (Deecke, Grözinger & Kornhuber, 1976, p. 111).

LRP appears around 300 ms after **RP**, but still precedes the conscious decision **W** to bend the finger with around 200 ms. It has been shown that **LRP** and **W** co-vary statistically, and are thus tightly bound together, which has been taken as evidence that the recorded unconscious activity is movement specific, rather than general (Haggard & Libet, 1999; Haggard & Eimer, 1999; Eimer, 1998; Kutas & Donchin, 1980).

W is the time reported by the subjects of awareness of intention to move. Libet et al. (1983) pointed out that there seems to be a time frame of around 300 ms (or even 700 ms according to later experiments) *after* **W**, but *before* actual motor action (**EMG**), during which a conscious decision can be made to stop the initiated finger movement, something they refer to as a **veto function** of the conscious mind (e.g.. Libet et al., 1983; Libet, 1985a, 1991, 1999). This would mean that even if the decision to move is unconscious, consciousness occurs before actual movement is executed, leaving some time for a conscious decision not to move, i.e., leaving room for a decision to prevent the intended movement from being executed, or a *free won't*, as it has also been referred to (e.g. Haggard, 1999; Claxton, 1999). The notion that unconsciously initiated acts could be stopped during the short period between conscious awareness of the oncoming movement and the actual movement was experimentally tested in

[1] Note that the times given are approximate since they differ between different sources. Deecke (1987b) mentions that "BP starts as early as 1–2s or more before the onset of movement" (Deecke, 1987b, p. 233). Vaughan, Costa & Ritter (1968) mention "as much as 2 sec" (Vaughan, Costa & Ritter, 1968, p. 1). Frith mentions "up to 1 s" (Frith, 2002, p. 484), while Haggard & Eimer (1999) mention "at least 700 ms before EMG onset" for RP and 296 ms "on average" for W (Haggard & Eimer, 1999, p. 128). Frith, Blakemore & Wolpert (2000) puts M "50–80 ms" before EMG (Frith, Blakemore & Wolpert, 2000, p. 1776), as do Blakemore & Frith (2003, p. 219), who also put LRP "around 500 ms before the movement" (ibid., p. 220). Becker et al. (1972) noted a BP of about –1 second for eye saccades. The timeline basically provides the reported *order* between the events, even if exact times and proportions vary slightly in the literature.

that subjects were instructed to move spontaneously, but then stop actual action at around 100 ms prior to movement (using the clock hands). The results indicated that the veto function is indeed a real phenomenon and the so-called **M-veto** is shown in Libet (1985a, p. 537).

M is the reported awareness of muscle movement. Note that this precedes actual muscle activity, a fact that has been referred to when ruling out sensory feedback monitoring of motor activity. Frith, Blakemore & Wolpert (2000) summarized previous findings, pointing out that "[t]hese observations imply that our awareness of initiating a movement is not derived from sensory signals arising in the moving limb. This information will not be available until after the limb has started moving. In terms of the model of motor control we are formulating here, the most likely representation relating to awareness of movement initiation is the predicted state of the system" (Frith, Blakemore & Wolpert, 2000, p. 1776). Blakemore & Frith (2003) point out that "our awareness of initiating a movement is not derived from sensory signals arising in the moving limb — because such signals are not available until after the limb has started moving. Instead our awareness appears to be linked, at least in part, to a signal that precedes the movement." (Blakemore & Frith, 2003, p. 220). In a similar vein, Haggard, Newman & Magno (1999) argue that "it appears to rule out the possibility that our knowledge of movement is based on peripheral feedback from the moving limb. Had that been the case, the perceived time of movement should be delayed relative to the actual movement onset, because of the neural conduction time for sensory information from the moving limb to the brain centers making the judgment" (Haggard, Newman & Magno, 1999, p. 292). Frith (2002) reached a similar conclusion: "Direct sensory feedback arrives too late to be useful for movement guidance" (Frith, 2002, p. 483). While not completely ruling out Levelt's outer (auditory) loop, the reported results seem to rule out proprioceptive or tactile feedback in speech production, at least so long as speech production is similar to finger tapping.[1]

EMG is the moment muscle activity starts as evidence of recorded activity in the muscle. For an early study, showing a larger positive deflection in the contralateral (to the activated limb) hemisphere, see Gilden, Vaughan Jr. & Costa (1966).

S An external stimulus used in some experiments to be described below. It has been shown that **EMG** must precede **S** by 40–360 ms to be judged coincident with **S** (McCloskey et al., 2003). It has also been shown that if **EMG** is *voluntary* (i.e., the subjects made the decision to move themselves), then **M** and **S** are "moved together" by the brain (which seems to be "cheating with the clock" again), so that a later **M** is reported (on average by 26 ms), and **S** is reported to have occurred earlier than it actually did (by 9 ms on average). If **EMG** is *involuntary*—using transmagnetic cranial stimulation[2]—then an earlier **M** is reported (by 9

[1] For further discussion, *vid.* Haggard, Newman & Magno (1999), who discuss possible explanations for the observed timing differences, including the **prior entry phenomenon** (Sternberg & Knoll, 1973), the fact that events attended to are detected earlier than events unattended to, Sternberg et al.'s (1978) motor programming model as well as the notion of **P-centers** (Morton, Marcus & Frankish, 1976; Aschersleben & Prinz, 1995). In discussing the latter, Haggard, Newman & Magno (1999) mention that "The concept of a P-centre has not previously been applied to sequences of several movements such as typing strokes. However, this case should be comparable with spoken words, which are essentially a sequence of several vocal tract articulations." (Haggard, Newman & Magno, 1999, p. 302). Also, Vos, Mates & van Kryusbergen (1995) argue that subjects tapping in synchrony with a metronome use P-centers rather than physical onsets as the synchronization cue. Penrose (1989), referring to neurosurgeon Chester Penfield, suggests that "the *desire* for movement might have more to do with the thalamus than the cerebral cortex" (Penrose, 1989, p. 493; italics in original).

[2] It has been argued that the few electrodes employed by e.g. Libet are not exact enough, which has led other to use transcranial magnetic stimulation, instead of and as well. Most such studies have replicated the results by Libet and others. See for example Deecke, Weinberg & Brickett (1982), Deecke, Boschert, Weinberg & Brickett (1983) and Brasil-Neto et al. (1992).

ms on average), and a later **S** is reported (by 15 ms on average) (Frith, 2002; Haggard, Clark & Kalogeras, 2002; Tsakiris & Haggard, 2003; Blakemore & Frith, 2003; Haggard & Magno, 1999). This has been taken as evidence that the brain cheats the internal clock to bind together voluntary internal events with contingent external events so as to reinforce causal binding, while involuntary internal events are dissociated with contingent external events. These observations were all done on endogenous, voluntary acts, for "freely willed" actions, which could roughly correspond to inner loops in that they all deal with how the brain monitors its own actions, or stimuli, if you will. But what about external stimuli? It suggests that the intricate timing relationships of external stimuli have also been studied, and the results will be summarized in the following section.

2.6.8.2 Peripheral stimuli: backward referral (or antedating)

In the mid-60s, Libet and colleagues conducted a number of experiments where conscious sensory experiences were elicited by electrical stimuli applied directly to the surface of the somatosensory cortex in awake subjects (e.g. Libet, 1965, 1966; Libet et al., 1964, 1967, 1979, 1991, 1992). They found that the pulse repetitions needed to go on for "the surprisingly long period of about 0.5 seconds or more" (Libet, 1965, p. 82) to effectively elicit a conscious sensory experience. Libet also observed that stimuli applied directly to the skin were effective with much shorter pulse trains, and needed only a few pulses to elicit conscious experience.[1] Libet summarized his observations thus:

> These findings lead me to formulate a general *hypothesis that a minimum period of suitable cortical activation, lasting 0.5–1 sec., is a necessary feature of any such activation* (at least when it is close to liminal level) *for eliciting any conscious experience. The corollary of this would be that shorter periods of such cortical activation may still elicit unconscious experiences*. (Libet, 1965, p. 83; italics in original.)

Libet continues:

> The hypothesis also provides some understanding of how it is that, in complex, integrative and creative thinking, the play, interaction, and juxtaposition of mental events are often carried through unconsciously /… / The requirement of relatively long duration, of some cortical activation at the near liminal level, for the onset of each increment in conscious experience obviously would impose a certain ponderousness on the thinking process. In contrast, if only short durations of liminal activities are needed in unconscious experiences, they would provide the kind of quick-acting only marginally intense nature that would facilitate complex interactions, rearrangements, and integrations of the type demanded. If a rather long period of activation, e.g. 0.5–1 sec., is a requirement for conscious experiences at near liminal levels, this would constitute a "latent period" between the onset of activation and the "appearance" of the conscious experience. This would mean that one is not actually aware of a sensory stimulus (at least of a near-liminal one) for a period as long as 0.5 sec. or so after its occurrence. /…/ A lag of conscious experiences behind the initiating events, which can be an order of magnitude greater than the delays involved in sensory and motor pathways, introduces a viewpoint about awareness which can have important psychological and philosophical implications. (Libet, 1965, p. 84.)

[1] Peripheral sensory input needs not be strong at all to generate a subjective sensory experience. Hensel & Boman (1960) exposed skin nerves in a human subject and monitored the electrical responses in the remaing nerve fiber to mechanical stimulation. They found that the weakest stimulus that was detected subjectively gave rise to one single impulse in the only remaining nerve.

Libet then points out that we all know that we can react to sensory stimuli extremely quickly (even as short as 5 ms), even if some decision-making is involved, and that such quick reactions must take place before conscious awareness of the action is experienced. It should be obvious by now that this has implications on the notions of free will and "choices", from the general perspective, but also from our, narrower, perspective of how consciously the brain reacts to the produced speech, and how that affects the proposed speech production models. The observation that there is a delay (of 500 to 1000 ms) before conscious somatosensory experience is reported prompted Libet and colleagues to test whether there is also a subjective delay in the conscious experiences of peripheral sensory stimuli. Or, as they formulated their question, "is there a delay in the subjective timing of the experience that would correspond to the presumed delay in achieving the neuronal state that 'produces' the experience?" (Libet et al., 1979, p. 193). This question above entailed two specific postulates, *viz.*, the existence of a subjective referral of the timing of sensory experience, and a role for a specific projection system in mediating the said subjective referral of timing. Libet et al. acknowledged the problem of determining the timing of subjective experience (and emphasized that it must be distinguished from behavioural responses in general, since these may be unconscious), and adopted a method where the subject reported the subjective timing order of two separate sensory experiences, the test stimulus and a reference stimulus. To keep things short, their study revealed some astonishing phenomena. Quoting:

> (1) Some neuronal process associated with the early or *primary evoked response*, of SI (somatosensory) cortex to a skin stimulus, *is postulated to serve as a 'time-marker*'. (2) There is an automatic *subjective referral of the conscious experience backwards in time* to this time marker, after the delayed neuronal adequacy at cerebral levels has been achieved" (Libet et al., 1979, pp. 201–202; italics in original.)

The terms **retroactive referral** and **antedating** of the subjective experience are also used to describe the observation (Libet et al., 1979, p. 201). Or, as they put it, "[t]he sensory experience would be 'antedated' from the actual delayed time at which the neuronal state becomes adequate to elicit it; and the experience would appear subjectively to occur with no significant delay" (Libet et al., 1979, p. 202). Thus, they concluded that subjective timing of a sensory stimulus is in fact retroactively antedated back to the time of the primary cortical response. Or, in other words, one does not become conscious when it happens, only later, but that later awareness is projected backwards in time so as to make the subject experience "immediacy" in the sensory experience. Thus, it is as if the brain plays tricks with the mental clock so as to make our conscious decisions with the physical reality.[1]

[1] Related research has shown that the brain is fairly creative, sometimes in retroactive ways, with other types of external stimuli. Geldard & Sherrick (1972) showed that mechanical impulses to the arm created a phenomenon dubbed the **cutaneous rabbit**, which I will let them explain themselves: "[I]f five brief pulses (2-msec duration each, separated by 40 to 80 msec) are delivered to one locus just proximal to the wrist, and then, without a break in the regularity of the train, five more are given at a locus 10 cm centrad, and then another five are added at a point 10 cm proximal to the second and near the elbow, the successive taps will not be felt at three loci only. They will seem to be distributed, with more or less uniform spacing, from the region of the first contactor to that of the third. There is a smooth progression of jumps up the arm, as if a tiny rabbit were hopping from wrist to elbow. /… / hopping can go down the arm as well as up it. Indeed, it is possible to have hopping in both directions at once." (Geldard & Sherrick, 1972, p. 178). The obvious problem with this phenomenon is how the brain, when experiencing the first five pulses can make them go up the arm, *before* the next five pulses are even administered. The only solution seems to be that the entire train of pulses is interpreted, and experienced, in a way that is at least partly retroactive, i.e., experienced timing referred backward. Such "apparent motion" has also been shown for vision by Kolers & von Grünau (1976) in what is known as the **color phi phenomenon** experiment, or for haptic experience by Sherrick & Rogers (1966). Similar extensions backwards in time have also been shown for saccadic eye movements (Yarrow et al., 2001).

So, to summarize the findings, let's cite Libet (1981):

> (1) *There is a substantial delay before cerebral activities*, initiated by a sensory stimulus, *achieve "neuronal adequacy"* for eliciting any resulting conscious sensory experience. /…/ (2) After neuronal adequacy is achieved, the *subjective timing* of the experience is *(automatically) referred backwards in time*, utilizing a "timing signal" in the form of the initial response of cerebral cortex to the sensory stimulus. (Libet, 1981, p. 182; italics in original.)

The antedating of subjective experience is further investigated, elaborated and discussed in a number of articles by Libet and colleagues, e.g. Libet (1991a, 1991b, 1992a, 1992b, 2002; Libet et al., 1991, 1992).

The controversial implication of these results is that they seem to lead to dissociation between brain states and mental states, a conclusion that also yielded a lot of critical work.

Early out was Churchland (1981), who, mainly on methodological grounds, was of the opinion that the data used by Libet and colleagues were not sufficient to draw such far-reaching conclusions. Churchland (1981) criticizes Libet's results on methodological grounds, claiming that "[t]here are many ways of tricking one's nervous systems such that false perceptual judgments are made about the perceived world" (Churchland, 1981, p. 165), something which Libet (1981) replies to (in a fairly astute way), stressing the methodological relevance of his results. Libet also points out that sensory illusions are to be distinguished from backward referral.

From a more philosophical angle, Honderich (1984) claimed that Libet et al.'s findings pose problems to monist (identity) theories of the mind/brain relationship, something Libet (1985d) denies. The fact that his results dissociate between physical/mental timing and actual physical/neural timing does not contradict monist theories of the mind/brain.

To summarize this section, suffice it to say that the notion of backward referral has received substantial criticism, both from methodological and philosophical standpoints, and there is still an on-going debate concerning the relation between neural and mental timing of events. For a recent debate, see e.g. Banks (2002), Bolbecker et al. (2002), Breitmeyer (2002), Gomes (1998, 1999, 2002), Stanley Klein (2002a, 2002b), Pockett (2002a, 2002b), Rosenthal (2002), Trevena & Miller (2002) and Libet (2002).

2.6.8.3 Philosophical implications

It should be clear to the reader that the results described in the two previous sections have considerable implications for human action, both as to agency of that action (if initiated internally) and the perceived timing of that action (if applied externally). From a speech production point of view, the crucial points concern mainly perceived timing, but there are other implications, as well. Before delving into a more detailed discussion concerning the philosophical implications, let us try to summarize the main points of the research results described in the previous two sections.

- Motor action is initiated subconsciously, occurring up to 500–2000 (or more) ms before any conscious decision to move is reported. This shows up as a readiness potential (RP) in the brain activity. Early hemispheric differences have been observed for speech production (Deecke et al., 1976). Further, there is a time difference depending on whether or not pre-planning occurs, in that pre-planned RP occurs earlier than fully spontaneous RPs.

- A motion-specific, lateralized, readiness potential (LRP) appears later than RP, at around 500 ms before the onset of movement, and has been shown to co-vary with the reported conscious decision, W.

- There is a time window during which a conscious decision can be made to stop a motor action that has already been initiated by the brain, but which has not yet been executed, i.e. the subjects could change their mind about moving a finger. This has been labeled the "veto period" (e.g., Libet, 1991, p. 685)—or "free won't"—and has received support experimentally (Libet, 1985a).

- The brain has some kind of "antedating function" that refers subjective timing of an event back to the moment it occurred, although actual activity took place later.

- Besides the antedating phenomenon mentioned above, the brain seems to employ some kind of *intentional binding*, which also behaves differently as a function of whether or not the movement is self-induced (voluntary) or other-induced (involuntary).

It goes without saying that the implications of these findings go beyond the possible problems they pose to speech production models, and not only Libet but also others have discussed what these mean to our notion of "free will". If our actions "begin" subconsciously, and the brain then "fools itself" (or us) into believing that we "did it", then what constitutes a "conscious decision", or indeed "free will"?[1]

When it comes to the role of RP when "gauging" voluntary acts, of interest here is the observation by Obeso, Rothwell & Marsden (1981) that involuntary tics in patients with **Tourette's syndrome** were *not* preceded by a readiness potential. When the patients were asked to produce the same movements voluntarily, RPs occurred at about 500 ms prior to the movement. Hoffman & Kravitz (1987) pointed this out as a potential problem for Libet, while Libet (1987) pointed out that Tourette's patients exhibit involuntary movement (Libet, 1987, p. 784), and later that "actions by a person during a psychomotor epileptic seizure, or by one with Tourette's syndrome, etc., are not regarded as actions of free will." (Libet, 1999, p. 52).

Given the possible implications for (among other things) human self-image, it is not surprising that these results have faced stark criticism from a number of different angles, some of which will be briefly summarized in the following. Although part of the criticism raised is methodological, the underlying reasons are basically philosophical.

Vanderwolf (1985) was of the opinion that Libet's clock control cannot be used reliably since mental processes are not available to introspection. Libet (1985b) defends the method, pointing out that he differentiates between subjective experience and externally observable processes (RP), and that he regards the former as a primary unit, something which cannot be broken down into smaller pieces, which also renders it unanalyzable in other ways than what he is doing. The only way to know when a subject became "conscious" about something is to ask the subject possessing the consciousness when he became conscious of it.

Latto (1985), Marks (1985) and Ringo (1985) suggested that consciousness probably occurs gradually, and that there is a threshold that must be passed before the subject reaches awareness of it. This should considerably reduce the time difference between RP and W, since

[1] For a recent synoptic review and discussion, see Spence (1996). See also Libet (1999), Libet, Freeman & Sutherland (1999) and Haggard & Libet (2001).

consciousness should "be there" much earlier, just not be fully amenable to the memory until having exceeded a certain strength.

Breitmeyer (1985), Latto (1985), Rollman (1985), Stamm (1985), Underwood & Niemi (1985) and Wasserman (1985) all point out that the clock reports/timing may be erroneous from a number of different reasons, including "delays" of visual stimuli and so on. Libet responds to this that such phenomena in no way affect primary phenomena like subjective consciousness, which is the focus of his studies. Or as he puts it:

> One should not confuse *what* is reported by the subject with *when* he may become introspectively aware of what he is reporting. /…/ This can explain, for example, why a runner in a race can take off within 50–100 ms after the starting gun, presumably well before he becomes introspectively aware of the stimulus, but later reports that he heard the gun *before* taking off. (Libet, 1985b, p. 559.)

Scheerer (1985) suggested that the veto experiment carried out by Libet et al. (1983) was equivalent to a simple visual reaction time test, to which Libet answers that their subjects knew beforehand that they were to react to a stimulus at a given moment in time, which made their experiment different from a simple reaction time paradigm.

Jasper (1985), pointed out that conscious action may take place without being accessible to memory at a later stage. Libet (1985b) responded that this is testable, and that he has taken great care to eliminate that particular confounding risk factor.

Näätänen (1985) questioned the notion of "spontaneous", while Eccles (1985) and Rugg (1985) are of the opinion that averaged RP values can mask fluctuations, something Libet points out has been taken into account.

Merikle & Cheesman (1985) pointed out that one should try to demonstrate the same phenomenon for subconscious actions while Van Gulick (1985) voiced the opinion that one should not regard RP as a subconscious state, but rather as a "conscious" state, but which does not become "self-conscious" until a later stage. Libet rebuts this as a mere play of words.

Mortensen (1985) agreed with Libet that the "veto" process is a plausible explanation as to why we do not do certain things. Even if actions are initiated subconsciously, we can still prevent them from taking place, given the time window mentioned above. However, the veto function has been questioned by several other researchers, including Danto (1985), Doty (1985), Latto (1985), Nelson (1985), Rugg (1985), Underwood & Niemi (1985) and Wood (1985). If conscious actions, like finger movements, are initiated subconsciously, why should not the veto, too, have its RP earlier? Libet acknowledges this as valid criticism, but points out that his research in no way excludes the possibility of conscious actions that lack an earlier RP.

Deecke (1987a) wrote that:

> Although the BP [RP] is widespread and can be recorded over both hemispheres /… / two principal generators seem to prevail. These are the supplementary motor area (SMA), which generates the early symmetrical component, and the rolandic motor cortex (MI), which generates the late asymmetric (i.e., contralateral) component preceding finger movement. /… / Topographical recordings are needed to distinguish between the two components; the few electrodes used by Libet are insufficient. (Deecke, 1987a, pp. 781–782.)

Penrose (1989) was of the opinion that the concept of "time" is really cumbersome when discussing things like consciousness. Or, as he puts it:

> I suggest that we may actually be going badly wrong when we apply the usual physical rules for *time* when we consider consciousness! /…/ Consciousness is, after all, the one phenomenon that we know of, according to which time needs to 'flow' at all! The way in which time is treated in modern physics is not essentially different from the way in which *space* is treated [footnote excluded here] and the 'time' of physical descriptions does not really 'flow' at all; we just have a static-looking fixed 'space–time' in which the events of our universe are laid out! Yet, according to our perceptions, time *does* flow /…/. My guess is that there is something illusory here too, and that the time of our perceptions does not 'really' flow in quite the linear forward-moving way that we perceive it to flow (whatever that might mean!). The temporal ordering that we 'appear' to perceive is, I am claiming, something that we impose upon our perceptions in order to make sense of them in relation to the uniform forward time-progression of an external physical reality. (Penrose, 1989, pp. 574–575; italics in original.[1])

Frith (2003) questions whether the finger-lifting task really is an example of "free will". Or as he puts it: "When Libet tells you to lift your finger whenever you feel the urge, you're well aware he could be cross if you never had the urge. So you're selecting from a specific sub-category of responses" (Frith, 2003, p. 46). Free will, according to Frith, occurs before the selection of a particular action.

After what might seem is a little detour from speech production, it should however be obvious to the reader that Libet's (and others) results should affect speech production models that include monitors that detect and react to the speech string being created. If detection and correction *is* indeed occurring, *can* they be conscious? Are we consciously capable of making split-second repairs, sometimes faster than the 300 ms time window mentioned above. After all, speech articulators are motor units, muscles that respond to brain commands in the same way as the fingers in the Libet experiments.

But, is it that simple? Can motor action like finger bending be compared to speech production, "just like that"? Can it at all be generalized to other willed actions? This has been questioned by e.g. Breitmeyer (1985), Bridgeman (1985), Danto (1985), Jung (1985) and Latto (1985). Jung (1985) and Breitmeyer (1985) also point out that *overlearned* activities not necessarily need to be conscious at all.[2] Perhaps speech production is such an over-learned activity?

The problem of consciousness and willing is of course a much more complex issue than there is space for here. For example, one cannot take for granted that willed acts are the same as involuntary acts. Indeed, as is pointed out in Kimble & Perlmuter (1970), a voluntary eyeblink differs both in form and latency from a conditioned, involuntary eyeblink. For a good review of different views on volition, the reader is referred to Kimble & Perlmuter (1970), who describe and comment on volition models since Sechenov (1863/1935) and James (1890).

---

[1] Penrose's theory received support, on various grounds, from e.g. Glynn (1990) and Hameroff (1998a, 1998b).

[2] This is also argued in Langer & Imber (1979), who, when studying a translation task, found that overpractice resulted in performance decrement when subjects were assigned an inferiority label. They claim that "… as overlearning leads to mindlessness, the individual components of a task become relatively inaccessible to consciousness and therefore unavailable to serve as evidence of task competence" (Langer & Imber, 1979, p. 2014). However, by making the task components salient, the detrimental effect could be prevented.

2.6.8.4 Brain potentials and speech processing

It should be clear by now that readiness potentials have serious implications for speech production models, or minimally have severe implications for speech production. Or, as Helen Neville succinctly titles one of her papers: "Brain potentials reflect meaning in language" (Neville, 1985).[1] The question that presents itself is naturally to what extent research has attempted to study readiness potentials in the production of speech. A fairly recent answer is given in Garnsey (1993), with the title "Event-related Brain Potentials in the Study of Language: An Introduction", but it mainly focused on technical and methodological aspects, rather than linguistically oriented issues.[2]

Haggard, Newman & Magno (1999) mention speech as an area that should benefit from such studies (without devoting research to speech themselves), and such studies have also been carried out during the last decades that covered speech production. It should be pointed out that while these studies all concern event-related brain potentials, not all study the "readiness potential", or "Bereitschaftspotential" as identified by Kornhuber & Deecke (1965), or Libet (op. cit.) but also other, similar and related, brain potentials, like the **Contingent Negative Variation** (CNV) (Walter et al., 1964; Rohrbaugh, Syndulko & Lindsley, 1976; Herning & Jones, 1984), the respiratory **R wave** (Grözinger, Kornhuber & Kriebel, 1973), the meaning-related **N400** (Kutas & Hillyard, 1980; see also Chwilla, Kolk & Mulder, 2000, p. 317), **Transcranial Magnetic Stimulation** (TMS; Rothwell, 1998) and so on. While the blanket name for these is **ERP**, for "event-related potential", the terms "evoked potentials" or "evoked responses" are also used (Garnsey, 1993, p. 338).[3]

Speech Production Early work on cortical activity in speech production was carried out by Ertl & Schafer (1967, 1969), Schafer (1967), McAdam & Whitaker (1971), Morrell & Huntington (1971, 1972), Grabow & Elliott (1974) and Szirtes & Vaughan (1973, 1977). These early studies found that bilaterally symmetrical potentials begun up to 500 ms prior to word articulation with larger negative potentials occurring over the left hemisphere (McAdam & Whitaker, 1971). Szirtes & Vaughan (1973) found electrical manifestations of muscle activity that began 500 ms before sound production, and that the potential shifts could be of either polarity (negative or positive). That the readiness potential preceding speech can have either polarity (as opposed to voluntary limb movement, where it is negative) was also observed by Grözinger, Kornhuber & Kriebel (1973). Grözinger et al. (1974) also found that the readiness potential preceding speech was asymmetric (in contrast to hand movements[4]), "accounting for hemispheric dominance involved in speech" (Grözinger et al., 1974, p. 435). In contrast with these results, Grabow & Elliot (1974) and Morrell & Huntington (1972) found no asymmetries for pre-speech activity.

However, one problem with the production of speech—which was noted from very early on—is that it is so much more complex an action than e.g. the bending of a finger. Indeed, speaking is amongst the most complex motor actions humans exhibit, if not *the* most complex motor action. There are many muscles involved, ranging from breathing muscles (Grözinger,

[1] Other papers that describe the value of ERP studies—within linguistics and generally—are e.g. Neville (1980), Coles (1988) and Picton & Cohen (1984).

[2] Garnsey (1993) is an excellent introduction to the field of event-related potentials research, and is highly recommended to interested readers.

[3] Event-Related Potentials are divided into five subcategories in MacKay (1969, *viz.*, *Evoked Potential*, *Motor Potentials*, *Long-Latency Responses*, *"Steady" Potential Shifts*, and *Extracranial Potentials* (MacKay, 1969, pp. 206–207).

[4] Note the difference between RP and LRP.

Kriebel & Kornhuber, 1974; Grözinger, Kornhuber & Kriebel, 1973; Grözinger et al., 1974), tongue, lips, glottis, laryngeal muscles and so on. Already Morrell & Huntington (1971) pointed out that "[r]ecording from speaking subjects is assuredly a formidable problem in EEG" (Morrell & Huntington, 1971, p. 1360). Grözinger, Kornhuber & Kriebel (1975) tried to disentangle all possible artifacts of speech production that could contaminate the study of brain potentials, and included galvanic skin response, head movements, eye blinks and other.

More recent work that analyzes artifacts is Grözinger et al. (1980), Wohlert (1993) and Wohlert & Larson (1991). Szirtes & Vaughan (1977) pointed out that "[d]ue to the extensive distribution of speech-related activity, the reference sites also importantly influenced the characteristics of the potentials" (Szirtes & Vaughan, 1977, p. 388). They also observed that there were "markedly different potentials associated with different speech sounds" (Szirtes & Vaughan, 1977, p. 388). Szirtes & Vaughan (1977) also observed that "the slow activity preceding speech may be either positive, negative, or absent altogether, depending upon the utterance and the individual subject" (Szirtes & Vaughan, 1977, p. 392). They conclude that "it seems naive to have expected that cortical potentials could be recorded uncontaminated by extracranial activity" (ibid., p. 394), and suggested that further studies must be carried out using intracranial recordings (rather than scalp recordings). Grözinger et al. (1980) also point out that "pre-speech activity is distributed widely over the head" (Grözinger et al., 1980, p. 803). From a linguistic perspective, Indefrey et al. (2001), using PET, reported on neural correlates of syntactic encoding during speech production.

Speech perception As we have seen, there are obvious problems associated with the recording of potentials associated with speech *production*, given the complexity of the activity. However, brain potentials have also been used, to study language and speech *perception*, both from a purely neurological perspective—e.g. the lateralization of speech in the brain—but also from a *linguistic* perspective, like how the brain reacts to different morphological, phonological, semantic, syntactic and prosodic aspects. In early work, Morrell & Salamy (1971) found hemispheric asymmetry when subjects were exposed to speech stimuli. From a linguistic perspective, studies have focused on **morphology** (McKinnon, Allen & Osterhout, 2003; Friederici, Pfeifer & Hahne, 1993), **semantics** (e.g. Burian, Gestring & Haider, 1969; Kutas & Hillyard, 1980,[1] 1984; Neville, 1980; Thierry, Cardebat & Démonet, 2003; Herning, Jones & Hunt, 1987; Holcomb, 1988, 1993; Rothenberger et al., 1987; Novick, Lovrich & Vaughan, 1985; Hagoort & Brown, 2000; Bentin, Kutas & Hillyard, 1993; Van Petten, 1995; Boddy & Weinberg, 1981; Connolly et al., 1992; Connolly, Stewart & Phillips, 1990; Friederici, Pfeifer & Hahne, 1993), **lexical/word-class processing** (Tyler et al., 2001; Chwilla & Kolk, 2000; Federmeier et al., 2000; Helenius et al., 1998; Van Petten & Kutas, 1987), **syntax** (Van Turennout, Hagoort & Brown, 1998; Van Petten & Bloom, 1999; Friederici, 1995; Friederici et al., 1998; Friederici, Pfeifer & Hahne, 1993), **phonology** (Rumsey et al., 1997; Van Turennout, Hagoort & Brown, 1998; Lee et al., 1999; Rugg, 1984), **prosody** (Steinhauer, Alter & Friederici, 1999), to mention but a few.

Röder, Rösler & Neville (2000) studied N400 in 11 congenitally blind and 11 sighted adults, and found that the N400 effect to semantically incongruous stimuli started earlier in the blind than in the sighted subjects, suggesting that blind people process auditory language faster than do sighted people.

[1] This is the first study to refer to the N400 component, which is an event-related potential related to semantic processing. It is shown to vary inversely with the semantic relatedness of target words, so that the closer the relationship, the smaller the N400 amplitude. Consequently, it is conversely related to the Cloze probability of a word (Strandburg et al., 1997, p. 597).

Finally, rather than just using brain potentials to localize areas involved in certain aspects of speech processing, MacNeilage & Davis (2000) acknowledge the problem the observed timing events pose, but simply by pointing out that the "Bereitschaftspotential" illustrates the complexity of "initiation of voluntary movement" (MacNeilage & Davis, 2000, p. 529), but they do not refer to any of the previously mentioned studies, only to Kornhuber's two-page article in *Encyclopedia of Neuroscience* (Kornhuber, 1987).

The thing of interest to note here is that different linguistic entities such as phonemes, semantics, gender (syntax) and so on, seemingly *do* have an observable neurological basis, and that this can be studied, both from a production-based perspective (despite problems of contamination) and from a comprehension/perception-based perspective. This is also probably as close as we can get to "opening the lid" to the brain and speech processing.

2.6.8.5 Brain potentials and disfluency

So, given the range of studies devoted to language aspects such as phonology, morphology, syntax, semantics and prosody, are there any brain potential studies devoted to speech disfluency? The answer to this question is that such studies exist, both directly and indirectly, as we shall see.

Stuttering Besides the studies already referred to concerning hemispheric lateralization in stutterers, there are studies that compare brain potentials in stutterers and nonstutterers. Zimmerman & Knott (1974) studied the contingent negative variation (CNV) in stutterers and nonstutterers in a verbal and a nonverbal task, and found that 80% of the nonstutterers showed a larger shift in the left hemisphere preceding speech, while only 22% of the stutterers showed a greater left hemispheric asymmetry. Also, there were differences between the two groups even when stutterers did not approach moments of stuttering, i.e., when the two groups had equal speech performance (at the surface level). Prescott (1988) also studied CNV in stutterers and nonstutterers and concluded that the stutterers had problems in setting up the parameters of a response, rather than in the ongoing control of speech.

In a recent study, Salmelin et al. (2000) had ten fluent speakers and nine developmental stutterers read isolated nouns aloud in a delayed reading test. During the test, brain activity was mapped using a whole-head magnetoencephalography system. During the test, the stutterers were mostly fluent. However, there were differences between the two groups in their brain activity:

> Although the overt performance was essentially identical in the two groups, the cortical activation patterns showed clear differences, both in the evoked responses, time-locked to word presentation and mouth movement onset, and in task-related suppression of 20-Hz oscillations. Within the first 400 ms after seeing the word, processing in fluent speakers advanced from the left inferior frontal cortex (articulatory programming) to the left lateral central sulcus and dorsal premotor cortex (motor preparation). This sequence was reversed in the stutterers, who showed an early left motor cortex activation followed by a delayed left inferior frontal signal. Stutterers thus appeared to initiate motor programmes before preparation of the articulatory code. During speech production, the right motor/premotor cortex generated consistent evoked activation in fluent speakers but was silent in stutterers. (Salmelin et al., 2001, p. 1184.)

As was discussed earlier, these observations support the notion that fluent speech in stutterers is not equal to fluent speech in normal speakers, even at the deepest level. This is also supported in Khedr et al. (2000), who compared stutterers and normal subjects using a

variety of different brain potentials (visual and auditory evoked, P300, encephalography), and found that the dominant EEG rhythm was slower in stutterers with strong interhemispheric asymmetry, as compared with the controls. They conclude that their findings "point to a possible role of an organic etiopathogenesis of stuttering" (Khedr et al., 2000, p. 178).

Disfluency Although no brain potential studies—to the best of my knowledge—have been devoted to speech disfluency proper, studies have been done on verbal *fluency*. Wise et al. (2001) re-analyzed four functional neuroimaging studies, and found that non-speech and speech sounds (including the subject's own voice) activated the supratemporal cortical plane, while activity "in its most posterior and medial part, at the junction with the inferior parietal lobe, was linked to speech production, rather than perception" (Wise et al., 2001, p. 83). More interestingly, from our perspective was their observation that:

> The second, more lateral and ventral part lay in the posterior left temporal sulcus, a region that responded to an external source of speech. In addition, this region was activated by the recall of lists of words during verbal fluency tasks. (Wise et al., 2001, p. 83.)

While studies of disfluency proper seem to be rare, there are a few studies that have included the effects of **pauses** on comprehension.

Lee et al. (1999) performed an fMRI[1] study of phonemic and semantic fluency in Chinese-speaking subjects, and observed that both tasks revealed strong left-hemisphere dominance, but also that there were "subtle differences in the representation of the central processing in the brain between ideographical-based and alphabetical-based systems" (Lee et al., 1999, p. 1062). Holcomb & Neville (1991) recorded brain potentials as subjects listened to spoken sentences. In a first series of experiments, all sentences were presented as connected speech, in a second series all words were separated by a 750 ms silent interval. Three types of sentence-ending words were used: *best completions* (contextually meaningful), *unrelated anomalies* (contextually meaningless), and related anomalies (contextually meaningless, but related to the best completions). Large N400 components were observed for both related and unrelated anomalies, relative for the best-completions final words (thus confirming the N400 effect). However, the auditory N400 onset earlier in the connected speech experiment than it did in the version with interword silent intervals.

Besson et al. (1997) studied ERPs in two experiments where the temporal patterns in reading and listening to sentences where disrupted by inserted pauses (with a duration of 600 ms). The two modalities resulted in different ERP responses, suggesting that processing of natural speech is different from reading, with the former more resembling earlier work on ERP responses in the processing of musical phrases. An interesting thing to note, not pointed out by Besson et al. (whose interest lies in comparing speech to text) is that the fact that brain potentials *are* observed at the location of pauses ("temporal disruptions") could be interpreted as evidence that unfilled pauses do have an effect on our speech comprehension system that is different from our processing of fluent speech. In other words, when the normal flow of a sentence is disrupted, the brain *reacts*. How this occurs varies as a function of the modality, but it still reacts to disfluency. This indicates that disfluency does indeed have an effect on speech comprehension, but also that the study of brain potentials is useful from the point of view of disfluency in general.

[1] The abbreviation MRI stands for *Magnetic Resonance Imaging*, a method that uses radio waves and a strong magnetic field to study inner organs. fMRI (for *functional Magnetic Resonance Imaging*) works by measuring the metabolic changes that take place in the active parts of the brain.

2.6.8.6 Integrating it all

The way I view things, and which I have tried to show here, is that there are two roads leading to an interdisciplinary street crossing, and also that the drive there is not even very long. One emanates in the study of slips-of-the-tongue, which has served as the basis for linguistically based models of speech production. The other has its starting-point in the neurological studies of how humanly initiated motor actions—of which speech is but one example—are reflected in the brain's electrical activity, as reflected in different kinds of brain potentials. Brain potentials have been studied both for speech production and speech comprehension, and have resulted in observations concerning the lateralization of language in the brain, but also in linguistically interesting observations concerning the dissociation between semantic, syntactic and acoustic information. Most interesting from our point of view is of course the result of Besson et al. (1997), since they show that unfilled pauses result in observable brain potentials.

That brain potentials can be studied without taking into consideration the backward referral à la Libet should be obvious from the presentation above. That does not mean, however, that it is without interest, and we shall devote some space to that issue in the following section. The burning question, as it were, is whether there are any attempts to "bring it all together"? Are there any attempts to bridge these interdisciplinary gaps? Perusing the literature, I have found a couple that could serve as candidates, and will shortly describe them in the following. To the best of my knowledge, Velmans's target article (1991a/1991b) seems to be the most important attempt to date to make a stab at bringing together human information processing in general, speech perception/production and Libet's observations concerning time events in human motor control. When discussing where (and how) consciousness enters into human information processing (including the notion of preconscious and unconscious processing, see Dixon, 1981), Velmans cites several works in the psycholinguistics literature (e.g., Cherry, 1953 and Moray, 1959, on dichotic listening tasks), and also discusses human speech *perception* (Velmans, 1991a/1991b, p. 655 et passim), pointing out that if words in context are recognized within 200 ms, then the "confluence of data-driven [bottom-up] and cognitively driven [top-down] processing cannot be conscious" (ibid., p. 657), given the experimental findings that "consciousness of a given stimulus does not arise until at least 200 ms *after* the stimulus has arrived at the cortical projection areas" (Velmans, 1991a, p. 658; italics in original) as observed by e.g. Libet et al. (1979), Neely (1977)[1] and Posner & Snyder (1975). It goes without saying that this affects the speed with which Levelt's outer loop could react to disfluent speech and make the necessary corrections.[2]

Of more interest, for the present purposes, Velmans then discusses speech production (Velmans, 1991a, p. 663). He discusses the lack of awareness of the "myriad of motor commands" the central nervous systems commands:

> In speech, for example, the tongue may make as many as 12 adjustments of shape per second – adjustments that need to be precisely coordinated with other rapid, dynamic changes within the articulatory system. According to Lenneberg (1967), "Within one minute of discourse as many as 10 to 15 thousand neuromuscular events occur." Yet only the *result* of this activity (the overt speech) normally enters consciousness. (Velmans, 1991, p. 663; italics in original.)

[1] Velmans erroneously spells the name "Neeley" (Velmans, 1991a, p. 657.)

[2] That speakers can make use of an outer, auditory, feedback is evident from the fact that speakers compensate for perturbed acoustic feedback of their own vowel production in order to produce perceptually "correct" target sounds (Houde & Jordan, 1998).

Velmans—who does not refer to Levelt—seems to preclude the possibility of an omniscient inner loop at work in speech processing, at least one that is available to consciousness or awareness. Referring to Bock's (1982, 1986) model of speech production (a precursor of Levelt), which divides speech production into six "arenas"[1] (cf. Levelt's modules), Velmans asks whether the planning of speech could be conscious? And this is where we go full circle in what may have appeared as a detour from the topic proper, since Velmans refers to Goldman-Eisler's (1968) and Fodor et al.'s (1974) work on hesitation pauses.

Let me cite Velmans in some detail:

> Hesitation pauses tend to occur within clauses and sentences and appear to be associated with the formulation of ideas, deciding which words best express one's meaning, and so on. In assessing whether the planning of *what* to say is conscious, it is hence instructive to examine what one experiences during a hesitation pause (where we have good reason to infer such planning to be taking place). This simple thought experiment reveals that during a hesitation pause one might experience a certain sense of effort (perhaps the effort to put something in an appropriate way), but nothing is revealed of the *processes* that formulate ideas, translate these ideas into a form suitable for expression in language, search for and retrieve words from memory, assess which words are more appropriate, and so on. In short, no more is revealed of conceptual or semantic planning in hesitation pauses than is revealed of syntactic planning in breathing pauses. The fact that a process demands *effort* does not ensure that it is *conscious*. Indeed, there is a sense that one is only aware of what one wants to say after one has said it! Nor is the situation any different if one expresses one's thoughts in covert speech through the use of phonemic imagery. Covert speech and overt speech bear a similar relation to the planning processes that produce them. In neither case are the complex antecedent processes available to introspection. (Velmans, 1991, p. 663–664; italics in original.[2])

Even if Velmans does not delve much further into speech production (or disfluency research), he acknowledges its role in consciousness studies, and its relation to Libet's findings concerning relative timing in the brain during motor events, where language is "at the upper end" of the "cognitive spectrum" (Velmans, 1991b, p. 704).

Velmans's target article was published as an open peer article in *Behavioral and Brain Sciences*, and is consequently discussed by a number of researchers (pp. 669–702). Regrettably, from our point of view, very few pick up the language/speech thread in any detail, which is understandable, since the main goal of the article is to present a model of human consciousness (or awareness) and its role in information processing in general, and that speech and language only serves as one example to support Velmans's arguments). However, among the comments made, Gray (1991) acknowledges Velmans's argumentation that language processing can occur without consciousness, but thinks that a corollary question is whether consciousness can occur without language (Gray, 1991, p. 679). Underwood (1991), argues that "there is strong evidence to suggest that without attention there is limited integration of the words in a sentence" (Underwood, 1991, p. 698). Van Gulick (1991) is of the opinion that some of Velmans's argument is off the target, especially concerning comprehension and production of speech, and claims that while language is generally considered the result of highly task-specific and informationally encapsulated modules (compare Levelt's model), consciousness is normally thought to "involve very general

[1] These are: 1. A referential arena, where some nonlinguistic code is generated and passed onto the ensuing linguistic arenas. 2. A semantic arena. 3. A syntactic arena. 4. A phonological arena. 5. A phonetic arena. 6. A motor assembly arena. Note the similarities with Levelt's model (1989).

[2] Nisbett & Wilson (1977) discuss the issue whether or not cognitive processes are accessible by introspection.

nonmodular and informationally nonencapsulated processes." (Van Gulick, 1991, p. 699). Given its wide scope and the ambitiousness of his undertaking, it is only natural that Velmans does not delve deeper into language production proper.

A second bridging article is Baars (1992c), who explicitly mentions Libet and the problems caused by his (and others') results and observations. Baars, when presenting his **Ideomotor Theory of Voluntary Control**, a neurologically as well as psychologically based theory, makes the following remark:

> Cortical activity *by itself* appears to be unconscious (e.g. Libet /…/ 1985[a/b]). (Baars, 1992c, p. 96; italics in original.)

And, a little later:

> Libet (1985[a/b]) has presented arguments that we may become conscious of some actions only *after* the brain events that immediately trigger them. But this cannot be true in every case; surely, there are a great many times when people are conscious of what they are about to do seconds or hours before they do it, as shown by the fact that they can accurately discuss and predict their actions beforehand. (Baars, 1992c, p. 104; italics in original.)

So, Baars, with an affluence of linguistic studies in his output—mainly as one of the discoverers of the lexical bias phenomenon in slips-of-the-tongue (Baars, Motley & MacKay, 1975)—and with his own theory of consciousness (Baars, 1988), appears in a way to be a tad "disturbed" by the Libet findings. However, his counter-argument seems a little misconceived. In the Libet experiments, the subjects *knew* beforehand that they were going to move their fingers. That was part of the instructions they were given. The fascinating thing was that when they later *made* the movement, RP *still* preceded the movement. This would have been the case irrespective of whether the subjects postponed their movements by "seconds or hours". Indeed, that this is the case is often part of the criticism Libet has received, since acting on directives cannot be considered a true example of spontaneous behavior (e.g. Näätänen, 1986; Ringo, 1986). Moreover, it has also been shown from early on that RPs (and similar brain potentials) are influenced by psychological states such as anxiety and stress (Knott & Irwin, 1973), attention, risk or reward (e.g. Hink et al., 1982; Foit, Grözinger & Kornhuber, 1982; McAdam & Seales, 1969), motivation (Irwin et al., 1966) that attention to a cued location in space leads to faster reaction times (Johnson & Haggard, 2003), and that there are differences between pre-planned and non pre-planned events[1] (Libet, 1993) Baars agrees that we are not aware of the details of our actions, or, as he points out: "[W]hat is the difference between pronouncing /ba/ and /pa/? Most people simply do not know" (Baars, 1992c, p. 105). Without further addressing the RP problem, he proposes a speech production models that resembles Dennett's *Multiple Drafts Model* (and is also mentioned by Dennett (1991):

> Just as in language there are often dozens of ways of saying the same thing, /…/ the action is carried out by specialists that know more about local conditions than we do consciously. Various unconscious specialists keep continuous track of our posture, balance, and relationship to gravity (Baars, 1992c, p. 105.)

[1] "[T]he readiness potential (RP) /…/ begins first at about –1050 ms when some pre-planning is reported (RP I) or about –550 ms with spontaneous acts lacking immediate preplanning (RP II)" (Libet, 1993, p. 128).

At least on the surface, this seems to resemble Dennett's demons, a notion further strengthened by Baars's remark that "[c]onscious goals seem to be inspected and edited by multiple, simultaneous, unconscious criteria" (Baars, 1992c, p. 106). No doubt is this different from Levelt's feed-forward, modular, conscious-editing, model.

Summing up, disfluencies, while serving as the basis for most speech production models (notably, Levelt, Nooteboom, Postma & Kolk and others), are not always present in the literature that discuss such proposed models (e.g. Dennett). Also, while the literature in neurology has treated the timing of action events in the brain for over thirty years, these findings do not seem to have been discussed to any extent in the speech production literature, or from any linguistic perspective in general. Indeed, Velmans and Baars might well be the only works that even approach including all the different fields mentioned above, and as such, it is not strange that speech is not mentioned to greater extent.[1]

It might be seen as "upping the ante" to semi-divert into such issues such the consciousness, free will and so on in a study of a (seemingly) narrow focus as the incidence of disfluencies within a constrained domain, but I would like to claim that there is a straight line from (or rather: between) the single *eh* to such phenomena, and that although they are not the main focus of *this* book, the results presented later on could be imported lock, stock and barrel into the fields of speech production, neurology, neurophilosophy, psychology and others.

To further illustrate the interrelatedness of disfluency and these ephemeral phenomena, let me cite Sellen & Norman (1992):

> [S]ome action takes place *in spite* of conscious desires and some takes place even in *the absence* of conscious awareness. However, these departures from desired behavior are still the infrequent case. After all, slips may be common, but they are not the dominant behavior. /…/ [T]he challenge becomes one of understanding and modeling a system in which the details of action can be specified without appealing to some type of all-knowing executive controlling agency. This is not to say that we can ignore the role of "volition," "will," "consciousness," and "intention." In many ways the study of slips forces us to confront these elusive concepts head-on" (Sellen & Norman, 1992, p. 319.)

Frith (2002) argues:

> Awareness of choosing one action rather than another comes after the choice has been made, while awareness of initiating an action occurs before the movement has begun. These temporal differences bind together in consciousness the intention to act and the consequences of the action. This creates our sense of agency. (Frith, 2002, p. 481).

How this relates to our sense of agency, or conscious will, when speaking, and exactly how disfluencies come into the picture remains to be tested.[2]

The relevance of reaction times *per se* to speech production models has been observed in the literature, without relating it to the aforementioned discussion on free will or consciousness. Blackmer & Mitton (1991) note that:

---

[1] Incidentally, Baars's (1991) criticism of Velmans (1991) is that the latter's position is epiphenomenalist (Baars, 1991, p. 669), something Velmans denies (Velmans, 1991b, p. 712).

[2] For further discussion on agency, the reader is referred to e.g. Bargh & Chartrand (1999), Wegner & Wheatley (1999) and Wegner (2002).

> Although existing [speech production] models can be seen to have different implications for the timing of detection and repair of problems in speaking, the authors have seldom explicitly stated the temporal implications of their models. (Blackmer & Mitton, 1991, p. 174.[1])

Consequently, in an attempt to remedy this state of affairs, Blackmer & Mitton studied reaction times for error-to-repairs and cut-off-to-repairs, and compared the results to three proposed speech production models, Laver (1980b), Levelt (1983a, 1983b, 1989) and Kempen & Hoenkamp (1987). They found that many of the cut-off-to-repair times were faster than were predicted (or even possible) in any of the three models, while Laver's (1980) model was found to be incompatible with the observations. As to Levelt (1989), the main problem was his proposed *main interruption rule*, that states that when a problem is (consciously) detected, speech is interrupted immediately. Levelt also provides a latency of 200 ms after detection for the halting to take effect.

Kempen & Hoenkamp (1987) were less specific about their monitor than was Laver (1980b) or Levelt (1983b, 1987), which makes their model harder to evaluate. However, the monitor resides outside the speech production components, and works in an incremental way, checking output from each of the production modules. If this is done serially or in parallel is not clear. However, given the observed cut-off-to-repair times (the shortest being 0 ms!), Blackmer & Mitton (1991) find support for some kind of incremental buffering, such as that proposed in Kempen & Hoenkamp (1987).

It would be of utter interest to try to encompass or incorporate detection/repair times of 0 ms in any proposed speech production model, in the light of what is known about initiation times in motor processes, especially since some monitoring is claimed to be conscious (e.g. Levelt). Also, several of the brain potential studies mentioned above also refer to reaction times, insofar as they mention event latencies upon external stimuli that prevent outer loops of very short reaction times.

To summarize the field of speech production is difficult, for a number of reasons, the most central being its inherent complexity. As is evident in e.g. Rapp & Goldrick (2000), even comparison of related models is not straightforward, and most observations, assumptions and conclusions are indirect. An underlying assumption in most studies of speech production, however, is that something goes wrong, which appears on the surface as disfluencies. The study of this malfunction of the speech production system is what makes us conjecture what this system might look like. This assumption makes the silent claim that disfluencies are not on a par with other linguistic items, such as words, that are successfully produced, i.e., the way they "should be", a notion that perhaps received some support from the findings made by Besson et al. (1997), but which all the same has been questioned, as will be shown later on.

To the extent that decisions can be said to be made during the process of speech production—irrespective or not whether these be made consciously or subconsciously—reaction times, and general motor functions need to be addressed. Whether or not one "believes" in the results of Kornhuber, Deecke and Libet (and others), the implications of their, and similar, research is so far-reaching that I find it difficult to ignore, and the study of speech production should, in one way or another, take a stand on their findings.

[1] In passing, this is also my own observation, and the reason for this rather "detailed synopticon" of the area. An exception is Donald G. MacKay (1987) who devoted much space to both timing implications and eletrophysiological bases for his theory. He does not, however, refer to Libet's findings.

Also, as was pointed out earlier, although Levelt's seminal work was titled *Speaking*, not all human languages are *spoken*, but perhaps an equal number of human languages are signed (which will be described later). Since none of the models above seem to put much responsibility of the speech production process on the articulators of spoken language (i.e., tongue, lips, velum, vocal folds and so on), the same hesitation or disfluency phenomena should show up in sign language as well.[1]

I am going to let Baars (1992b) have the final word in this section, with a final twist on the speech (language) production problem:

> [N]o current theory, for example, can account for the elementary fact that people can repeat their own slips voluntarily. Yet speech errors are often repeated spontaneously, as when someone says, "Did I say X? I really meant Y!" Note that the slip and its voluntary imitation are behaviorally identical—but psychologically they are vastly different." (Baars, 1992b, p. 4.)

… and, in the same vein:

> [L]ocal theories of speech and action do not account for the fundamental distinction between slips and their voluntary imitations: They have no mechanism for showing how an error could be intentional, or how a generically correct phrase could, in principle, be unintentional. A complete account of slip phenomena must be able to represent this difference. (Baars, 1992b, p. 22.)

As if it was not complicated as it is!

2.7 Inner speech: evidence from schizophrenia?

As should be clear from the previous paragraphs, the most common way to form hypotheses concerning how speech is produced in the black box (the brain) that produced it, has been, and still is, to study how it goes wrong. An additional source of "distorted" language can be found in the study of schizophrenic speech, and some such studies have indeed been carried out, which will be discussed in the following paragraphs.

2.7.1 Covert schizophrenic speech

While no one has attempted to explain, at least in any detail, what "language" Levelt's Conceptualizer "speaks", the status of inner, or covert, speech has been extensively discussed in the literature, especially with regard to whether or not it is conscious, preconscious or subconscious. Evidence to the effect that covert speech could very literally in fact be *speech* comes from studies on schizophrenics (Jaynes, 1976/2000, 1986, 1990; Hamilton, 1985). Among the most common symptoms of schizophrenia are verbal auditory hallucinations (WHO, 1975;[2] Mellor, 1970; Chapman, 1966; Johnson & Miller, 1965; Miller, Johnson & Richmond, 1965; Loftus, Delisi & Crow, 2000). The argument basically boils down to the hypothesis that the covert speech in schizophrenics is like that of normal speakers, but in schizophrenic patients it gets misattributed to external sources—as is also the case with other motor actions or events in schizophrenics (Blakemore & Frith, 2003; Blakemore, Oakley & Frith, 2003; Frith & Done, 1989)—and the (literal) voices of the brain are heard as coming from the outside. As is pointed out in Frith (1979), "[n]ot only are the majority of

[1] Sign language will be treated in section 2.8.
[2] WHO (World Health Organisation). 1975. *Schizophrenia: A multinational study*.

schizophrenic hallucinations auditory, but they specifically involve the hearing of words" (Frith, 1979, p. 228). Frith (1999), in later work, mentions that "[t]he patient does not hear just sounds but fully formed verbal communications that appear to emanate from a particular speaker or group of speakers" (Frith, 1999, p. 414), and draws the conclusion that "[a] number of authors /…/ have suggested that these experiences have their origin in the patient's own inner speech or thought" (Frith, 1999, p. 415). That schizophrenics have genuine auditory experiences is further supported by studies that show that the primary auditory areas of the brain (notably Heschl's gyrus) show increased activity during hallucinations (Dierks et al., 1999; McGuire, Shah & Murray, 1993). Thus, it has been claimed that verbal hallucinations in schizophrenic patients depend on defective self-monitoring (Hoffman, 1986; Johns et al., 2001; Frith, Blakemore & Wolpert, 2000). That verbal, auditory, hallucinations also occur in the *normal* population has also been shown (e.g. Jaynes, 1990; Posey & Losch, 1983; Posey, 1986; Bentall & Slade, 1985), which further strengthens the hypothesis of inner speech in a literal sense. Also, in a number of studies on epilepsy,[1] Penfield and colleagues (summarized in Penfield & Perot, 1963) found that electrical stimulations of the cortex elicited auditory hallucinations, verbal, musical and visual.

From a language planning perspective of schizophrenia, Hoffman (1986) is the most exhaustive work to date. He presented a model of speech (dis)organization—based on verbal hallucinations of schizophrenics—with four main claims: 1) That sensory properties of verbal hallucinations are not distinct from ordinary verbal imagery. 2) The verbal hallucinations are accompanied by a feeling of unintendedness. 3) That the (characteristic) unintendedness in point 2 is caused by disruptions in the language-planning processes. 4) That this unintendedness is the basis for the conviction of the patient that the verbal imagery has a non-self origin. Although Hoffman's argument basically treats disordered speech at higher levels, such as discourse levels, he does include slips-of-the-tongue, and problems associated with e.g. lexical retrieval, likening a schizophrenic's view that verbal hallucinations are unintended with a normal subject's view that slips-of-the-tongue are unintended, although not perceived as having a non-self origin in healthy subjects. Hoffman also pointed out that schizophrenics sometimes also produce involuntary overt speech, often experienced as unintentional (or even having a non-self origin).

Akins & Dennett (1986) claimed that Hoffman's (1986) model "is in effect the sketch of a theory of slips of thought" (Akins & Dennett, 1986, p. 517), when an easier way to view things would be the view that "verbal imagery /…/ is always the execution or misexecution of communicative intentions. /…/ it is quite possible to make middle-level production errors – or word choice, for instance and recognize and correct them." (Akins & Dennett, 1986, p. 517).[2] Akins & Dennett believe that a view based on unintended *speech acts* (in discordance with the intended ones) is an easier way to describe that the verbal imagery in schizophrenics is not interpreted as mere verbal slips, mispronunciations or spoonerisms, but are attributed to external sources. (In effect, this view is not too far from Hoffman's own account.)

Bentall & Slade (1986) raised the opinion that Hoffman's view of speech disturbance and verbal hallucinations as the result of one and the same deficit, a disorder of discourse

[1] That epileptics can experience auditory hallucinations has been known for a long time. Penfield & Pharot (1963, p. 600) refer to the Arabian physician Abulquasim (10th century) who reported two cases of experiential hallucinations in epileptics.

[2] Akins and Dennett also ask whether one "can 'mispronounce' a word in verbal thought?" (Akins & Dennett, 1986, p. 517), and promptly give the answer: "Yes – think of reading the surnames in Russian novels." (Akins & Dennett, 1986, p. 517.)

planning, has little, if any, convincing evidence to support it. They also pointed out that Hoffman fails to take into account that the distinction between "real" and "imaginary" events is influenced by context.

Brand (1986) pointed out that overt speech is intentional (it is generally assumed) in that it requires cognitive schemata and discourse planning, and that if one assumes, as Hoffman does, that covert speech is, in most respects, similar to overt speech, then the former must also be considered intentional (cf. e.g. Levelt, 1989). Verbal hallucinations would then be actions that are nonconcordant with the cognitive schema for the concurrent covert speech, and, as such exhibit "mock intentionality".

Deese (1986) pointed out that Hoffman's model (that schizophrenic's hearing of voices is a result of the failure of unconscious discourse planning processes) bears resemblance to the model of consciousness proposed by Jaynes (1976/2000).[1] Deese further raised the questions as to the intentionality of dreams[2] or the biological foundation of verbal hallucinations, both unaddressed by Hoffman.

Flor-Henry (1986) picked up the biological thread by referring to the findings that verbal hallucinations are accompanied by electromyographic activity in the vocal apparatus,[3] which has also been show to be the case during silent thought or silent reading. Flor-Henry argued that the fundamental defect in schizophrenia relates to impaired-dominant hemispheric functions, and that "auditory hallucinations are reflections of altered neural structures responsible for verbal–linguistic expression /…/ 'hallucinations of inner speech'" (Flor-Henry, 1986, p. 523).

Harley (1986) agreed with Hoffman that speech errors are caused by "fragments of conversational plans" that intrude into the speech output. He even goes so far as to suggest that the speech errors of normal speakers are in fact "the same" as the hallucinations of schizophrenics.

Jaynes (1986) pointed out that some kind of inner speech, in a literal sense, must exist, citing work by Hamilton (1985) that cerebral palsied spastic-atheoid nonverbal congenital quadriplegics, who have never spoken in their lives both are capable of understanding speech, and also report "hearing voices", normally of the same sex as the patient, normally with the voice of a relative, but most often identified as "God". Jaynes, who is very much in agreement with Hoffman, argued that Hoffman has missed an important point in his proposal, *viz.*, what the hallucinated voices actually *say*! Many (most?) of the voices are admonitory, and around 75% of the men hear commands, and most women hear criticisms (of the own person). A theory that fails to include the communicative aspects cannot be considered complete, according to Jaynes.

Schwartz (1986) questioned hallucinations as such, and wonder whether they may in fact be rationalizations of disturbed speech planning. He makes a comparison with the results obtained by the study of **split-brain patients**, who have had their two hemispheres separated by severing the corpus callossum (Akelaitis, 1944; Bogen, 1969; Bogen & Vogel, 1962; Bogen, Fischer & Vogel, 1965; Gazzaniga, 1967, 1970, 1983, 1992, 1998, 1999, 2000, 2002;

[1] Cited as "Jaynes (1977)" by Deese.
[2] Compare Foulkes's observation that "dream speech typically is both grammatically correct and appropriate to the imagined situation in which it is embedded" (Foulkes, 1990, pp. 39–40). See also Foulkes (1991).
[3] A point also made by Junginger (1986, p. 528).

Funnell, Corballis & Gazzaniga, 2000; Iacoboni et al., 2000; Gazzaniga & Hillyard, 1973; Gazzaniga & Sperry, 1967; Sperry, 1966, 1967, 1968; Sperry & Gazzaniga, 1967, Gazzaniga, Bogen & Sperry, 1965; Ivry & Robertson, 1998; see also Reuter-Lorentz & Baynes, 1992). This enables researchers to expose the two hemispheres with different information, with astonishing results. It shows that while the right hemisphere is capable of processing both linguistic and other information, and results in appropriate reactions in the subject (e.g. blushing as a reaction to erotic material), subjects fail to realize *why* they blush, nor can they verbalize what they have perceived, since it has never reached consciousness. What then happens is that rather than saying that they do not know why they blush (for example), subjects rationalize, and provide an explanation for the reaction out of the information that is consciously available at the time, which might result in more or less far-fetched accounts. This is interesting for several reasons. Schwartz's main point is that verbal hallucinations might be related to such rationalizations, in that erroneous speech plans might be attributed to external sources in an attempt to explain why it went wrong. However, extending beyond Schwartz, split-brain studies could of course also be used to test claims about hemispheric function within stuttering research, or speech production models in general, with obvious bearings into the notion of a "unity of mind". Luckily, one should say, there are very few patients around who have been subject to this drastic operation (people suffering from severe epilepsy), so this might be out of reach.

Zivin (1986) questioned that verbal *imagery* should be involved in the monitoring of "inner speech", and thinks that Hoffman must have misinterpreted some of the neurological references he based his assumptions on.[1] The point raised by Zivin is indeed crucial, and the "scant knowledge of the neurology of inner speech" (Zivin, 1986, p. 534), which she rightly brings to the fore, is evident in all speech production models. While most proposals include some kind of inner speech, any formal, or scientifically justified, definition of what inner speech *is* is indeed hard to disinter in the works cited above. Whether or not the "voices" of schizophrenics (or normals) are examples of "normal" inner speech (just more obvious), or something altogether different is of course hard to tell, but could at least be said to constitute an intriguing possibility.

2.7.2 Overt schizophrenic speech

So what about overt speech of schizophrenics? Does it exhibit specific traits from a disfluency perspective? That the speech of schizophrenics differs from normal speech in general is well-known, but what about structural studies? I will review a few such studies in this paragraph.

Frith (1987) listed "poverty of speech" as a trait typical of "Type II (chronic)" schizophrenics, thus making it a "negative sign", i.e. lack of a trait (Frith, 1987, p. 631). Forrest, Hay & Kushner (1969) studied schizophrenic speech in general, and described features such as "word salad", alliteration and over-inclusiveness. Feldstein (1962) studied the speech of schizophrenics, using Mahl's Speech Disturbance Ratio, which includes the ah and non-ah ratios, to gauge disfluency as compared to normal subjects, and found no significant differences in disfluency rates between schizophrenics and normal subjects. Gerald Silverman (1973) used the Cloze method to study redundancy, repetition and pausing in schizophrenic speech, and found very low Pause–Speech Ratios, "suggestive of a quite distinct type of language disturbance" (Silverman, 1973, p. 413).

[1] Among others Luria (1960, 1961) and Sokolov (1972).

Brown (1973) claimed that there was no such thing as schizophrenic speech, and that the language disorders encountered in the speech of schizophrenics mainly reflected disordered thinking.

Chaika (1974) analyzed speech in a schizophrenic patient, and found that one distinctive feature was the production of "gibberish", i.e. phrases of non-existing "English" strings, similar to long stretches of mispronunciations, or "slips-of-the-tongue". Like slips produced by normal subjects, the produced speech adhered to the phonology of the speaker's language, but unlike healthy subjects, the schizophrenic patient did not react to the speech produced, and the strings produced were also much longer than slips observed in healthy subjects.

The notion that schizophrenics fail in the editing phase of speech production was also supported by Smith (1970). Smith based his study on previous work by Rosenberg & Cohen (1966), who forwarded the idea that speech production makes use of a comparison phase where speakers compare what they are about to say with how the listener is likely to interpret the utterance about to be produced. Smith (1970) concluded:

> To sum up, schizophrenics communicated poorly, not because they produced deviant associations nor because they were unable to properly assess the relative strengths of associative relations, but because they failed to edit adequately their responses by considering the relation between what they did not want to communicate and what they were about to say. (Smith, 1970, p. 186.)

Chaika (1974) also mentioned that **opposite speech**, i.e. saying *yes* instead of *no*, which is observed in healthy subjects (especially children), as well, is also observed in schizophrenic speech (Laffal, Lenkoski & Ameen, 1956; Kaplan, 1957).

Chaika (1974) met with some opposition. First, Fromkin (1975) argued that the language disturbances listed by Chaika are typical of human speakers in general, and are not exclusively signs of schizophrenic speech.

Fromkin:

> [I]n arguing against the opinion of Brown (1972)[1] that schizophrenics do not reveal a breakdown in language so much as a breakdown in thought, she [Chaika] contends that the psychological or mental aberrations are paralleled by "a disturbance in those areas of the brain concerned with linguistic production." This is an interesting avenue of research in that it may shed light on the relationship between non-linguistic mental breakdown and language disruption. If it is always the case that a disruption of one leads to a disruption of the other, then one may conclude that there is an interdependence of thought and language. If, on the other hand, language can remain intact despite mental or psychological disturbance, this argues for a greater independence of language processing from other mental functions." (Fromkin, 1975, p. 498.)

[1] The year is erroneously given in the quote. The reference is: Brown (1973). See References. The original quote is: "I would have to conclude that there is no such thing as schizophrenic speech. I hasten to add that I encountered plenty of schizophrenic thought, but that is another matter." (Brown, 1973, p. 397. Brown later adds. "I do not know how to make a deep distinction between language and thought" (Brown, 1973, p. 398).

Having said this, Fromkin argued that the linguistic traits mentioned by Chaika (1974) also exist in the speech of normals:

> [I]t is clear that normal speakers show performance errors and deviances identical with those of the schizophrenic patient cited in the [Chaika's] paper. (Fromkin, 1975, p. 500.)

… and:

> Chaika is wrong in her claim that "the kinds of errors associated with schizophrenic speech are different from 'normal' performance errors. /…/ Chaika is also incorrect in stating that normal people who make "slips of the tongue" always attempt to correct themselves. While such corrections do occur, a large percentage of errors go undetected by the speaker. (Fromkin, 1975, p. 502.)

Fromkin finally concluded that:

> If the characteristic features singled out by Chaika are unique, then they are unique to the class of *human* speakers. (Fromkin, 1975, p. 503; italics in original.)

Chaika (1974) is further criticized by Lecours & Vanier-Clément (1976), who generally agreed with Fromkin (1975) that the phenomena mentioned by Chaika (1974) also occur in the speech of normal speakers. However, Lecours & Vanier-Clément (1976)—in what is a very elucidating listing of different kinds of aphasias—also stated that "in most cases, one identifies schizophasic and jargonaphasic behaviors rather confidently when witnessing one or the other; and one seldom has major difficulties in distinguishing either from normal speech" (Lecours & Vanier-Clément, 1976, p. 517). Summing up their detailed overview of different aphasic behaviors, they suggested that

> [O]rdinary speakers think and talk standard, that (most) jargonaphasic speakers think standard but talk deviant, that schizophasic speakers think quaint and talk accordingly (Lecours & Vanier-Clément, 1976, p. 516).

Chaika (1977), in a reply to both Fromkin (1975) and to Lecours & Vanier-Clément (1976), pointed out that most of Fromkin's points were misconceived:

> Fromkin (1975, pp. 502–503) said that because some of the deviant language produced by some schizophrenics has a surface similarity to normal slips of the tongue, there is nothing unique about schizophrenic speech. Furthermore, she eschewed error at the level of discourse as the proper domain of linguistics (Fromkin, 1975, p. 501). These assertions are of the utmost importance, for, if Fromkin is correct, much, if not all, linguistic research into speech pathology, including aphasia, is compromised. If an occasional appearance in normal speech is enough to render an utterance type nonaberrant, then there may be no such thing as deviant language. (Chaika, 1977, pp. 464–465.)

Chaika (1977) proceeded to go through Fromkin's examples, one by one, to elucidate how errors in the speech of schizophrenics is in fact different from normal speech errors. Chaika notes that normal errors, at both phonological, morphological and syntactic levels, are "usually rendered transparent by reference to the uttered context" (Chaika, 1977, p. 465), while schizophrenic deviance usually is not. Chaika also highlighted the fact that Fromkin cited Chaika's examples out of context:

> This is a distinction completely overlooked by Fromkin. Her "slips of the tongue" or "speech errors" are all isolated utterances after which the speaker in question ostensibly proceeds normally. In Chaika (1974a), in contrast, the disruption was shown to stretch over an entire discourse. Fromkin failed to note that the utterances she quoted from Chaika (1974a) were embedded into larger wholes with which they were consistent. Thus, she made the schizophrenic data seem more like normal error than they actually are. (Chaika, 1977, p. 467.)

Chaika (1977) also expressed the impression that a difference between schizophrenic and normal speech is that the former results from a "*cycling through levels of errors*" (Chaika, 1977, p. 469; italics in original), and that this combination or errors is what makes schizophrenic speech different from normal speech.

Turning to Lecours & Vanier-Clément (1976), Chaika (1977) contended that she "would revise my [Chaika] original analysis in that I [Chaika] now assume less difference between schizophasia and jargonaphasia than before. However, my assumption remains of a true break in normal language competence of schizophasics" (Chaika, 1977, p. 474).

Maher (1972) reviewed the literature on schizophrenic language, and found that the speech of schizophrenics exhibited a larger degree of part-word and whole-word repetitions than do the speech of normal subjects. However, he also pointed out that many schizophrenic patients do not exhibit any language disturbance whatsoever. Hoffman et al. (1985) reported longer pauses at clause boundaries in the speech of schizophrenics than in the speech of normals.

Clemmer (1980) compared silent and filled pausing in twenty schizophrenics and twenty (matched) normals, who were asked to read out aloud paragraph-long stories and then retell them. He found that schizophrenic speech was marked by longer and more frequent silent pauses within constituents, but also that when the semantic content of the stories did not agree with the commonly held presuppositions, the speech characteristics of normals were similar to that thought disorder of the schizophrenics, thus concluding that the schizophrenic dysfunction was cognitive rather than linguistic.

Kremen et al. (2003) studied phonemic and semantic fluency in 83 schizophrenia patients, and compared them to 15 bipolar disorder patients and 83 normal controls. They observed that both fluency types were impaired in the schizophrenia patients, and that the schizophrenia patients as a whole manifested disproportionate semantic fluency as compared to the controls. They conclude that their study confirms the literature in the observation that schizophrenics exhibit a small degree of fluency impairment as compared to normal speakers.

2.7.3 Schizophrenic speech and brain potentials

It probably comes as no surprise that various neurological studies have been carried out on schizophrenic patients. Some of these have been devoted to speech, and have used both brain potentials and other, related methods. Timsit (1970) showed that RP was similar between a group of schizophrenics and a control group, but that the positive potentials that immediately follow movement were either completely lacking or were very weak in the schizophrenic group. McGuire, Shah & Murray (1993) used single photon emission tomography (SPET) to measure blood flow in the brain of schizophrenics during auditory hallucinations. They found an increase in blood flow in Broca's area during hallucinations as compared to non-hallucinating states. McGuire et al. (1996) used positron emission tomography (PET), and concluded that predisposition to verbal hallucinations is associated with a failure to activate areas known to be implicated in the monitoring of inner speech. Sommer et al. (2001), using

dichotic listening tasks, reported that schizophrenics exhibited less lateralization in a listening test when they were exposed to speech sounds (see also Maddox, 1997; Spence et al., 2000; Crow, 2000). Shergill et al. (2003) used functional magnetic resonance imaging (fMRI) while they varied the rate of inner speech generation. They concluded that in schizophrenic patients who are prone to auditory hallucinations, increased demands on inner speech processing is associated with attenuated activity in brain areas used in verbal self-monitoring.[1] Strandburg et al. (1997) used event-related potentials in a study of linguistic information processing in schizophrenics. They used an idiom-recognition task involving judgments of meaningfulness of idiomatic, literal and nonsense phrases. They found a larger than normal (compared to the controls) N400 component to idioms and literal phrases, but no differences for the nonsense phrases. They concluded that linguistic context, as provided by the first word in two-word idiomatic and literal phrases, is reduced in schizophrenia.

2.7.4 Summary

As we have seen, schizophrenic language lies on the border between several different disciplines, and has been discussed from various perspectives, ranging from the notion of inner speech (in speech production) models to actual covert disfluency behavior. The main interest is of course what schizophrenic speech can tell us about inner monitoring, and interaction between proposed modules in the speech production process.

2.8 Sign language: another mode of *language* production

The underlying assumption of disfluency studies from a speech production perspective is, of course, that disfluencies reveal the deeper processes of *language* production, rather than speech production. Whether or not, at the final stage, the articulators actually are organs such as tongue, velum, lips and the like is of minor, or no, importance, the crucial thing being that it is on-line and real-time, i.e., fast enough to reveal how the linguistic message is produced and processed as it is created. This leads us to the conclusion that to the extent phenomena like stuttering and "normal" disfluencies are *language* production traits, they should show up in other communication modes, as well, provided these are real-time.

An interesting field here is that of sign language. It is reasonable to assume that there are as many sign languages in the world as there are spoken languages, although the language borders of sign languages do not fully correspond to those of spoken languages.[2] If, as it is claimed, pauses (silent or filled), prolongations, truncated words, repetitions, restarts, mispronunciations and so on, in one way or another, reveal deep-lying processes in the language production process, then the said phenomena should also occur in sign languages. On the other hand, if disfluencies, be they stuttered or "normal", are mainly failed executions at the motoric level, then signed disfluencies could perhaps be expected to take on a different guise, since the articulators are not the same.[3]

---

[1] Both McGuire et al. (1996) and Shergill et al. (2003) cite Creutzfeldt, Ojemann & Lettich's (1989) observation that the temporal cortex is modulated by vocalization. This occurs on the order of several hundreds of ms before overt speech, which suggested that its role is associated with the *intention* to speak, anatomically linking areas that generate and perceive speech.

[2] The listing at **http://www.ethnologue.com/show_family.asp** contains 114 deaf sign languages. This, however, is most likely a grave underestimate of the actual number.

[3] Interestingly, it is generally assumed that there are few congenitally deaf stutterers (Starkweather, 1987, p. 243).

Early work by Covington (1964/1973) showed that several kinds of junctures, corresponding to junctures in spoken language, could be found in **American Sign Language** (ASL). Also, mechanisms for "floor-holding" were found.

To further support the validity of sign language as an interesting of object of study, it may be pointed out that McGuire et al. (1997), using positron emission tomography (PET) found that the region of the brain (the left inferior frontal cortex), which is associated with overt speech and silent rehearsal of letter strings in hearing subjects, is also activated during "inner signing", i.e., the covert "speech" of deaf subjects. They concluded that "the left inferior cortex /…/ participates in the generation of language, whether it is covert or overt, spoken or signed" (McGuire et al., 1997, p. 697) and pointed out that this confirms previous studies and observations that signing is markedly impaired by left hemispheric lesions, and that sign aphasia can be induced by inactivation of the left hemisphere using the Wada test.

That the neural organization of signers is not identical to that of speakers was shown in Bavelier et al. (1998a, 1998b). Using functional magnetic resonance imaging (fMRI), they found that in contrast to spoken English, ASL strongly recruited right hemisphere structures, indicating less lateralization in sign language than in spoken language.

Newkirk et al. (1980) studied disfluency in ASL. However, the focus of their study was mainly "slips" at levels corresponding to words or sounds in spoken language and did not discuss things like pauses, prolongation, repetitions (in any detail) and so on. Nevertheless, they concluded that although entire signs were occasionally affected by a slip, far more often "sign-value" errors were produced that most often resulted in non-existing, but *possible*, signs, i.e. they did not violate the structure of ASL.[1]

Grosjean (1979) found pauses in ASL to be as numerous as in English. Grosjean (1980b) reviewed comparative studies in spoken language and ASL and reported that mean pause duration in signing was shorter than in speech. He also pointed out an interesting difference in that the sign concerned remains visible (frozen) during pausing, whereas pausing in speech is the *absence* of sound. He also observed that no systematic study of the equivalents of filled pauses, drawls (prolongations), repeats or false starts had been done. Grosjean & Collins (1979), Grosjean, Grosjean & Lane (1979) and Grosjean (1980a) studied pause distribution in English, and found similarities between English and ASL. Grosjean (1979) found that while speakers tend to breathe at clause boundaries, signers breathed at locations independent of syntactic importance. Grosjean & Lane (1977) differentiated between three kinds of holds (equivalent to unfilled pauses in speech) in ASL: *long holds* that appeared at the end of sentences, *intermediate holds* between conjoined sentences and *short holds* between internal constituents.

Lars Wallin at Stockholm University, Department of Sign Language (personal communication), confirms that—seemingly—the "same" disfluencies do in fact occur. **Unfilled pauses** correspond to halted signs, or signs that are stopped dead mid-sign as it were, and are then continued after a period of a frozen movement. **Filled pauses** correspond to empty gestures, and possibly also to broken eye contact, a floor-holding device employed by signers, thus making it tantamount to the floor-holding function filled pauses are claimed to have in speech. **Prolongations** could be said to have their counterpart in sign language production when part of a sign is reiterated/looped before continued. **Repetitions** of

[1] In English, this would be the difference between e.g. *snobe*, a possible English word which does not exist, and *nsboe*, not a possible English word.

words/signs, one or many, also occur, and so do **restarts**. As for **truncations**, they could be said to correspond to signed movements that are not completed, but instead followed by another sign. Finally, **mispronunciations** also seem to occur, where a sign is erroneously executed. It must be borne in mind that all of these observations/comments must be taken very cautiously as nothing more than preliminary, since only *types* of disfluency are discussed, and not at all frequency or distribution in the language (i.e. sign) stream. No formal studies of disfluencies in sign language in this sense have so far been made, and not until controlled studies have been made on the occurrence of disfluencies in sign language will it be possible to compare the observations with those of spoken (or written/typed) language. However, it goes without saying that the study of sign language from this perspective would be of utter interest from several research points of view in order to gain deeper insight into human language production. Not only would a typological study be interesting, but even *if* the same categories/types of disfluencies do in fact occur in sign language, the frequencies—both absolute and relative—might be different, as might their distribution in the utterances.

2.9 Application-driven approaches

So far, we have only discussed disfluency from an exclusively *human* perspective. With the inclusion of speech interfaces in computers, disfluency processing has become more and more interesting from the point of view of automatic applications. Although one could easily argue that speech synthesis would benefit from the inclusion of disfluency insofar as disfluency helps listeners comprehend speech, the main interest so far has come from the other end of the speech chain, in automatic speech recognition (ASR) and tagging/parsing of the recognized strings.

Automatized systems, using ASR and text-to-speech (TTS) or speech synthesis systems, are becoming household events these days. However, with a couple of notable exceptions, disfluency handling is at best rudimentary in most such systems, and given the frequency with which human speech is disfluent, it goes without saying that automatic handling of disfluencies could improve the performance of ASR systems, or at least the perceived naturalness of such systems if they allow the same type of input that is possible in human–human interaction. Much recent research on human disfluency has also been prompted by technological requirements, i.e., in order to include phenomena typical of spontaneous, human, speech, one needs to know what spontaneous, human, speech looks like. For example, the previously referred work by Lickley & Bard (1996, 1998b) explicitly mention ASR systems as one of the underlying rationales for carrying out the studies.

2.9.1 Disfluency in automatic speech recognition

The first, immediate, problem facing an automatic system exposed to human speech input is recognizing the acoustic waveforms into speech, *including* whatever disfluency might be present. While recognizers have been able to (decently well) handle "ideal" linguistic input for quite some time now, disfluencies have been posing a problem to such systems, and still do. Although Lickley, McKelvie & Bard (1999) reported that disfluencies adversively affect both human and machine comprehension of speech input, one could safely assume that the problem to date is more acute in machine speech recognition. An alternative, or complementary, view is given by Siu & Ostendorf (1996), who claimed that disfluencies provide valuable information that could indeed be used by a system for more accurate language modeling.

Early work was carried out by O'Shaughnessy (1992a, 1992b, 1992c, 1993, 1994, 1999), on detection and correction of false starts and hesitations. Gabrea & O'Shaughnessy (2000) described automatic detection of filled pauses using a combination of duration, fundamental frequency and spectral information.

Extensive work on automatic disfluency detection has been carried out by Shriberg, Stolcke and colleagues (e.g. Shriberg & Stolcke, 2004; Shriberg, Stolcke & Baron, 2001; Stolcke & Shriberg 1996; Shriberg, Bates & Stolcke, 1996, 1997; Stolcke et al., 1998; Shriberg et al. 1996). Shriberg, Bear & Dowding (1992) and Bear, Dowding & Shriberg (1992) described automatic detection of repairs by integrating knowledge from various sources, such as pattern matching, syntactic and semantic analysis, as well as acoustic information. Liu, Shriberg & Stolcke (2003) found that automatic detection of disfluency interruption points was best achieved by a combination of prosodic, word-based, and part-of-speech-based cues.

Levow (1998) found that one could identify speech corrections by using acoustic-prosodic cues such as duration, pause and pitch variability.

Spilker, Klarner & Görz (2000) reported a system that detects and correct speech input integrating information at acoustic, lexical, syntactic and semantic levels. An acoustic module generates hypotheses about potential repairs, a stochastic model then suggests corrections, and a lattice parser then decides what suggested correction should be used.

Opperman et al. (2001) discuss the problem of *off-talk*, i.e. speech that is not directed to the system. They found a significantly higher percentage of filled pauses in off-talk than in system-directed speech. Consequently, they conclude that filled pauses could be used as an indication that the utterance in question is not directly meant as an instruction, question or feedback to the system.

More recently, it has also been shown that (one-syllable, high-frequency) function words are less reduced and longer in duration when they precede or follow a disfluency, such as a filled pause (Alan Bell et al., 2003).

2.9.2 Disfluency in automatic tagging and parsing

Assuming that ASR is no longer posing a problem to the handling of spontaneous speech input, there is still the problem of linguistic analysis. Given that grammars of languages from days of yore have been written from a text perspective, more or less tantamount to Chomsky's competence, instead of language as human interlocutors use it, there are vast lacunae in our knowledge of what language *actually* looks like, at least from a syntactic perspective. It has been argued that human speaker–listeners pay less attention to "ideal" grammars than to communicative goals.[1] So, then, what does actual human linguistic input (to a system) look like, and how do we handle it?

To a language component of a system, there is the need to assign words with word classes as well as with grammatical information. One also needs to make some kind of semantic analysis, look for communicative incoherencies, or repairs, and try to arrive at a conclusion as to what the intended message (from the speaker's perspective) might be, irrespective of how "cluttered" the message might appear the way it is delivered.

[1] E.g. by Debaisieux & Deulofeu (2001; title of paper) who ask "[g]rammatically unacceptable utterances are communicatively accepted by native speakers, why are they?".

Hindle (1983) was among the first to consider speech disfluency from a technical-computational perspective. He presented an implementation of a deterministic parser capable of dealing with speech disfluency. However, Hindle assumed there is a phonetically identifiable editing signal present in the speech input, something that has later proved to be a not-so-common phenomenon as he assumed was the case.

Bear, Dowding & Shriberg (1992) used a simple pattern matching technique for repair detection. Unlike Hindle (1983), they did not assume explicit cues in the acoustic input. They also noticed that "cue words" like "no", that may or may not be error-signaling words, could be distinguished as to function (error signal or not) using prosodic information alone. Oviatt (1995) reported that only 5% of all spoken disfluencies in her material were marked with an explicit editing signal, and Eklund & Shriberg (1998) also noted that explicit edit signals were not common.[1]

Nakatani & Hirschberg (1993) also used acoustic information to detect the interruption point in speech repairs. They found that inter-word pause duration and word fragments were useful cues, among others. Nakatani & Hirschberg (1994) used acoustic and prosodic cues to identify speech repairs, and reported a precision rate of 91% and recall of 86% on a corpus.

Kikui & Morimoto (1994) presented a similarity-based algorithm that identifies the onset of the reparandum in tagged utterances, and report a 92% success rates.

Heeman, Allen and Loken-Kim have done extensive work on repair detection (Heeman, 1997; Heeman & Allen, 1994a, 1994b, 1997, 1999; Heeman & Loken-Kim, 1995, 1999; Heeman, Loken-Kim & Allen, 1996; see also Hirst et al., 1994). In the early work (Heeman & Allen, 1994a) they build the repair pattern "on the fly", without using prosodic information, although it was argued that prosody could well be incorporated into their model (Heeman & Allen, 1994b). In later work (Heeman & Allen, 1997, 1999), they also presented a system that not only detects intonational phrases, but also includes discourse marking. They argued that by identifying parts-of-speech, discourse markers, speech repairs and intonational phrases simultaneously, they obtain better results than solving each of these tasks independently.

In the same vein, Core and Schubert presented a model for the handling of speech repairs (Core, 1996, 1999; Core & Schubert, 1997, 1998, 1999a, 1999b, 1999c). In Core (1996), simple modifications to a chart parser are presented, to enable handling of hesitation and repairs. Overlapping speech, however, is not handled. In Core & Schubert (1997, 1998) the notion of *metarules* was introduced, that allow interpretation of overlapping speech in dialogues, by having the metarules allow syntactically separate constituents to interleave or straddle each other. Repairs are handled by building parallel phrase structure trees that separate the reparandum and the reparans (alteration). In Core & Schubert (1999c), they discussed extending their work to include prosodic information.

Also, Levow (2002) studied differences between original utterances and their corrected forms as a response to speech recognition failure, and found that pause durations increased in the corrections, which means that user corrections diverge even more from the recognizer's underlying model.

[1] It should be pointed out that the frequency is highly affected by whether or not one includes filled pauses in what counts as an edit signal, so figures are not completely comparable across corpora.

2.9.3 Designing dialogue systems

While all the work referred to in the previous two paragraphs has focused on the detection and "cleaning up" of disfluencies, there are alternative views on how systems could adapt to the up-and-coming phenomenon of human–machine applications. One alternative view is that instead of trying to achieve as high a detection (and repair) rate as possible in the functionings of the system, one could try to minimize disfluency input from the users by designing the system so that the (human) users become as fluent as possible.

Oviatt (1995) observed that not only were humans less disfluent when addressing an automatic system than when speaking with other human beings, their disfluency was associated with two parameters: utterance length and presentation structure in the system. Consequently, by designing a dialogue system so that it encourages short sentences from the user, and by presenting the information in a "structured" way, disfluency rates should drop considerably in the speech of the users.

Bell, Eklund & Gustafson (2000) compared disfluency rates in a unimodal, telephone corpus and a multimodal one (Bell et al., 2000). They found that disfluency rates were overall higher in the telephone corpus. However, while disfluency rates were highly speech act dependent in the telephone corpus (even more so than utterance length; compare Oviatt above), this was not the case in the unimodal corpus. That speech disfluency is dependent on the channel was observed as early as in Kasl & Mahl (1965), who reported that subjects were much more disfluent in a telephone-like setting than in a face-to-face setting.

Asp & Decker (2001a, 2001b) carried out work on a subset of the same data set that is studied in this thesis, and proposed a method to reduce disfluency in the speech of the users by careful design of the dialogue moves of the system.

Oviatt, MacEahern & Levow (1998) noticed that disfluency rates went down as a response to recognition errors, as a side effect of human users' adaptation to the system (which also included hyperarticulation and changes in fundamental frequency).

Moreover, since more and more systems will probably be targeted to children, given that children are notoriously disfluent in general—known in stuttering research, and repeated in a human–machine interaction study (Oviatt, 2000)—knowledge about the disfluency of children (already present in the literature) in general, and children's interaction with automatic systems in particular, is of importance in the design process. Finally, Narayanan & Potamianos (2002) analyzed conversational interfaces for children, and included disfluency in their study. They specifically studied false starts, mispronunciations and filled pauses. They concluded that:

> Although disfluencies and hesitation phenomena occur more frequently in children than in adults, our experience showed that ASR performance does not suffer significantly due to these effects, hence requiring no special or acoustic modeling strategies. (Narayanan & Potamianos, 2002, p. 72.)

This obviously runs counter to some of the opinions expressed above, but could perhaps be regarded as a source of relief for developers of dialogue systems. Whether or not the results of Narayanan & Potamianos (2002) hold true will have to be confirmed or rebutted in future studies, but without doubt the speech produced by children will be increasingly studied given the increase of domains where speech-based applications are introduced—some of which with

children as the main target—and the associated development of recognizers tuned to include the speech of children.

2.9.4 Summary

So, how are we to regard the application perspective? Should automatic systems "clean up" the incoming data and make (educated) guesses as to the intended message? Or should automatic system include everything that is being said, and try to make use of it? It boils down to whether or not one regards human–machine communication as tantamount to human–human communication. This, perhaps, is a sliding scale. A very constrained domain with very specific goals might be different from a more open-ended dialogue system, where linguistic form and content are less predictable. While it could be argued that Clark, Allwood and others are right in pointing out that human communication is profoundly collaborative (or cooperative), and that descriptions such as OCM is what comes closest to an accurate formalization of human communication, one could perhaps claim that there is little need for an automatic system to be aware of the types and tokens of "editing" that has occurred, no more than readers of a newspaper would benefit from knowing all the previous, corrected, versions of the newspapers articles they read. Automatic systems need not be aware of the "mental status" of the people addressing them (indeed, many users would probably prefer them *not* to be!). Such a view would, seemingly, justify a "detect-and-dispose-of" approach of application-driven research within a field.

However, a system that *would* be able to make inferences from overt (or perhaps even covert) editing the way humans obviously do, would of course in a sense be a "smarter", more human-like, system, and for that reason most likely also more successful given a communicative task. This would probably lead to more attractive systems, even from a commercial point of view, since human users would need to adapt less to the system, at least no more than humans need to adapt to different human interlocutors, given various communicative or contextual circumstances.

Also, there might be a difference between what is of short-term interest, and what is of long-term interest. Perhaps there is a current need to clean up messages to be able to launch automatic services that perform well enough to make them commercially interesting, while the said service could then be "boot-strapped" further down the road along the lines advocated by Clark, Allwood and others who stress the communicative functions of disfluency.

2.10 Disfluency in a nonnative language

Although "words" like filled pauses are among the commonest words in any language (*vid.* Shillcock et al., 2001), language education does not normally "teach" them, so even if a speaker of a foreign language obtains a very high level of fluency, and perhaps even a near-native accent, the disfluencies will reveal a foreign origin. So when it comes to production of disfluencies, given their informative role, one can safely conclude that well nigh nothing is done, and that the attitude is to teach "fluent" language, i.e., something similar to edited text.

So, what about *perception* of disfluencies? Do disfluencies pose a problem to a foreign listener? Once again, little work has been done within the field. Voss (1979) found that German speakers of English indeed had perceptual problems as a function of the degree of disfluency in English stimuli.

Voss pointed out that:

> From the point of view of speech perception, the native speaker as listener will, on the whole, have no difficulties with them [disfluencies]. In fact, similar to slips of the tongue he will often not even notice their occurrence. (Voss, 1979, p. 130.)

In order to study whether disfluencies per se would cause misinterpretations, Voss had 22 German students transcribe a stretch of spontaneous English speech. Voss then analyzed transcription errors while looking at the occurrence of repeats, false starts, filled and unfilled pauses. Five of the students tried to capture the disfluencies by including *ahm*, *ah* and so on in their transcriptions, but repetitions "were hardly ever written out" (Voss, 1979, p. 132).

Voss concluded that nearly a third of the transcription errors are likely to have been caused by a misinterpretation of hesitation phenomena (excluding unfilled pauses). He divided them into two groups:

1. Cases where repeats or filled pauses have been mistaken for words, or parts of words.

2. Cases where words (or parts of words) have been mistaken for disfluencies, and consequently been filtered out (*vid.* Bond, 1973; Laver, 1970).

Voss summed up the study concluding that:

> [H]esitations present a major perception difficulty for the non-native speaker confronted with spontaneous speech. More attention to this area in language teaching, e.g. in the form of more exposure to genuine, spontaneous speech, should help remove or at least reduce a considerable source of perceptual problems for the non-native speaker." (Voss, 1979, p. 138.)

Raupach (1980) compared pausing patterns in native language (L1) and foreign language (L2) in German and French speakers, and found that "[t]he hypothesis that speakers are likely to transfer idiosyncratic performance from L1 to L2 find support /…/ except in the use of filled pauses (Raupauch, 1980, p. 267).

Deschamps (1980) studied syntactical distribution of silent and filled pauses, as well as "drawls" (prolongations) in English as a second language, spoken by French students. He found, among other things, that the speakers did not increase the length of silent pauses, but rather increased the number of pauses.

2.11 Disfluency and bilingualism

If disfluencies are part of the communicative competence and performance of speakers, and if disfluencies are not exactly the same across languages, what kind of disfluency behavior do we encounter among bilingual speakers?

Dale (1977) studied four bilingual Cuban-American male adolescents who were "dysfluent" in Spanish only, while being able to speak English fluently. The boys were all born in the United States in Spanish-only speaking homes, and they all reported that they had begun to "stutter" in Spanish at around the age of 12, while maintaining full proficiency in English. She concluded that the pressure put on them by their parents to speak Spanish perfectly

exposed them to too much negative reinforcement, something which was lacking in the English-speaking environment, and that that made their Spanish deteriorate.

Ratner & Benitez (1985) analyzed a bilingual (Spanish and English) stutterer. They concluded that syntax was a more important determinant than phonology to explain the frequency and location of fluency breakdown. For instance, the most common constituent-initial errors were made on introductory noun phrases, which are more common in English than in Spanish. On the other hand, since Spanish allows subject-dropping, thus moving the verb-phrase to an utterance-initial position, there was a tendency to stutter more on verb phrases in Spanish than in English. As for phonology, 70% of the errors in Spanish were made on word-initial vowels, while only 38.9% of the errors in English were made on word-initial vowels.

Rieger (2003) investigated hesitation strategies of learners of German as a second language, and found that the more advanced the speakers were, the more complex were the hesitation strategies employed, i.e. "the students who perform best on a linguistic or grammatical level also perform best on conversational or discourse level" (Rieger, 2003, p. 44).

Thus, it would seem that a number of factors are at play when speaking a second language, that both concern the level of proficiency in the weaker language, but also purely linguistic factors such as typological issues (Ratner & Benitez, 1985). Disfluencies are typically not taught in language education, despite their being among the most common "words" in a given language, but the results of Rieger (2003) seem to indicate that they might become a part of the language once the level of proficiency increases.

2.12 Crosslingual studies

While some disfluency studies have been done on more than one language, a comparatively small number of studies have had cross-lingual comparison as the explicit objective. It goes without saying that cross-linguistic studies may reveal underlying functioning, as well as universal traits, of disfluency production. Indeed, Dechert (1980) suggested a speech production model with cross-linguistic observations as the basis.

Donald G. MacKay (1970) looked at data from German, English, Latin, French, Greek and Croatian, and concluded that the "phoneme repetition effect" (in spoonerisms) was language independent, and may reflect a universal mechanism.

Grosjean & Deschamps (1972, 1975) found that the pause–time ratio in French and English was almost identical, but that the distribution differed in that there were fewer but longer pauses in French, and more and shorter pauses in English.

Faure (1980) compared French and German, and concluded that the previous observation (for English) that pauses at the beginning of a phrase tend to occur either before the first or second item also seems to hold for German. He also concluded that filled pauses seemed to be a idiosyncratic feature of the speaker in both French and German.

Fox, Hayashi & Jasperson (1996), comparing repairs in English and Japanese, found that there were differences between the two languages at the morphological level. In the Japanese data, they found that the speaker went back and repeated (changed) only the inflectional verb ending (a bound morpheme) of a word, something they did not observe in the English data. They argued that this could be dependent on the fact that verb suffixes in Japanese are not

agreement markers, which means that they do not refer back to the subject of the verb, which is the case in English, which possibly makes it easier to change the ending only.

Eklund & Shriberg (1998) compared Swedish and American English human–human and human–machine corpora of spontaneous speech within the same domain (air travel information). However similar as to dyad characteristics and domain, there were some differences as to mode (the Swedish corpora were all telephone conversation, while one of the American English corpora was a push-to-talk conversation), as well as task details (pictorial instructions for Swedish, actual travel plans for American English, and so on). They concluded that, on the whole, disfluency is similar in the two languages, both as to type and distribution. There were some minor differences, however, such as the occurrence of filled pauses inside compound words in Swedish,[1] something that was not observed in American English.

Tseng (2000) compared Mandarin Chinese and German spontaneous speech and found that while German words and Chinese characters seem to play a similar role in speech repairs, it was more common for noun phrases (NP) to be repaired directly within the NP, while Chinese repairs were often composed of more than one phrasal category.

Eklund (2000a) compared Swedish and Tok Pisin human–human air travel authentic dialogues (Eklund, 2000b), and observed that the two languages exhibited similar traits on the macro level, but dissimilarities on the micro level. For instance, unfilled pauses are commonly found inside lexical roots in Swedish,[2] but were not found in the Tok Pisin data. A truncated word was never continued, but always restarted. Moreover, segment prolongation ratios as to phone-position, which was 30–20–50 for initial, medial and final phone in a word in Swedish,[3] proved to be 15–0–85 in Tok Pisin. Eklund (2001), took a closer look at prolongation in the two languages, and reported that different segments were prolonged in the two languages, but "for the same reason" (Eklund, 2001, p. 7), in that the speakers hesitated at the same places, which meant that the hesitation (prolongation) affected the segments that occurred in those positions in the phrase, e.g. the final segment of prepositions meaning "to" or "from".

Comparable to the Tok Pisin figures, Den (2003) report a 10–5–85 ratio for Japanese, and Lee et al. (submitted for publication) report a 4–1–95 ratio for Mandarin.

Eklund (2001) summed up his study by saying that the observations made for Swedish "probably do not hold for all languages, and that more cross-linguistic studies /…/ need to be done in order to gain deeper insights with regard to the role and function /…/ in speech production" (Eklund, 2001, p.8). I fully agree with Eklund on that point, and look forward to more cross-linguistic studies of disfluency in the future.

[1] Filled pauses inside lexical compounds have also been reported for German (Lüngen et al., 1996; Althoff et al., 1996; Althoff, 1997).

[2] Unfilled pauses inside lexical roots have also been observed in German (Lüngen et al., 1996; Althoff et al., 1996; Althoff, 1997) and Tagalog (Rubino, 1998).

[3] The same ratio was observed for American English by Eklund & Shriberg (1998).

2.13 Disfluency and gestures

It is well known that humans not only communicate by means of spoken language, but also employ gestures, both facial and bodily. It has also been known since long that body movement and speech rhythm are related, and that movement also closely follows disfluencies (e.g. Dittman & Llewellyn 1969). That gestures are fundamental in human communication is suggested by the fact that congenitally blind speakers gesture, even when they speak to other blind listeners (Iverson & Goldin-Meadow, 1998). Furthermore, Kelly (2003) studied ERPs in subjects who watched video segments of people who were speaking and gesturing. He found that when exposed to mismatching speech and signs—e.g. saying the word tall while gesturing to a short, wide object—the well-known N400 effect appeared, known from experiments when subjects are exposed to semantically anomalous speech (e.g. Kutas & Hillyard, 1980).

It has been suggested that gestural motor activity and speech production are linked neurologically (Rimé & Schiaratura, 1991; Feyereisen & de Lannoy, 1991). It has been shown that hand gestures tend to co-occur with speech hesitation and pausing (c.f. Ragsdale & Silvia, 1982), and the same observations have been made for head movements, that also tend to occur with speech disfluency (e.g. Hadar, Steiner & Rose, 1984).

Kendon (1972) showed that in those cases where gesture occurred with speech, the gesture preceded speech. He also noted that "[t]he larger the speech unit, the earlier and more extensive are the preparatory movements" (Kendon, 1972, p. 205).

Butterworth & Beattie (1978) also reported that gestures preceded speech. They concluded that "[g]estures are products of lexical preplanning processes" (Butterworth & Beattie, 1978, p. 358). Their explanation as to why gestures occur earlier than speech is that the mental lexicon (of words) probably contains up to 30,000 words, while the set of gestures must be much smaller, which in turn means that the gesture selection process is much faster.

Turning to speech disfluency, Seyfeddinipur & Kita (2001) found that halted gestures preceded halted speech, indicating that speech errors are detected *before* speech is stopped. On average, gestures stopped 240 ms before speech stopped.

Esposito, Duncan & Quek (2002) studied holds in gestures and speech in American English and Italian. They found a 28% overlap between gesture holds and speech holds (including prolongation of segments) for Italian and a 45% overlap for American English. They also noted that the Italian speaker used more gestures, more filled pauses and fewer silent pauses than did the American speaker.

Finlayson et al. (2003) examined disfluency rates in spontaneous speech in three different experimental settings: one hands free, one with one hand immobilized and with both hands immobilized. They found that as gestures were restricted, disfluency rates went up, thus supporting a link between speech and gesture production.

Given that future automatic systems might well include visual information as well as acoustic information (not only for on-line interfacing and translation of sign language), it is of interest to study in what way disfluency is apparent in gesturing for inclusion in such systems.[1]

2.14 Disfluency in writing

Although the notion of disfluency is most often discussed as a feature or characteristic of speech, all other kinds of motor action may be more or less fluent. One such mode of linguistic communication is writing, and the question that presents itself is whether disfluency can be found in writing, as well, and if so, whether or not it is similar or dissimilar to speech disfluency.[2] That writing should not simply be regarded as just a different surface form of language produced is suggested by the observation that speech and writing *can* be residing in different hemispheres, as is shown in a split-brain patient whose left hemisphere could read and speak out items aloud, but not write them, but whose right hemisphere could write items, but not read them out aloud. (Strauss, 1998). However, it is hard to extrapolate from just one patient, of course, but it is still suggestive that such functioning at all may occur.

Similar to their studies on manipulative factors in speech production, Blass & Siegman (1975) compared three different channels: speech, dictation and writing, finding amongst other things that the writing condition contained the smallest amount of silences (unfilled pauses). The explanation would be that writing takes much longer time to execute than speech, so when a person is writing a word he has the time to formulate the next one, which means less interruption in the planning of communication.

Oviatt (1995), comparing writing to speech, found that content corrections were more common in writing than in speech, although no overall frequency differences between speech and writing were found, in contrast to Hotopf's (1983) claim that disfluencies are more common in writing.

Wengelin (2001; see also 2002) performed typing studies in three groups: ten (normal) university students, eleven dyslexic adults, and nine congenitally deaf subjects. She observed that the number of pauses alone were more frequent than all disfluencies taken together in reports from studies of spoken language, and that the dyslexics produced more pauses than the other two groups combined. An interesting observation was that the congenitally deaf subjects exhibited something that could be interpreted as a "filled pause" in typing. Deaf people are used to communicate in real-time and on-line in a text telephone setting, and while writing they occasionally paused by writing a sequence of full stops/periods instead of just

[1] The role gestures play in human communication is of course more far-reaching than can be covered here. One outstanding issue is in what way brain lateralization, handedness, gestures and speech are related (or "what came first"), i.e., the origin of language and speech. For a recent debate, see the open peer article by Corballis (2003a/2003b), with critical comments by Annett (2003), Arbib (2003), Arcadi (2003), Armstrong (2003), Beaton (2003), Bradshaw (2003), Breitenstein et al. (2003), Code (2003), Cook (2003), Corbetta (2003), Dale, Richardson & Owren (2003), Dickins (2003), Faurie & Raymond (2003), Feyereisen (2003), Fouts & Waters (2003), Gillett (2003), Holloway (2003), Hopkins & Cantalupo (2003), Iverson & Thelen (2003), Johnson-Frey (2003), Jones & Martin (2003), Josse & Tzourio-Mazoyer (2003), Jürgens (2003), Kelly (2003), Knight (2003), Leavens (2003), MacNeilage (2003), Michel (2003), Pearce (2003), Pedersen & Vereijken (2003), Raz & Donchin (2003), Rönnqvist (2003), Sommer & Kahn (2003), Walker (2003), Woll & Sieratzki (2003) and Wolpert (2003). Frost et al. (1999), in a recent study, showed that language is strongly lateralized in the brain in both sexes. See also Eling (1986) on a discussion on handedness and lateralization.

[2] For an exhaustive (and cautious) analysis of the (alleged) differences between speech and writing, the reader is referred to Hotopf (1983).

stopping. This should perhaps be seen in the light of the proposal that filled pauses in spoken language are seen as a commitment signal, or floor-holder, indicating the speaker is not done speaking yet, and it is not far-fetched to view the production of "…" in a text telephone channel as a means serving the same purpose.

I will not delve deeper into writing, or typing, here, but once again it seems that corroboration of proposed speech (or language, really) production models can be obtained from other fields, as was previously suggested for e.g. sign language or the hallucination or speech of schizophrenics.

2.15 Disfluency as a paralinguistic segregate?

In two classic studies, Trager (1958, 1964) divided communication into three different parts: language, paralanguage and kinesics. Paralanguage was further divided into voice qualities and vocalizations, the latter including sounds like *ah*, *er* and *uh-huh*, i.e. what is called interjections or filled pauses (etc.) in the literature. Levin & Silverman (1965) set out to describe the incidence of a group of paralinguistic variables such as **vocal aggregates** such as *uh*, *er*, *um*, as well as sentence corrections, sentence incompletions, repetitions, slips, omissions, parenthetical remarks (words like *well*, *oh*, *see*), **zero aggregates** (unfilled pauses) and "drawls" (segment prolongations), correlating these with speaking situation and personality characteristics. They tested 48 children in two speaking situations, either speaking to an audience, or to a microphone as no one was listening and concluded that "deliberate hesitations" were predictable for boys as a function of the personality characteristic exhibitionism, while stressful hesitation was responsive to speaking situation. While certain authors, e.g. Clark & Fox Tree (2002) argue that *uh* and *uhm* should be regarded as conventional English words, it is obvious that some disfluencies lend themselves easily to Trager's paralanguages scheme, i.e., vocalized behavior with a signaling function.

2.16 Disfluency among the elderly

We have seen that a large number of studies have been devoted to children, mainly for the purpose of diagnosing early stuttering, but also from an application-driven angle, given that an increasing number of interfaces of automatic services, or toys, are directed towards children. We also know that speech continues to change throughout life, so it is only natural to ask whether disfluency is different in the speech of the elderly, Comparatively little research has been done here, but let me just mention a couple of studies, lest this large demographic group be entirely forgotten.

Yairi & Clifton (1972) compared disfluency in three groups, preschool children, high school seniors and geriatric persons. They found that high school seniors were significantly less disfluent than preschool children and geriatric persons, and that these two latter groups were similar as to disfluency rates. This suggests that disfluency rates go down during early adulthood (which has been proposed from early on), but then rises again later on in life. Yairi & Clifton (1972), speaking from a stuttering perspective, also noted that the type of disfluency exhibited by geriatrics were typical of nonstutterers, and was characterized mainly by a large number of interjections (filled pauses) and revision-incomplete phrases.

Kemper (1992) reached similar results. She studied sentence fragments in two groups of elderly subjects, a "young-old" group of 60 to 74 years of age, and an "old-old" group of 75

to 90 years of age. She observed no general increase in the number of sentence fragments with age, but while the young-old group was more prone to produce false starts, the old-old group showed more filled pauses. Kemper (1992) also reviewed the literature on the speech fluency of elderly people, and summarized previous findings thus:

> Older adults are generally found to be less fluent than young adults with regard to slower rates of speaking; increased speech errors, such as stuttering and stammering; longer and more frequent hesitations and pauses; increased speech fillers, such as *well* and *you know*; more ambiguities of reference; and increased revisions, paraphrasing and redundancy (Kemper, 1992, p. 444)

Kemper argued that her own findings were linked to syntactic complexity, like the production of complex sentences. Both filled pauses and false starts were more common in embedded clauses, for example.

Leeper & Culatta (1995) studied the relationship between speech rate and speech fluency in 78 elderly speakers (ages 55–92 years), but found only few significant effects between the variables.[1] They concluded that "most old speakers in normal health produce speech that is within accepted standards of normal fluency" (Leeper & Culatta, 1995, p. 11).

Of course, this presentation is not in any way exhaustive as to how the speech of elderly exhibits special characteristics. However, the main point has been to show that one can not take for granted that the speech of elderly looks exactly like the speech of younger subjects, who have been devoted enormous amounts of research over the years, and that one needs to study elderly speech in more detail to establish what possible, unique, characteristics it might demonstrate.

2.17 Effects of disfluency

Given the veritable cornucopia of disfluency research, but also given the consistency in the typologies suggested within the wide variety of field within which disfluencies have been studied, one obvious question that has not been dealt with so far is: *do disfluencies matter?* To be more specific, do disfluencies affect the listener in any way, either concerning the linguistic content or comprehensibility of the speech, or concerning listeners' reaction to the speaker along other dimensions, such as credibility, seriousness, competence, trustworthiness, intelligence and so on and so forth. Several studies have addressed these questions—once again prompted by stuttering research—as one of the main problems for the stutterer has been fear for given a bad impression on the listener. However, similar such studies have also been devoted to the speech of nonstutterers, and I will summarize some of the results from those latter studies in this section. Although most of these studies have treated both linguistic content as well as listener rating concerning extralinguistic, or psychological, traits attributed to the speaker, I will separate these in the following paragraphs.

[1] It is striking that Leeper & Culatta, as late as in 1995, both refer to and use Johnson's categories from 1961. Moreover, Franklin Silverman (1995) published a paper with the title "Can Disfluencies Be Categorized Reliably Using Wendell Johnson's Scheme", and concluded that "I [Silverman] certainly would not claim that the procedure we used [Johnson's scheme] for classifying instances of disfluency was completely error free. /…/ I would claim, however, that it has been of considerable heuristic value. Much of what we have learned about the disfluency behavior of stutterers and nonstutterers during the past 50 years has been from studies that employed Johnson's scheme." (Silverman, 1995, p. 586.)

2.17.1 … as to extralinguistic factors

Miller & Hewgill (1964) studied audience ratings of speaker competence, trustworthiness and dynamism for two types of disfluency: *vocalized pause* (i.e. filled pause) and *repetition*. They found that speakers who produced fluent speech were rated as more competent and dynamic than people who produced nonfluent speech. This effect was more marked for repetition than for filled pauses. Trustworthiness was not seriously affected by disfluency.

Sereno & Hawkins (1967) studied audience reactions according to the same parameters as Miller & Hewgill (1964), but included slips of the tongue (i.e. mispronunciations) and repeated phonemes in the material. They obtained the same results: audiences rated speakers as less competent and dynamic as a function of disfluency, but only a slight effect on trustworthiness ratings. They also studied audience reactions concerning their attitudes towards the topic discussed in the speech data, in this case a speech favoring Black Muslims, but found no attitude changes towards the topic per se as function of disfluency, i.e. whether or not the speech contained disfluencies did not affect any changes towards Black Muslims in the audience. Thus, it seems as if disfluency more affects audience reactions to the speaker than to the topic.

McCroskey & Mehrley (1969) studied audience attitude changes and ratings of source credibility as a function of disfluency (filled pauses and repetitions) and organization of a speech (sentences shuffled around as compared to the original). They found that both disorganization and disfluency led to a less convincing speech (i.e., less attitude change) and less source credibility. The most detrimental effect was when speech that was both disorganized and disfluent was presented to the audience.

Duffy, Hunt Jr. & Giolas (1975) studied the effect of disfluency on information transfer, attitude and ratings of the speaker's competence, trustworthiness, dynamism and delivery. The included disfluencies were broken (truncated) words, repetitions, prolongations and interjections (i.e. filled pauses). They found that disfluencies negatively affected audience ratings of speaker competence, dynamism and delivery. Overall, type of disfluency made no difference to the ratings.

Christenfeld (1995) studied the effect of filled and unfilled pauses on listeners' perception of the speaker. The material consisted of an authentic speech sample (a radio talk show caller) in three different versions: In the first, no changes were made, leaving all filled pauses untampered with. In a second version, the filled pauses were replaced with silent pauses of the same duration. In the third, filled pauses were removed, making the tape nine seconds shorter. Both the edited versions sounded perfectly natural. Subjects listened to the different versions and were asked to rate the speaker on 15 adjectives (intelligent, comfortable, educated, interesting, competent and so on) using a five-point Likert scale. On a second page, they were asked to estimate the number of filled pauses. In one condition, the subjects were asked to attend to the content of the conversation, in another conditions they were told to focus on the style of the speech. There was also a control condition, where no instruction was given. The first observation is that when subjects attend to style, they are conscious of filled pauses, but when they attend to content they are not aware of any *ums* occurring. On a rhetorical eloquence scale, no pauses gave the best impression, but the use of filled pauses created a better impression than did silent pauses. However, on an anxiety scale, although filled pauses were regarded as a sign of anxiety, the tapes with filled pauses were perceived as relaxed as the perfectly fluent tapes, while the tapes with silent pauses were considered more anxious.

Consequently, Christenfeld (1995) suggested that speakers, in order to give a relaxed impression, should learn to say *um* instead of being silent when confronted with difficult choices.

Susca & Healey (2002) found that although listeners attend to fluency in the speech signal, one has to weigh in a number of other, paralinguistic, factors as well, and conclude that it is difficult to assess the effect of disfluency as separated from other features of the speech. However, their listeners were more concerned about thought organization than vocabulary issues in the speaker in disfluent speech samples than they were when they listened to fluent samples, proving that disfluency affects listener attitudes to speech.

2.17.2 … as to linguistic content

As we saw in the previous paragraph, Duffy, Hunt Jr. & Giolas (1975) found that disfluency affected ratings of speaker competence, dynamism and delivery, irrespective of disfluency type. However, they found no effect on information transfer (audience recall), listener attitude towards the topic or the trustworthiness of the speaker. They concluded "that disfluency does not affect the information transmitted in a verbal message but that it can negatively influence the listener's evaluation of the style of delivery and the competence of the speaker" (Duffy, Hunt Jr. & Giolas, 1979, p. 112).

Hulit (1976) studied the effect of prolongation and "double-unit phoneme repetition" on listener comprehension of a passage. Both types of disfluency were used on "key words" (nouns and adjectives) and "lesser words" (closed word classes). He found that both types of disfluency adversely affected comprehension of the passage. Prolongation on key words was slightly less detrimental, but had still a negative effect. However, it must be noted that the disfluencies were simulated and possibly do not reflect authentic disfluency.

Fox Tree (1995) used a monitoring test to study the effect of false starts and repetitions on comprehension speed of words in a stretch of spontaneous speech in English and Dutch. She found that, for both languages, false starts were detrimental to speech comprehension (i.e., slowed down reaction times) while repetitions were not detrimental, and even exhibited a tendency to be beneficial, i.e. speed up reaction times. Fox Tree's explanation is that while false starts force the listener to rebuild the parse tree, repetitions do not.

As has already been referred to, Lickley & Bard (1996) found that words were harder to recognize in disfluent utterances than in fluent utterances, showing a clear detrimental effect of disfluency on speech comprehension.

Likewise, Bortfeld et al. (1999) found that comprehension was faster with filled disfluencies, than when disfluencies were excised and replaced with silences of the same duration.

2.17.3 How we do not notice disfluencies

It has been known for a long time that speech perception is "creative", as it were, and that there is no one-to-one correspondence between acoustic-linguistic stimulus and perceived-interpreted item. Helfrich (1980) goes so far as to stating that there "is no doubt that both digital and analogue-acoustic [automatic pause extraction methods] are superior to perceptual methods in terms of reliability" (Helfrich, 1980, p. 251).

Martin & Strange (1968) found that it was difficult for subjects to attend to acoustic and message/contents of a speech signal simultaneously, and that these were "incompatible operations" (Martin & Strange, 1968, p. 438). They also found that subjects displaced within-constituent disfluencies to constituent borders.

Warren (1970) showed that if one cuts out a phoneme from a word and replaces it with a cough, people still hear the indented, no longer existing phoneme, and displace the cough to another location, quite often outside the word (even between words). If the excised sound was replaced with silence of the same duration, the absence of the sound was perceived and correctly located.

Cole (1973) exchanged phonemes in a read text with phonemes of various degrees of differences from the original phonemes as to the number of differing phonetic features. He found that phonemes that differed only by one feature were detected much more seldom than phonemes that differed by two or more features. However, when the words were played in isolation, rather than as parts of a meaningful text, all phoneme changes were detected. Cole concluded that human speech perception makes use of phonetic-features, and that semantic context hampers mispronunciation at lower levels.

Bond & Small (1983) had their subjects shadow passages containing mispronunciations, and showed that words containing voicing errors typically were "corrected" when repeated, while stress and vowel errors caused more problems to the subjects.

The phenomenon of not recognizing disfluencies have further been studied by e.g. Duez (1981/1982, 1982, 1983/1984, 1985, 1995), Duez & Carré (1983), as well as Lickley and Bard (and colleagues) in a number of experiments (e.g. Lickley, Shillcock & Bard, 1991; Lickley, 1994, 1995; Lickley & Bard, 1992, 1996, 1998a; Bard & Lickley, 1997; Bard & Lickley, 1998a, 1998b). As Bard & Lickley point out:

> Transcribing disfluent speech verbatim is inordinately difficult: the contents of the disfluency seem strangely evanescent. Without many replays of the material, even the location of the disfluency is difficult to ascertain. (Bard & Lickley, 1998b, p. 108.)

The fact that transcription of disfluent speech is cumbersome, paired with the knowledge that speech interpretation is dependent on both preceding and subsequent context (*vid.* Cole, 1973), prompted Bard and Lickley to conduct a number of word-gating experiments. Lickley & Bard (1992) found that disfluency was most often recognized on the first word after the interruption point, and that disfluency was often detected before the first word in the continuation (reparans) was finished, i.e., disfluency recognition preceded word recognition. Lickley & Bard (1996) found that words never recognized were more common in disfluent utterances than in fluent utterances (showing that disfluency has a detrimental effect on speech comprehension), and that recognition failures clustered around the point where the disfluency in question interrupted the flow of the utterance. It was argued that disfluencies distort both the preceding and subsequent context of an utterance. Bard & Lickley (1997) showed that words close to the interruption point are the hardest to recognize, and that words in the reparandum are harder to recognize than word in the continuation (repair).[1]

[1] Bard & Lickley (1997) use the previously mentioned three-part model of speech repairs, in which a disfluency is said to have a **reparandum**, the part which is to be deleted or substituted, an **interruption point**, the point where the speech flow is interrupted (with or without explicit signaling), and a **continuation** (also called **reparans**), which is the "fresh" speech that replaces the erroneous speech of the reparandum.

That disruption of speech flow is worst in the reparandum was also observed in Bard & Lickley (1998b). Bard & Lickley (1998a) examined whether listeners were able to predict disfluency from cues in the reparandum, based on a claim made by Hindle (1983) that identifiable acoustic editing signals occur before the interruption point.

Bard & Lickley (1998a) did not find that listeners were able to detect imminent disfluency (i.e., detect disfluency before it occurred), only disfluency that had already begun. Moreover, even if there was an editing signal at the end of the reparandum, listeners did not seem to be able to use it to predict ensuing disfluency. However, the offset of the continuation (reparans) enabled listeners to identify disfluency, even in the absence of explicit editing signals.

Duez & Carré (1983) found that recognition rates of pauses were strongly correlated with duration. Pauses longer than 900 ms had high recognition rates whatever their prosodic or syntactic distribution. Duez (1985) concluded that the prosodic structure is of importance for the detection of pauses, as well as the duration of the pause. Duez (1995) found that while silences (unfilled pauses) were not detected, filled pauses, lengthening (prolongations) and repeats were detected, and stated that hesitation phenomena are not beyond reach for listener, but can be heard.

Cohen (1968/1973), Hill (1973) and Fromkin (1971/1973) pointed out that slips of the tongue (mispronunciations) often go undetected by listeners. Tent & Clark (1980) investigated error detection of phonemic and nonphonemic slips of the tongue—using a definition given in Nooteboom (1973)—and found that phonemic slips were harder to detect than were nonphonemic slips.

Bond & Small (1984) examined whether three types if mispronunciations could be detected by listeners, *viz.* voicing errors (voiced obstruents replaced by unvoiced obstruents and vice versa), vowel place (front vowels replaced by back vowels) and stress (wrong syllable in the word stressed) and found that stress errors were harder to detect.

2.18 Terminology and definitions

Throughout this chapter we have seen disfluencies studied from an array of different perspectives and angles, for different reasons and with various motives. We have also seen that the phenomenon under scrutiny is being referred to with a variety of different terms. Terms are never neutral, and whatever word you decide to use reveals some underlying assumptions and definitions of the phenomena you desire to describe. Moreover, even in cases where people agree on what term to apply for a given phenomenon, there is sometimes disagreement as to the exact denotation of that term in the real world.

Before closing this chapter I would just like to briefly summarize the terminology slash definition issue. It might seem odd to *close* this chapter with a discussion on terminology and definitions, but to me this issue is so much clearer if one is provided with the various backgrounds that serve as the basis for the research that has been carried out on disfluency. The way I see it, this knowledge is needed in order to understand why there is such an issue in the first place.

2.18.1 Disfluency… or what?

First, terminology *has* been debated openly, at least within the stuttering community, which will briefly be reviewed in the following.

The earliest research specifically devoted to disfluencies, and the first major attempt of classification of disfluency in speech was made by Johnson and colleagues at the University of Iowa (Johnson et al., 1948; Johnson, 1955; Johnson and Associates, 1959; Johnson, 1961). A list of "nonfluencies" and subcategories was first presented in Johnson and Associates (1959, p. 201) and during the years that followed, many researchers adopted both the classification and terminology proposed by Johnson and his group, making the terms household names within the field. In 1961, the term **disfluency** appeared (Johnson, 1961)[1], and Brutten (1963) writes that "disfluency is defined as interruptions and breaks in the flow of the speech signal" (Brutten, 1963, p. 41).

An early comment on terminology proper is given by Neelley (1961), who discussed the use of the term "stuttering" as it appears in Johnson and Associates (1959):

> The work of Johnson /…/ and his students suggests that it is not unusual for the term *stuttering* /…/ to be employed as a label for other types of disfluent speech. The most inappropriate and disadvantageous usage of the word probably occurs when the generally observed disfluencies of childhood speech are referred to as 'stuttering' and reacted to as unusual or abnormal. Speech can be disfluent for several reasons, but disfluencies due to one cause may be qualitatively different from disfluencies due to another cause. (Neelley, 1961, pp. 79–80.)

MacDonald & Martin (1973) argued that disfluency and stuttering by definition are different phenomena, and that there is no overlap between the two. In their study they asked subjects to judge speech material as disfluency, stuttering, both, or neither. They also pointed out that "what distinguishes stuttering from disfluency is the way people evaluate certain behaviors" (MacDonald & Martin, 1973, p. 692).

Wingate (1984b) discussed the use and misuse of the four terms *disfluency*, *dysfluency*, *nonfluency* and *fluency*. Wingate first pointed out that disfluency is the most obfuscated of the terms, and that the prefix *dis-* denotes reversal of the item it specifies. Thus, the meaning of **disfluency** would simply be everything that is not fluent. The term has been used at least since Johnson (1961), and has often been used as equivalent with stuttering.

Similarly, this is also the case with the term **nonfluency**, which is etymologically equivalent with disfluency, the only difference being that the prefix is in Latin instead of Greek. However, as Wingate points out, to use either of these terms interchangeably with stuttering would imply that speech therapists would aim at improving the fluency of all children, which is not the case.

Franklin Silverman & Williams (1967a) made the following distinction between *disfluency* and *stuttering*: "Disfluencies include all types of disruptions in the rhythm of speech, whereas judgments of stuttering do not necessarily" (Silverman & Williams, 1967a, p. 1085; footnote).

[1] For an overview of early terminology the reader is referred to Wingate (1987).

A third term in use is **dysfluency**, whose improper use Wingate (1984b) ascribes to its homonymousness with "disfluency".[1] As Wingate correctly pointed out, the Greek prefix *dys-* has a completely different meaning from *dis-*, the former meaning "abnormal", which is why it is mostly used within medical contexts denoting pathological conditions. Thus, the phrase "normal dysfluency" (Ratner & Sih, 1987, p. 278; Dale, 1977, p. 312; Shapiro & DeCicco, 1982, p. 109/title of paper; Westby, 1974, p. 133) is a contradiction in terms, as is "dysfluency" of "stutterers and nonstutterers" (Floyd & Perkins, 1974, p. 279/title of paper). Thus, to the extent that stuttering is considered a pathological condition, the term *dysfluency* would be adequate, but not *disfluency*.[2]

In the same vein, Quesal (1988) also pointed out that the terms *dysfluency* and *disfluency* are often confused and incorrectly used in the literature, with basically the same argumentation as Wingate (1984b) used. Quesal, however, also made it clear that both terms could be used to describe the speech of stutterers:

> For a number of reasons, the use of dysfluency has become more popular in recent years. This, unfortunately, has apparently led to the belief that the terms dysfluency and *disfluency* are synonymous and can be used interchangeably. This is not the case. The prefix *dis-* is used to form words that define the opposite of something or lack of something. In this sense, the word disfluency means a lack of fluency in speech or simply speech that is not fluent. We all exhibit disfluent speech at various times. I also would imagine that a good proportion of a stutterer's speech could contain disfluencies. On the other hand, the prefix *dys-* means bad, ill, difficult, abnormal, and so forth. Therefore, the word dysfluency would refer to speech that is abnormal. Most likely, our primary concern with stutterers' speech is the dysfluency they exhibit. (Quesal, 1988, pp. 349–350; italics in original.)

What all these terms have in common is the tacit assumption that there is a phenomenon **fluency** that everybody recognizes (Wingate, 1984b, p. 166). Although it would seem that most people, researchers, linguists and laymen alike, would indeed recognize and acknowledge that there is a certain feature fluency that could be applied to spoken language.

Wingate concluded:

> But fluency is an abstraction—an abstraction that reflects a perceptual extrapolation from truly flowing samples of a person's speech that are, however, typically brief. /…/ In reality, fluent speech does not match its abstracted definition. "Fluency" is an illusion, a fact borne out by a considerable amount of research on normal speech (Wingate 1984). Speech perceived as normal fluent speech typically contains a variety of "disfluencies"; in fact, it is characterized by "disfluencies" other than those appropriately referred to as instances of "dysfluency". (Wingate 1984b, p. 167.)

Quesal (1988), on the other hand, stated that:

> My personal feeling is that the term disfluency is the more useful and descriptive of the two, simply because there is no evaluation of fluency involved. (Quesal, 1988, p. 350.)

[1] Which of course is the case in English, but not in Swedish, French, German and other languages.

[2] Culatta & Leeper (1988) pointed out that dysfluency is not synonymous with stuttering, since there other ways in which speech might suffer from lack of fluency.

Among authors who use the term disfluency as a hyperonym are Horii & Ramig (1987), who pointed out that:

> Typically, the term disfluency connotes normal mistakes of speech whereas dysfluency refers to abnormal. For simplicity, however, the authors of this study have used disfluency to refer to both normal and abnormal speech errors. (Horii & Ramig, 1987, p. 257; footnote.)

It could also be pointed out here that Ratner (1988) defended the use of dysfluency and disfluency interchangeably by referring to definitions given in the *Random House dictionary of the English language* (1987) and the technical dictionary *Terminology of Communication Disorders* (Nicolosi, Harryman & Kresheck, 1978/1983/1989).[1]

That no consensus has been established is evident from e.g. Prins (1991) who used both disfluency and disfluency in the same article without explicitly defining if there is any intended difference as to meaning. Also, Kolk (1991) used "normal disfluencies" to make clear that they are not stutterings, and Postma & Kolk (1993) make a difference between "normal and stuttered disfluencies", which seems to be what Quesal (1988) proposed. Also, Postma, Kolk & Povel (1990) distinguished between "speech errors (deviations from a speech plan), disfluencies (interruptions in the execution of a speech plan), and self-repairs (corrections of speech errors" (Postma, Kolk & Povel, 1990, abstract).

Given that disfluencies have been studied for such a long time, and within several different disciplines with little or no contact, it is not surprising that terminology is inconsistent. Besides the argumentation summarized above concerning various kinds of fluency, other terms have also been used, such as *speech disturbance*, *discontinuities*, *own communication management*, *hesitation phenomena*, and so on and so forth.

Given the lack of agreement and the cornucopia of terms to choose from, the term **disfluency** will be used throughout this thesis (which should be obvious by now), mainly for the reasons provided by Wingate and Quesal. This does not mean that I do not acknowledge that the phenomena under scrutiny might indeed *contribute* to perceptual fluency, rather than be a detriment, on the contrary. The term has mainly been opted for since it is the de facto most widely used (which a web search makes evident), and since its denotation is in fact the *most* neutral, albeit not *completely* neutral, since no such thing (probably) exists.

Below, just a brief discussion of some the major subcategories will be made, mainly to facilitate comparison with previous research.

2.18.2 Unfilled pauses… or what?

A problem with unfilled pauses is that they range from the very obvious, like a seconds-long silence in the middle of the word, to hardly noticeable silences between e.g. phrases or even sentences. What counts as an unfilled pause? This is also why several disfluency studies exclude unfilled pauses, despite the fact that "[o]f the hesitation phenomena, unfilled pauses are the most frequent" (Martin, 1970, p. 75).

Cowan & Bloch (1948) recorded twenty minutes of continuous discourse and had twenty subjects mark all perceived silent pauses. They compared the thus marked pauses with the

[1] Ratner (1988) gives the year of publication as 1987.

acoustic signal, and noted the by now well-established lack of a one-to-one relationship between the perceived and the physical. Or, as they observe:

> [A] comparison of the observers' reports and the physical record shows that some of the 'perceptual pauses' were located at points where there was no actual interruption of the physical speech energy, and that on the other hand some relatively long interruptions of the physical energy were not detected as pauses. (Cowan & Bloch, 1948, p. 92.)

They also observed that pauses were more likely to be perceived at certain grammatical locations than at other positions in the speech string, and that a full grammatical analysis of the text would be needed in order to determine the role of syntax in the perception of silent pauses in speech.

Martin (1970) also compared listeners' perception with the acoustic-physical signal from a grammatical perspective. He used a lower cut-off duration of 50 ms. He found that 87% of the places where there were actual silences in the speech signal were also perceived by his subjects as pauses. The duration of these ranged from 50 ms (the lower limit included in the test) to 4970 ms (the longest pauses). Actual, physical, silences that went unnoticed ranged from 59 ms to 110 ms. Martin (1970), like Cowan & Bloch (1948) observe that pauses are perceived where there is no silent interval in the signal, while some silences go unnoticed by the listeners. He concludes that factors like speech rate, grammatical junctures and elongated speech sounds all play a role in the detection of unfilled pauses.

Rochester (1975/1976) summarized previous studies on unfilled pause detection—including Cowan & Bloch (1948), Goldman-Eisler (1968), Boomer & Dittman (1963) and Martin (1970)—and stressed that the phonemic clause play a crucial role in the detectability of silent intervals to listeners. She summarized that:

> Long pauses are always detected and no further variables are needed for explanation, while detection of short pauses (50–200 msec in Cowan & Bloch's work; 50–110 msec in Martin's study) depends on linguistic cues. (Rochester, 1975/76, p. 3.)

The aforementioned studies are, of course, not the only works that have concerned themselves with the "lower limit" of unfilled pauses. While sometimes cut-off durations have not been motivated or stated at all, others have given a diverse array of motivations. Verzeano & Finesinger (1949) used a cut-off duration of 500 ms for their automatic analyzer, while Goldman-Eisler excluded pauses shorter than 250 ms. Levin & Silverman (1965) used a one-second limit for unfilled pauses (which they refer to as *zero aggregates*) since "the accuracy of using a stop watch did not permit greater precision than one second" (Levin & Silverman, 1965, p. 72). One second was also the lower limit for Lay & Paivio (1969). Hieke, Kowal & O'Connell (1983) argued that Goldman-Eisler's cutoff duration of 250 ms was not tenable, and that pauses with durations between 130 and 250 ms can be systematically related to psychological factors.

Further, as was mentioned above, both Cowan & Bloch (1948) and Martin (1970) used a lower cut-off duration of 50 ms, something which was perceived by the subjects as silent pauses. Martin & Strange (1968) found that what was perceived as unfilled pauses was not necessarily silent stretches in the speech signal, but could also be changes of pace (Martin & Strange, 1968, p. 437). Holmes (1988) used 200 ms as the lower limit. Duez & Carré (1983) also observed "subjective pauses" that did not correspond to pauses in the acoustic signal. While some of these pauses seemingly depended on fundamental frequency, some

subjectively perceived pauses seemed to have no such causes. This is head-on to Helfrich's (1980) view that automatic pause extraction methods are superior to human perception. More recently, Hansson (1998) defined (silent) pauses perceptually, i.e. "a perceived pause, regardless of whether it can be associated with a silent interval in the speech signal or not" (Hansson, 1998, p. 158).

So, given all this, the issue boils down to whether the unfilled pause should be defined perceptually or acoustically. Rochester finished her paper by concluding that:

> It is now clear, for example, that a simple duration measurement of brief pauses is naïve to the extent that it ignores the linguistic context in which the pause occurs. For this reason, some (e.g. Martin, 1970; Boomer & Dittman, 1962) have argued that listener judgment is generally preferable to physical recording of pauses. This conclusion seems overstated however, since the only reliable data on interacting systems pertains to pauses less than 200 msec[1] in duration. Since most of the experimental investigations of pauses focus on intervals longer than 200 msec, it seems safe to ignore factors other than duration when defining pauses from the listener's point of view. (Rochester, 1975/76, p. 4.)

Silences, *or unfilled pauses* are cumbersome, since it is hard to tell whether they reflect disfluency or not. Nivre, Allwood & Ahlsén (1999) mentioned that whether pauses should be regarded as "part of an utterance or not, is to be decided on the basis of the context" (Nivre, Allwood & Ahlsén, 1999, p. 9). Fox Tree (1995) pointed out that disfluency counts vary a lot in the literature depending on whether pauses are included or not, and preferred disfluency figures exclusive of pauses simply because "not all pauses are disfluencies" (Fox Tree, 1995, p. 709).

Bell, Eklund & Gustafson (2000) considered silent pauses as phenomena on a scale from sure-fire disfluency (e.g. inside lexical roots) to almost certainly consciously intended devices, e.g. between grammatically well-formed sentences (Bell, Eklund & Gustafson, 2000, p. 627). Goldman-Eisler (1968) excluded all silences shorter than 250 ms, counting only silences of longer duration.

In this work, unfilled pauses have been included, trying to bear in mind all the complications associated with this category. Silences inside lexical roots leave little doubt, but between grammatical sentences within one utterance (which in itself escapes most attempts of a clear-cut definition), they have been less clear. Prosody has played a role in the judgement of when to regard something as a disfluent unfilled pause, or not. Generally, labeling has been conservative. That being said, let me finish by citing Deese (1978), who astutely points out:

> Unfilled pauses, of course, are not always disfluencies. We use pauses to mark sentence boundaries and other segments of discourse. The usual practice is to treat very long pauses—more than one-quarter of a second—as disfluencies, but this leads to problems. /…/ The truth of the matter is that some pauses represent failures of fluency and others do not. It is impossible to tell which is which without the full meaningful context of the discourse, and even then there are doubtful cases. (Deese, 1978, p. 318.)

I could not agree more, but that is no excuse for avoiding the category altogether.

[1] I have added a space between "200" and "msec" which is lacking in the original.

2.18.3 Filled pauses… or what?

As we have seen, *filled pauses* are arguably different in many ways from other types of disfluency. They have convincingly been shown to occur for different reasons, and they do not react to manipulation in the same way other types of disfluencies do, which led e.g. Mahl to create his *non-ah ratio*, which has been widely used in disfluency research.

The term used to describe *ah*, *um*, *er* and similar orthographic rendering vary widely in the literature (as should be obvious from the sections above), and include such terms as *interjection* (often lumped together with words and phrases such as *well*, *kinda* and other structure-forming devices, which are excluded in the definition used here), *vocalized pause*, *fillers*, and so on. The term *filled pause* appears to be the most common, but could definitely be considered as not ideal, since the word *pause* has certain connotations that not everyone would ascribe to. However, there are also good arguments for the view that *er* at times does function as a pausing device in speech production, while at the same time the view that it should be regarded a full-fledged *word* of English can also be easily defended.

In this work, I have opted for the term **filled pause**, although I am more than willing to acknowledge its flaws.

2.18.4 Prolongations… or what?

Prolongations, once thought to be a tell-tale sign of early stuttering, occur early in the literature, and are quite often included in the category *dysrhythmic phonations* (which also included e.g. cut-off words or syllables). Other terms that appear are *elongations*, *stretched-out sounds*, *drawls* and so on. This term seems less problematic than most others, and will be employed to describe phones that are longer than should be expected in normal-paced, fluent, speech.

2.18.5 Explicit editing terms… or what?

Shames & Sherrick (1963)—referring to Skinner's work—claimed that "speakers very often 'compose' and 'edit' and 'prompt' themselves for their verbal behavior" (Shames & Sherrick, 1963, p. 8). The notion is that some repairs are explicitly signaled by the speaker with words like *oops*, *sorry*, *wrong* and so on. Early production work (e.g. Hindle, 1983) was partly based on the detection of such signals. Sometimes what are here called filled pauses have been included in the category of editing terms. The term *explicit editing term* will be used in this work to described meta-linguistic words that refer to other words, and not to the message proper, including items such as e.g. *wrong*, *sorry*, *no, I mean* and similar. Filled pauses are not included in this category.

2.18.6 Mispronunciations… or what?

The term *slip-of-the-tongue* is similar to the term *mispronunciation*, employed in this work. However, in early linguistics research, slips-of-the-tongue often included spoonerisms, i.e., when an intended word is replaced by another, actual, real, word, leading to an unintentional, utterance. Indeed, the *lexical bias hypothesis* claims that most slips result in real words. In this work, the term *mispronunciation* will only refer to uttered words that are not actual, real words. This means that the oft-repeated (in the literature) example *barn door–darn bore* will not be considered a mispronunciation (but will instead be described as a repair (see below), while e.g. *barn door–dran (door)* would be considered a mispronunciation.

2.18.7 Truncations… or what?

As was mentioned above, cut-off words (or syllables) were included in the category *dysrhythmic phonations* (together with prolongations), and are often simply called *cut-offs* in the literature. Other terms, such as *interrupted words/syllables* occur, as well as other descriptions. The term *truncation* will here be applied to all linguistic items that are not fully executed/finished, whether or not they are finished later. This means that an item such as *bilj ... ett* (*tick ... et*) will count as a truncation, even though the word is completed later (after an unfilled pause).

2.18.8 Repairs… or what?

The term *repair* appears throughout the literature, quite often in application-driven research (e.g. Heeman and others). The idea is that something needs to be corrected, and that there is a structure to repairs themselves, with a *Reparandum*, and *Interruption Point* (sometimes an editing term), and a/the *Repair* (or *Reparans*). A repair can include other phenomena, such as *repetitions*, *substitutions*, *insertions*, *deletions* and so on.

These latter terms have sometimes been studied without the notion of *repair*, e.g. within stuttering research, where part-word repetitions are legion. Indeed, several early studies include such terms—especially *repetitions*—in lists of disfluencies under scrutiny. However, their "status" has been under discussion. Wingate (1994) stated that "whole-word repetitions should not be considered as stutter events" (Wingate, 1994, p. 581).

While initial-sound repetition has often been considered a sign of stuttering, Lebrun & Borsel (1990) noted that *final*-sound repetitions are rare, both in stuttering and in normal speech. Perusing the literature, they concluded that "there seems to be very little reason to consider final sound repetitions in children to be intentional and phonologically motivated. Rather, they appear to be real, pathologic dysfluencies." (Lebrun & Borsel, 1990, p. 112).

In this work, repetitions, insertions, deletions and substitution will always constitute parts of a repair. It must be pointed out, however, that repairs can also incorporate e.g. prolongations, filled pauses and so on, but that these do not always constitute parts of repairs (the way labeling has been applied in this work).

2.18.9 Summary

As we have seen, as was the case with terminology at the highest level, various terms have been employed at the lower levels, as well. However, it is of interest that despite the fact that much of the research described in this chapter has been carried out in parallel, as it were, without any noticeable interflow of information between the different fields, and despite that terminology in most (if not all) cases is colored by the specific research angles and fields where it is applied, with the consequently different connotation carried by the specific terms used, the denoted categories are very much "the same", which validates analyses throughout the literature. It may be the case that not all categories are included in all studies, and in some cases a term is being employed to cover more than one category (as was the case with *dysrhythmic phonation*, which denotes both cut-offs and prolongations), but a closer look quickly reveals that all studies, and all terminology described a fairly limited number of easily identifiable phenomena.

2.19 Chapter summary

So, then, and without further ado, let us summarize this chapter.

2.19.1 Stuttering

To summarize almost a century of stuttering research in a few pages necessarily entails skimming the surface, and, which should be obvious from the paragraphs earlier in this chapter, it would almost be easier to list what aspects have *not* been considered in stuttering research than those that have. Consequently, in order to (as the underlying objective) develop a diagnostic for stuttering, speech in stutterers and nonstutterers has been studied in the light of personality features (both children and their parents), developmental factors, different speaker settings (informal, formal, to toys, at school, at home and so on), fluency-enhancing conditions (masked noise, self-pacing, metronome pacing, speaking in chorus, singing, slowing down, speaking to an animal, speaking while swinging the arm, changing the dialect), brain scans have been carried out (using e.g. the Wada technique) in order to discover any hemispheric deviations, breathing patterns, laryngeal functioning, reaction times in different tasks (both speech and nonspeech), and even the chemical balance of the speakers. And so on ad infinitum. While obviously of utter interest to researching in stutterering, this research has also brought with it enormous amounts of knowledge of fluency and disfluency in nonstutterers, since the bulk of the work that has been carried out has included control groups of nonstutterers. Consequently, we now know under what conditions nonstutterers become more, or less, disfluent. We also know what categories of disfluency stutterers and nonstutterers exhibit, and to what extent the relative frequencies differ, but also what the absolute values are in different contexts, settings and conditions.

Although most studies show that there is an overlap between disfluency behavior in stutterers and nonstutterers, both concerning categories and frequencies, there are also indications to differences between the two groups. Nevertheless, it has been shown over and over again that it is very hard to pinpoint what *normal* fluency is, as opposed to e.g. *stuttered* (dis)fluency. Although the general view (with exceptions, of course) seems to be that there *is* indeed a difference, laymen and professionals alike exhibit the same problems in identifying stutterers, and tell them apart from normal speakers, and an “acid test” of stuttering seems ever-elusive Also, there seem to be at least some differences that do not overlap, like the general effect of DAF speech (which makes stutterers more fluent, and nonstutterers less fluent, although the results depend on the time delay setting), or the relapse effect after adaptation in stutterers. It is my contention here that from a linguistic, or application-driven, perspective, there is much to learn from the enormously rich literature on stuttering, especially as regards fluency in children at various ages and developmental stages. Finally, as was also shown, there are indications that allegedly fluent speech of stutterers is different from fluent speech in nonstutterers, which lends further support to the hypothesis that there is a qualitative, categorical difference between stutters and disfluencies.

2.19.2 Psychotherapy and psychology

Studies within this field, which during the early period might have been among the most influential and commonly spread ways to analyze and categorize disfluency phenomena, show us that disfluency production is dependent on a number of psychological and individual, factors. When we speak, the topics discussed, and our attitudes towards them, influence not only our choice of words, or what we say, but also *how* we speak, i.e. how fluent we are.

Moreover, these changes in fluency might be very local, implying that speech production is a fast-changing phenomenon, where fluency may be different from one sentence to the next. Also, phenomena such as "choking under pressure" reveal that expectations—our own or other's— do play a significant role in speech production. It has also been shown that such effects depend on individual personality traits, i.e., the proportions of different *kinds* of self-awareness a speaker possesses, private, public or social self-awareness (or self-consciousness). It has also been shown that disfluency is susceptible to manipulation of sorts, indicating that disfluency production to some extent is under speaker control. That different kinds of stressful situations affect different kinds of disfluency was noted already by Mahl, who showed that filled pauses in many ways behaved differently from all other kinds of disfluencies (which has since then been confirmed by others). It has also been shown that disfluency to some degree is under speaker control, in that phenomena like punishment and instruction have immediate and significant effects on disfluency ratios of speakers. There is no doubt that much of what we know about under what conditions disfluency occurs is to be found within psychological literature.

2.19.3 Physiological factors

A slightly less studied area is to what degree disfluency is related to physiological conditions. That inebriation has an effect perhaps comes as no surprise, but that its interrelatedness with the psychological phenomena like self-awareness (and consequently self-monitoring) under some circumstances make speech *less* disfluent might be more unexpected. Several studies have pointed to gender differences in disfluency production, while some have failed to observe such differences. It has also been shown that biological cycles seem to effect disfluency production, and it has also been proposed that hesitation vowels are merely an artefact of physiological breathing mechanisms. With the introduction of speech recognition software in high-performance environments like aircraft cockpits or space stations, both physiological and psychological parameters have to be considered. Not only do heavy *g*-forces affect speech and performance, since high cognitive load might be co-occurring with other tasks, speech detriment under multidimensional conditions needs to be studied, and this field is also given more and more attention.

2.19.4 General linguistics

Much of the early work was carried out either on slips-of-the-tongue or spoonerisms—both examples of mispronunciations in this work—two types of speech errors that are extremely rare. Moreover, several of these were not spontaneous, but elicited in the laboratory. The fact that slips *can* be elicited is in itself interesting, but the bulk of disfluency as a phenomenon consists by and large of other categories, which oddly enough were far more studied within stuttering research or psychotherapy, at least during the first decades. However, early linguistics did also devote much effort to hesitation, roughly filled and unfilled pauses, which combined constitute the majority of the disfluency encountered in spontaneous speech. The notion that disfluency should not be seen simply as performance errors was one of the steps within linguistics where speech *as it occurs* started to be the object of studies. While most of the early studies focused on either pausing or hesitations, on the one hand, or slips-of-the-tongue (or tip-of-the-tongue), on the other, they lay the foundation for the incorporation of disfluency into a linguistic framework in general, where it is now argued that disfluency not only should not be seen as performance errors (à la Chomsky), but rather as full-fledged communicative tools to enhance both linguistic comprehension, and also paralinguistic transfer between interlocutors. Consequently, what is called disfluency are (probably) often

examples of *fluency*. To really augment our knowledge about disfluency as a universal language phenomenon, other languages need to be studied, like Tok Pisin, Tagalog or Ilokano. Disfluency in different social groups has also been studied, hinting at small differences as to verbal planning. That prosody plays a major role in speech communication is clear, and several studies have pointed towards interaction between disfluencies and prosody. One of the more interesting stances forwarded in the literature is the notion that disfluency is in fact beneficial in human communication in that it *helps* the listener decode the message conveyed by a speaker. It has also been proposed that different types of disfluencies also serve different purposes in that respect.

Summing up this section, it must be pointed out that most of the studies described in this chapter have been on English. While it may be the case that English is indeed the most studied language, it is by no means the *only* language studied. Other European languages such as Dutch, German, Swedish, Finnish, French, Spanish and so on have also been devoted a large number of studies. Chinese and Japanese are also represented to a large degree. However, the said languages represent only a small fraction of the multitude of language spoken, and from a global point of view there is little difference between English, Dutch, German and Swedish, for example.

2.19.5 Speech production

Perhaps the most interesting question (depending on one's personal inclinations and penchants) of all possible things we could ask about disfluencies is what they reveal about *how* we produce language, i.e., how the brain, or our "psyche", does it. Or *what* language really *is*. What role does language play in consciousness? To what extent are we aware of what we are saying, or what goes wrong? Is language production conscious, preconscious, subconscious or perhaps not very conscious at all? Strikingly, most speech production models that have been proposed over the years are to some degree based on the study of disfluencies, since the only way (more or less) to study the inner workings of an "inaccessible" system, like that of language production, is to focus on the ways it can fail. From early on, it was noticed that there were regularities in speech errors, and also separation between different aspects of language, such as the previous mention of a separation of lexical retrieval and prosodic stress assignment (the one seemingly independent of the other, at least judging from some recorded speech errors).

The speech production field deals with the profoundest issue underlying all other approaches to speech disfluency, and is for that reason closely interrelated to a variety of other fields devoted, partly or exclusively, to human behavior, such as biology, neurology, computer science, philosophy, cognitive psychology and so on and so forth. There seems to be a slight asymmetry in the research carried out concerning this area, in that neurologists, neuroscientists, philosophers, computer scientists and so on perform research on language and speech production and perception, while psycholinguists do not refer to findings within work on neuromotor functions, brain potentials (ERPs, CNV, fMRI, EEG, EMG and so on) studies of motor execution or brain functioning, many of which have significant implications for the linguistically motivated theories and/or models, as we have seen.

To me, speech—or more correctly, *language*—production models would benefit from the inclusion of what is known about timing events and general processing in the brain, and how recorded brain potentials (or similar) relate to actual executed actions in the human organs, be they fingers, feet, hips, toes and so on, or the speech organs. The problem, as I see it—which

was also pointed out by Blackmer & Mitton (1991)—is that temporal features are not treated in detail (if at all) in existing speech production models, and that phenomena like inner and outer loops, conscious, preconscious or subconscious detection repair cannot be discussed without taking into consideration what is known about reaction times and neurological functioning and so on.

Prima facie, it might seem that concepts like consciousness, free will, and the way they are affected by e.g. Libet's observations, might be peripheral from speech disfluency per se, but to me, and to some of the researchers within the speech production field, speech would be the most interesting of all motor executions one should study. Other fields that could shed further light both on speech (language) production and perception, consciousness and concepts like inner speech and monitoring, are not devoted more than a small number of interdisciplinary studies. The approach of Velmans to try to integrate all (or most) of this knowledge from different fields is indeed a major undertaking, but of utter interest, the way I see it.

2.19.6 Schizophrenic speech

As was shown, within the field of speech production, the notion of "inner speech" is frequently treated, in various ways. Whether inner speech is conscious or subconscious, or, put another way, whether it is to be taken *literally* as speech, has been studied within schizophrenia, since many schizophrenic exhibit auditory hallucinations. Studies on both covert and overt schizophrenic speech have yielded interesting insights into speech production and consciousness, as have—to some extent—studies of split-brain patients.

2.19.7 Sign language

While most studies refer to *speech* production, to me it seems clear that what we are talking about is *language* production, which most often happens to be speech. Disfluency in sign language could without doubt point to underlying processes in language production, and some such studies have also pointed towards universal traits, although more studies are needed before any far-reaching conclusions may be drawn.

2.19.8 Application-driven approaches

The fact that we communicate more and more often with machines, either one-way or two-way, in simple command language or in more complex dialogues, more resembling full-fledged human–human conversation, is a relatively new phenomenon, and its appearance on the arena is an accelerating trend. Advances in automatic speech recognition and synthesis are constantly yielding increasingly human-like systems that not only *sound* more and more human-like, but also *behave* like humans. Until recently, however, such systems were based on idealized language, and did not take spoken-language phenomena, such as disfluency, into account. Current systems are becoming increasingly capable of handling disfluencies, such as filled pauses, or correcting repetitions and changes, and this trend is sure to continue.

That both recognizers and dialogue systems that are able to cope with truly spontaneous speech will have an edge on systems that require "discipline" on behalf of the speaker goes without saying. As regards synthesis, an issue can be raised whether or not speech synthesizers should be disfluent. As we have seen, several studies seem to indicate that comprehension is actually helped by e.g. filled pause at the right locations. Whether or not disfluent computers will appear natural to human users will have to await further research,

and like other technological advances, such as hot air balloons, railways or mobile phones, one could assume that human attitudes might change over time, so that short-term hostility might develop into long-term acceptance, or even desire. However, the dialogue systems of the future might well exhibit disfluency, to a lesser or higher degree.

2.19.9 Disfluency in a nonnative language

Given that an increasing number of automatized services are launched in more than one country, nonnativeness is also of interest. In what way do people bring their native disfluency with them into a second, or third, language? It has been argued that disfluency should be taught at schools, together with syntax and pronunciation, and perhaps this is the case. But given foreign accents at all levels, even speakers who have studied disfluency in the target language could be expected to have an accent in their disfluency production, as well.

2.19.10 Disfluency and bilingualism

The study of bilingual disfluency could provide insight into universal, and language-specific aspects of speech production, and could also be used as the starting-point for speech production models, as we have seen.

2.19.11 Crosslingual aspects of disfluency

Like bilingual studies, crosslingual studies—comparing disfluency in different languages—could reveal potentially universal features in human speech production. This is but also of interest to developers of speech-based applications, since similarities and dissimilarities between closely or remotely related languages would affect how e.g. automatic speech recognition might be optimized for, or tuned to, a given language.

2.19.12 Gestures

That we do not exclusively communicate with speech, but also with facial expressions or arm and hand gestures is well known. A number of studies have aimed to describe how gesture communication is related to speech communication, and showing that there are indeed very stable interactions between the two modes. It is easy to envisage future systems that not only react to head nods (for confirmation), but can interpret hand and arm gestures, such as pointing. So from an application point of view, gesture recognition is of interest, which entails that gesture disfluency is of interest, especially given the results that show that gesture disfluency precedes speech disfluency, which means that by combining the two, more robust disfluency handling should be achieved.

2.19.13 Disfluency in writing

As was the case with sign language, humans also (to some extent) communicate in writing. While raising different issues concerning the particular mode of writing (typing, by hand and so on) or the particular type of writing system (alphabetic, iconographic and so on), studies of disfluency in writing provides yet another source of knowledge with regard to underlying processes in human language production.

2.19.14 Paralinguistic aspects of disfluency

Trager divided communication into three different parts, language, paralanguage and kinesics, and the issue raised by that division is where disfluency belongs in that trichotomy. Trager himself considered e.g. filled pauses to be a part of paralanguage, while others have considered at least filled pauses to be words in their own right.

2.19.15 Disfluency among the elderly

It is well known that language changes as a function of developmental factors and age, and it has been shown that this also occurs with regard to disfluency. While speech and language of the young is well studied, comparatively few studies have been devoted to the elderly, and such studies would be welcome in order to enhance our knowledge about disfluency, as it occurs in human speech.

2.19.16 Effects of disfluency

So, given the cornucopia of observations as to the etiology, characteristics, frequency, distribution, taxonomy (and so on) of disfluency, the pending question is: *do they matter?* Several studies have shown both that we to a large degree do not notice them, but instead filter them out from perception. Other studies have shown that they are detrimental to understanding, although some claims to the opposite have been made. Finally, some studies have shown that while disfluencies do not affect understanding proper to any larger degree, they do affect listeners' attitudes towards the speaker with regard to personal qualities.

2.19.17 Terminology and definitions

I ended this chapter by setting aside some place to the issue of terminology, and most of what I wanted to say within that field was said there and then. Some people are more concerned about terminology than others, and proponents of one term are sometimes willing to go at lengths to argue their point in trying to convince others to convert. Others are less concerned (and perhaps more prone to changing?). As was pointed out before, the term employed in this book is *disfluency*, simply because it has been around for more than 40 years, and seems to be (by far) the most common term, irrespective of field of research. This does *not* mean that I necessarily regard it as a completely felicitous term, or that I reject all other suggested terms to be found in the literature. I do find terms like *discontinuity* more appropriate from certain angles, and that *Own Communication Management* is a good way to describe the ways in which speech *is* indeed "managed" (which is not always the case, however, as should be obvious from some of the preceding paragraphs). Suffice it so say that the reader is free to question the choice without me protesting heavily.

Concerning the terminology at more detailed levels, I basically hold similar views. Whether it be called a *filled pause* or *filler* (word), or whether it is grouped together with words like *well* in a category 'interjections', or whether *prolongations* and *cutoffs* belong in a category *dysrhythmic phonations*, or whether *hesitations* include both *silences* and *fillers* or not, and so on, is not the most important issue in my view. However, what I tend to find more frustrating is when it is not *clear* what a certain term refers to. So long as a term is well defined and delimited, I am fully prepared to live with it throughout an article or a book. I have chosen a set of terms to be used here, and I hope I have succeeded in defining them to the reader.

2.20 Concluding remarks

As we have seen in this chapter, what is here called **disfluency** can be studied from an extremely wide variety of perspectives, for a plethora of reasons, and give insight into a cornucopia of different fields of human behavior in general, and communication in particular. So, what then, would be the correct way of looking at it? Are disfluencies signs of "problems" in human speech production? Are they detrimental or beneficial to human (or machine) speech comprehension? Are they evidence of erroneous motor planning or execution? Is there a continuum from severe stuttering to the "perfect" rendering of human speech as defined within rhetorics? Are disfluencies merely signs of psychological breakdown or stress, of lesser or graver status?

Perhaps this is stating the obvious, but the stance taken here is that there is a point to *all* of the above approaches. As has (hopefully) been shown, there is no doubt that disfluency sometimes is detrimental, making it harder to process language for listeners. There is also compelling evidence that disfluency also reflects psychological stress, or errors in speech production, be it at the deepest level of the conceptualizer or later in the chain, in lexical retrieval or motor execution. On the other hand, there is also convincing evidence to the effect that much of what is called disfluency is actually sign of **fluency**, insofar as it beyond doubt is beneficial to human communication, and conveys not only linguistic information proper, but extra- or paralinguistic information of central importance in human–human interaction, where information transfer is not only constituted by linguistic-semantic units.

Much of the research described in this chapter is beyond reach in this study, for obvious reasons, such as not having had electrodes in the brains of our subjects, having excluded stutterers or non-native speakers from our data, being confined to mock-up telephone conversations within a very restricted domain, not having measured heart-rates or palmar sweat during the sessions, not having video-taped the subjects and so on and so forth ad infinitum. Moreover, phenomena such as developmental factors cannot be studied since children were not part of the data collection. This naturally delimits the number of studies that can be done on the data studied in this thesis.

However, other areas, such as the validity of speech production models, and in connection with this also models of human consciousness, could be elucidated given the corpora here studied. Phenomena such as categories, distribution, reaction times and so on could illuminate, corroborate, confirm or rebut certain aspects of features of specific proposals. This, however, would to a lesser or higher degree require additional analysis and labeling of our data, and will consequently not be included here. The same goes for communicative aspects such as speech act theory, which would require an analysis of both interlocutors in our dialogues, not only the subjects.

The goal of this chapter has been to show how much study has been devoted to disfluency over the decades, and also how speech disfluency is a valid object of study within a vast array of different disciplines, with a more or less direct interconnectedness. After having thus "set the stage" in that we now know—to some degree—what the phenomenon disfluency is, or can be, it is time to turn to the data studied in this thesis.

PART III

3 Data collection and corpora

This chapter describes the data that are the main object of study in this thesis. All four corpora are described with regard to method, subjects, channels and general design. Since data collection methodology *per se* highly affects the characteristics and quality of the data collected, and consequently whatever observations can be made based on the data to some degree (lesser or higher) reflect *method* rather than the desired object of study, a fairly detailed account is provided. Also, during previous work, it was shown that some methodological characteristics, or task details, did indeed affect the data (Bell, Eklund & Gustafson, 2000), and that more such effects might be there to be discovered.

All corpora were collected as part of a number of projects, outlined in **3.1**. The method used to collect the data—the so-called **Wizard-of-Oz** method—is described in **3.2**, while the corpora will be described in sections **3.3** through **3.6**. Section **3.7** describes how the data were post-processed. Section **3.8** briefly summarizes the total amount of data collected. Section **3.9** describes the eight subjects that participated in both the WOZ-2 and the Nymans corpora, enabling inter-corpus comparisons.

3.1 The Spoken Language Translator

The data used in this thesis were collected during three projects that ran during the period 1992 through 1999, all under the blanket name *The Spoken Language Translator* (SLT) project. The goal of the projects was to create a functional speech-to-speech translation system between Swedish and English within the ATIS domain (Hemphill et al., 1990), financed by the **Telia Networks Division** and **Telia Research AB**, and they were carried out by Telia Research AB (Sweden), **SRI International**, Menlo Park (California), SRI International, Cambridge (UK), the **Technical University of Crete** (Greece), and the **Swedish Institute of Computer Science** (SICS), Kista (Sweden). Besides creating functional speech-to-speech translation between American English and Swedish, SLT also yielded translation to and from the said languages and French, and from these three languages to Danish text.[1] A brief summary of the different parts of SLT will be given in the following passages. For a detailed description of the different parts of SLT, the reader is referred to Rayner et al. (2000).

[1] The work involving French was carried out as a separate project involving **SRI International** and **ISSCO/TIM**, funded by SRI and Suissetra. Work involving Danish was carried out as a separate project involving SRI International and Handelshøjskolen in Copenhagen, under internal funding from both parties. A Danish speech synthesizer was not included in the project.

One thing that must be borne in mind is that the data collected within SLT were not collected with the objective to study disfluencies, but to train automatic speech recognizers and to build language models. This probably has both pros and cons. The pros part could be that the incidence and frequency of disfluencies is more "natural" than it would have been if the tasks were designed so as to elicit disfluent speech, a method used within the field of speech production studies. The cons part could be that no pre-collection notions about the distribution of disfluencies, e.g. how they are correlated with certain speech acts, were controlled for.

3.1.1 SLT-1

The first SLT project was a one-year project that ran from mid-1992 to mid-1993, as a collaboration between **Telia Research AB** (Sweden), **SICS** (Sweden) and **SRI International** (California) and SRI International (UK). During SLT-1 a functional speech-to-text system was developed which translated from English to Swedish. Neither a Swedish recognizer nor a Swedish synthesizer were employed at this stage. No speech data were collected during this phase of the project. A full description of SLT-1 is given in Agnäs et al. (1993).

3.1.2 SLT-2

The second phase of the project, SLT-2, ran from mid-1994 to late 1997. During this phase, a Swedish recognizer was developed. During this part of the project, **SICS** left, and **ISSCO/TIM** (University of Geneva) joined—under independent funding—as did the **Technical University of Crete**. During SLT-2, a Swedish concatenative synthesizer was also added, although it was not developed as a formal part of SLT but as a separate project at Telia Research AB, Sweden. As a result, the project yielded a fully functioning speech-to-speech translation system between English and Swedish. The first speech data collection, WOZ-1, was carried out during this phase of the project. A full description of SLT-2 is given in Becket et al. (1997).

3.1.3 SLT-3 / Database

The third phase, SLT-3, started shortly after SLT-s, and ran until mid-1999, with the partners described in **3.1.2**. The domain was expanded from ATIS proper to business travel bookings, and the domain was also changed from American air travel bookings to full business travel booking in Sweden. The system employed a real travel database, *TravelLink*™ (see **References**)—used by professional travel agents). Three speech corpora were collected during this phase of the project: WOZ-2, Nymans and Bionic.

3.2 Human–machine communication: a short history

When building a system for human–machine communication there are at least two underlying issues to heed. The first, general, observation is that several factors affect how humans interact linguistically as interlocutors, depending on factors such as task, channel, role and who the (other) interlocutor is (e.g. human or machine). The second issue is that there is no way of telling or knowing *a priori* exactly how humans will interact as a function of the parameters listed above.[1] So what entails is that in order to build a well-functioning human–

[1] Sperry (1976) makes the following remark on animal research: "One of the earliest rules for animal behavior stated that, when rigorous conditions are established in which all sensory input can be strictly controlled, one

machine system, one simply needs to study the behavior of human–machine interaction before being able to tune the application in question to the behavior of its intended human users.

3.2.1 Interactive communication: early studies

The idea of conversing with an automatic system in unrestricted English was there, as we have seen, already in 1969 (Coles, 1969). Another pioneer was Malhotra (1975) who wanted "to create conditions in which the subjects could behave as naturally as possible, unhampered by technological restrictions" (Malhotra, 1975, p. 843). However, both Coles and Malhotra only studied typed communication where parts of the system were simulated. Coles focused on semantic and syntactic problems associated with natural language conversation, but Malhotra reports that most of his subjects commented that the system "would be very useful if it could be implemented" (Malhotra, 1975, p. 843).

The idea of conversing with computers in natural language was out of the box. So how would one go about creating such a system? Smith (1980), discussing Weizenbaum's simulated therapist ELIZA (Weizenbaum, 1967) points out the human tendency to anthropomorphize mechanical objects, covering basically everything "from cars to airplanes. However, this effect seems particularly strong with computers" (Smith, 1980, p. 13). A possible reason for this strong tendency is suggested by Turkel (*vid.* Smith, 1980, p. 13), who points out the epistemological irreducibility of computation, that computers have no obvious analogies in the real world (like comparing airplanes with birds), or as Smith puts it: "shorn of physical referents, people have to resort to attributing purposiveness and other human attributes to the computer".

So, it seemed that people both wanted to be able to converse with machine in natural language, and were willing to expect "human" behavior from them, but how would humans react to computers that were *speaking*?

Early research on interactive communication including a voice channel was carried out by Chapanis and colleagues (Chapanis, 1971, 1973, 1975, 1981; Chapanis, Ochsman, Parris & Weeks, 1972; Chapanis, Parrish, Ochsman & Weeks, 1977; Chapanis & Overbey, 1974; Ochsman & Chapanis, 1974; Stoll, Hoecker, Krueger & Chapanis, 1976; Weeks & Chapanis, 1976; Weeks, Kelly and Chapanis, 1974). They compared how humans communicated with other humans under four different channels—**body movements** (in this case, face-to-face communication), **speech**, (hand-)**writing** and **typing**—either singly, or in different combinations. The tasks were chosen so as to provide different psychological challenges to the subjects, but were all formulated so that their solutions required the joint efforts of two individuals collaborating, which they could do by using one or several of the channels listed above. Some of their findings are summarized below:

- Tasks were solved significantly faster if the interlocutors had access to speech than if they were not able to use speech. (Voice settings were about twice as fast as handwriting or typing settings.)

---

may predict for any measured stimulus that an animal will respond 'as it damn pleases.' " (Sperry, 1976, p. 9). Similarities to speech research are probably not coincidental.

- Tasks were not solved faster in face-to-face settings than in voice-only settings. The two settings were about the same speed. (Face-to-face was slightly, but not significantly, faster than voice-only.)

- Voice channels were much wordier than non-voice channels.

- Face-to-face settings were wordier than voice-only settings.

- The number of words used by the subject was not influenced by the channel available to the other interlocutor, solely by the channel available to the subject.

- Freedom to interrupt ("barge-in") did not affect the time it took to solve a problem, it only affected the way messages were packaged, with more and shorter messages if interruption was allowed.

- Interruption was far more likely to occur if the system had a voice channel.

Having observed that speech allowed for quicker problem solving, they noted that the language used did not obey the grammar rules taught in school, but indeed was "extremely unruly and often seems to follow few grammatical, syntactic, or semantic rules" (Chapanis, 1981, p. 106), Chapanis almost lamented

> Most people know that ordinary communication tends to be somewhat disorganized, but few of us really appreciate how disorganized it can be. (Chapanis, 1981, p. 106.)

On a slightly more positive note, Chapanis then continues:

> If we are ever to have computers that can interact with their human counterparts in natural English, by typewriter, by voice, or by handwriting, we will somehow have to discover at least some of the rules that apply to natural, unconstrained communication. Discovering those rules is, in my opinion, one of the most fascinating and challenging problems facing both basic and applied scientists in this area of man–computer interaction. (Chapanis, 1981, p. 111.)

Also, Cohen (1984) compared spoken and keyboard communication in instruction giving and found that speakers aimed for more detailed goals in the spoken setting than when using keyboards. Moreover, those goals were expressed "indirectly" by dint of utterances where the surface form did not explicitly convey the speakers' intent. He concluded that "intent recognition will need to be a central focus for pragmatics/discourse components of future speech understanding systems" (Cohen, 1984, p. 97).

So, not very surprisingly, the more the situation resembled a human–human communication setting, the faster the tasks were solved. But there was still a catch here: the human subjects in the previously mentioned studies knew they were not *actually* communicating with a computer, at least not more than in a very primitive and rudimentary way, and that the systems they were conversing with were in fact run by humans. Chapanis et al. had shown that channel matters, could it be the case that interlocutor also mattered? Perhaps it is the case that humans not *necessarily* communicate with machines in exactly the same way they would if the other conversant were another human being?

To answer that question, one simply needs to study the behavior of humans interacting with a real, up-and-running automatic system. After having done that, one can create a system which will be tuned to the *actual* behavior of humans interacting with a computer. The obvious problem here is of course that one cannot study human–machine interaction before such a system exists, leading to a cumbersome "Catch 22": *One needs to have the system to collect the data one needs to build the system*. So, how does one proceed? The obvious solution to the problem is to *pretend* one has the system, and this method has also been used since the 1970s, in various forms.

3.2.2 Wizard-of-Oz simulations

The method of "deceiving" users by having them interact with a mock-up system was first named *PNAMBIC*, short for "*P*ay *N*o *A*ttention to the *M*an *B*eh*I*nd the *C*urtain", by Klein (1981). The phrase was taken from the novel "The Wizard of Oz" (Baum 1900). Kelley (1983, 1984) refers to the technique as the "OZ paradigm", which includes an experimenter, acting as "wizard", and the technique has thenceforwards become known as the **Wizard of Oz**, or **WOZ**, technique to collect data.

Early WOZ simulations studied human–machine interaction using keyboards (Kelley 1983, 1984; Dahlbäck & Jönsson 1986, 1988; Jönsson & Dahlbäck 1988; Reilly 1987; Kennedy et al. 1988; Peckham, 1990; Dahlbäck, Jönsson & Ahrenberg 1993), while others compared typed input with spoken input (Beun & Bunt 1987; Guindon 1988; Hauptmann & Rudnicky 1988). Fully oral WOZ simulations were conducted by Richards & Underwood (1984, 1985), Delomier, Meunier & Morel (1989) and Amalberti, Carbonell & Falzon (1993), just to mention a few. Hauptmann (1989) went even further and compared gestural input, voice-only input, and a combination of the two. He concluded that almost 60% of the subjects preferred to use a combination of the two modalities. Studies have also been devoted to differences between linguistic behavior of the users as a function of the interlocutor, i.e., whether the conversation partner is a real human being or is (believed to be) a computer (e.g. Morel, 1989). For a review of early WOZ methodology, the reader is referred to Fraser & Gilbert (1991).

So, do WOZ simulations solve the problem? Critical voices have been heard. Tennant (1979) voices the opinion that instead of focusing on *linguistic* coverage, the problem lies in the description of *conceptual* coverage, and that more work should be devoted to the latter. Newell (1984, 1989) was critical to the idea that human–human conversation should serve as the basis for human–machine conversation, mainly since the communicative situation is different. Another critical voice was von Hahn (1986), who claimed that "[e]vidence from mock-up systems, simulated by persons, is methodologically vague and mostly too isolated from real application" (von Hahn, 1986, p. 523), and similar views were held by Dahlbäck (1995) and Whiteside, Bennett & Holtblatt (1988), who pointed out that: "In the laboratory, subjects perform tasks prescribed by the experimenter. In the workplace, people perform tasks important to their careers and livelihood" (Whiteside, Bennett & Holtzblatt, 1988, p. 806). In the same vein, perhaps the most interesting critical comments on WOZ methodology are found in Allwood & Haglund (1992), who pointed out that in a WOZ simulation, both the subjects and the wizard(s) are still playing roles, occupied and assigned. The researcher acting as the wizard is occupying the role of a researcher interested in obtaining "as natural as possible" language and speech data, while playing the role of the system. The subject, on the other hand, is occupying the role of a subject in a scientific study, and playing the role of a client (or similar), communicating with a system while carrying out tasks that are not genuine

to the subject, but given to them by the experiment leader (who might be identical with the wizard). I will not dive deeper into this discussion here, but I concur that the critical points raised by Allwood & Haglund are indeed relevant.

As we have seen, there are claims that the social situation when speaking with a machine is different from normal conversation between humans, and also that the subjects in a simulation are acting out assigned roles. Thus, it would seem obvious, for instance, that humans would feel no need to be polite to computers. However, things seem to be more complicated than that. Reeves & Nass (1996) and Nass & Moon (2000) have shown that humans indeed *are* polite to computers under certain circumstances, and their more general claim is that humans, being basically social creatures, interact with media (TV, computers and the like), in more or less the same way they interact with other human beings. Consequently, everything we know about human (interactional) behavior in general can—in theory—be ported to human behavior with machines. I won't take a stand here on the issue whether or not the linguistic, or other, behavior of computers should try to mimic human behavior completely, or whether or not it is preferable to make it clear to users, in one way or another, that they are indeed communicating with a machine. Suffice to say here that lacking an up-and-running system, WOZ simulations do come in handy, since they come as close to human, or machine, behavior as one wishes to present to the users, while at the same time making it possible to compare subject behavior in different settings.

A final point to make here is that user behavior is hard to predict, and that simulation design should be as open as possible if one wants unconstrained data (*vid.* Furnas et al., 1987). Consequently, it is crucial to present the task to the users/subjects in ways that do not affect the linguistic behavior of the subjects in unwanted ways. It has been shown that written instructions tend to govern the linguistic behavior of the users (MacDermid & Goldstein 1996), so in order to avoid linguistic biasing, MacDermid & Goldstein (1996) proposed what they called the **Storyboard Method**, where instructions/command are given in entirely iconic form. This is also the method employed in three of the data collections that are studied in this work (WOZ-2, Nymans and Bionic), while the first data collection (WOZ-1) employed a method that combined iconic and written instructions.

3.3 WOZ-1 / human–"machine"–human (ATIS)

3.3.1 Introduction

In order to create a Swedish recognizer, one of the first concerns of SLT-2 was to collect Swedish language and speech data. A first a shot at obtaining **language data**, needed for language processing, was to have a large number of Swedes translate American ATIS sentences (Bretan et al., 2000; Bretan, Eklund & MacDermid, 1996). Basic **speech data** were collected at around 40 different locations around Sweden, and consisted of various sets of sentences in order to cover Swedish phonemes, allophones, phonological processes, as well as ATIS sentences, unique text (for each speaker) and so on. The collection of Swedish speech data is described in Eklund et al. (2000). The data thus collected, however, consisted entirely of read material, and in order to obtain more authentic Swedish data (both language and speech), it was decided to carry out a WOZ simulation, which was done in early 1996.

3.3.2 Goal

The goal of the WOZ-1 simulation was to collect between 3000 and 5000 Swedish utterances. A full description is given in MacDermid & Eklund (1996).

3.3.3 Scenario

The subjects were given ten tasks each. They were to book travels and thus needed to gather information concerning cities, hours, dates, prices and so on. The tasks were given in both written and iconic form. A task sheet is shown in **Plate 3.1**.

3.3.4 Subjects

The subjects were all Telia employees, working at Telia Data AB. A total of 52 subjects participated, six of whom have been omitted from analysis, either for technical reasons, or because the subject in question was not a native speaker of Swedish. The age span was between 18 and 55. Some of the subjects had never booked a business trip, whereas some usually did it twice a month. The subjects were not familiar with speech technology in any way. Post-collection evaluation showed that 34 subjects believed they had been talking with a computer, or a pre-recorded taped voice, while 10 subjects suspected they had been talking with a human. Four subjects were not sure, and four subjects did not understand the question.

3.3.5 Set-up

The subjects were told that they were talking on the phone with human travel agents in England, Germany or France who did not speak or understand Swedish. The utterances made by the subjects, however, were translated by a computer into English, German or French and read out by the same computer to the agent at the other end of the phone line. All parts of the "system" were carried out by humans, who impersonated speech recognition, understanding, translation and synthesis. The only part of the system that was "real" was that actual travel data were collected over the phone.

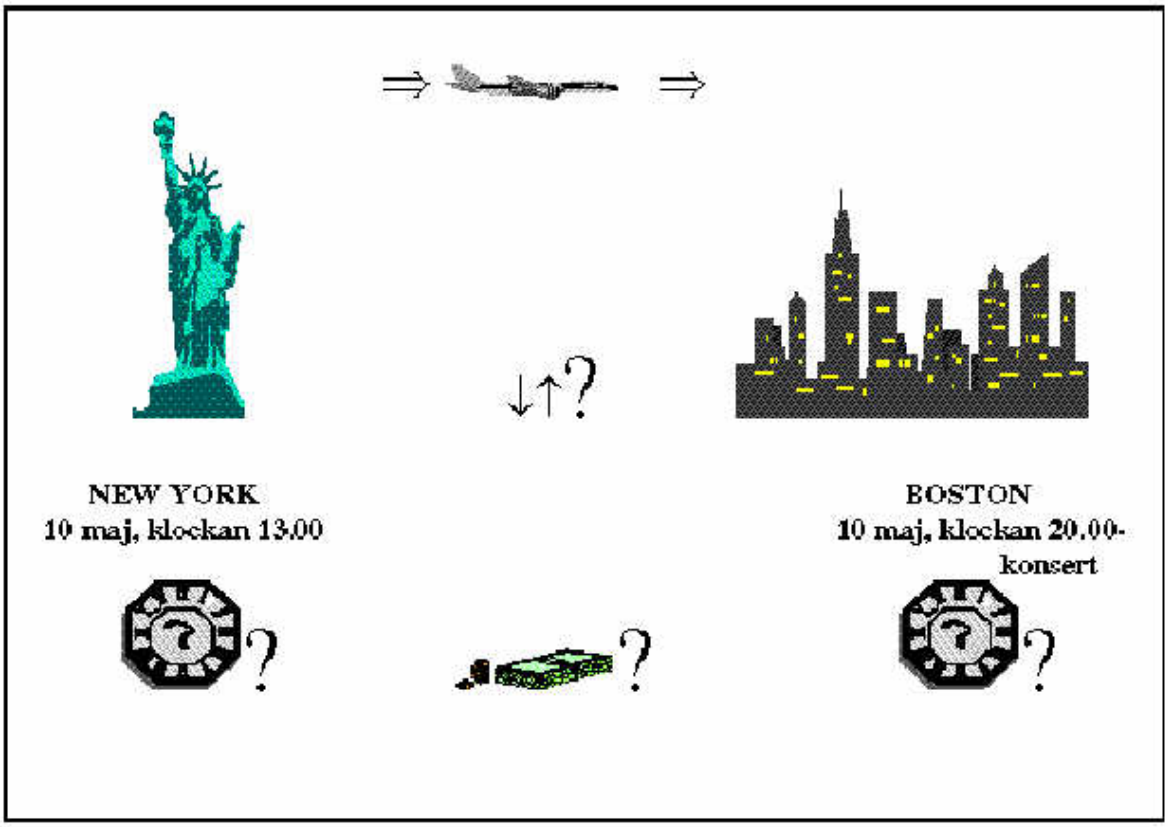
Two wizards were used, Wizard 1 and Wizard 2. The subject talked to Wizard 1, who made some modifications to the utterance (mainly filtering and simplification), and passed it on to Wizard 2, who in turn passed it on to a real travel agent over a phone line. The same procedure was then reversed, so that Wizard 2 modified the travel agent's response, and passed it on to Wizard 1 who provided the subject with the information.

Wizard 1 and Wizard 2 could talk freely with each other, in case of unclear information, in which case a mute button on the handset of the phone was used, lest the subject hear what was being said. A behavioral psychologist and a computer scientist took turns in serving as Wizard 2 (i.e., the "wizard" interface of the travel agent). Both were familiar with WOZ simulations.

Similarly, two professional actors took turns in playing the role of Wizard 1. In order to make the simulation believable, the two actors were given some training in learning how to imitate computer speech.[1]

[1] For instance, Richards & Underwood (1984) artificially distorted the wizard's voice by using a vocoder. Since the quality of speech synthesizers had improved considerably, this was not deemed necessary.

Bokning 2) Efter en knapp vecka i New York visar det sig att din favoritartist ska spela i Boston den 10 maj klockan 20.00. Du kan resa efter klockan 13.00 den 10 maj. Du undrar över tider, om det är några stopp på vägen till Boston och dessutom vill du resa så billigt som möjligt. *Ring och boka resan!*



Resplan: _____

_____

_____

_____

_____

Plate 3.1. WOZ-1 task sheet number two (out of ten different task sheets used). The text reads: "After a week in New York you are told that your favorite artist is going to perform in Boston May 10th, at eight o'clock in the evening. You want to know the times, whether there are any stopovers on the way to Boston, and you would also like to travel as inexpensively as possible. *Make the call and do the booking!*".

Their task was to read out the three utterances:

> *Välkommen till flygresebyråns översättningstjänst, var god dröj.*
> "Welcome to the Air Travel Agency's translation service, please hold on."
>
> *Var god repetera.*
> "Please repeat."
>
> *Kan du formulera dig kortare.*
> "Could you rephrase that and make it shorter."

Besides these three utterances, Wizard 1 only repeated what Wizard 2 or the client said.

Wizard 1 was instructed to reject utterances longer than twenty words, and use one of the two appropriate phrases above to ask the subject to rephrase the utterance just made. Apart from these three sentences, Wizard 1 only repeated the utterances of the subjects or Wizard 2. The utterances of the travel agents were based on interviews with professional travel agents at the travel agency Nyman & Schultz, at the time the travel agency employed by Telia Research AB for business travel. The set-up is given in **Figure 3.1**.

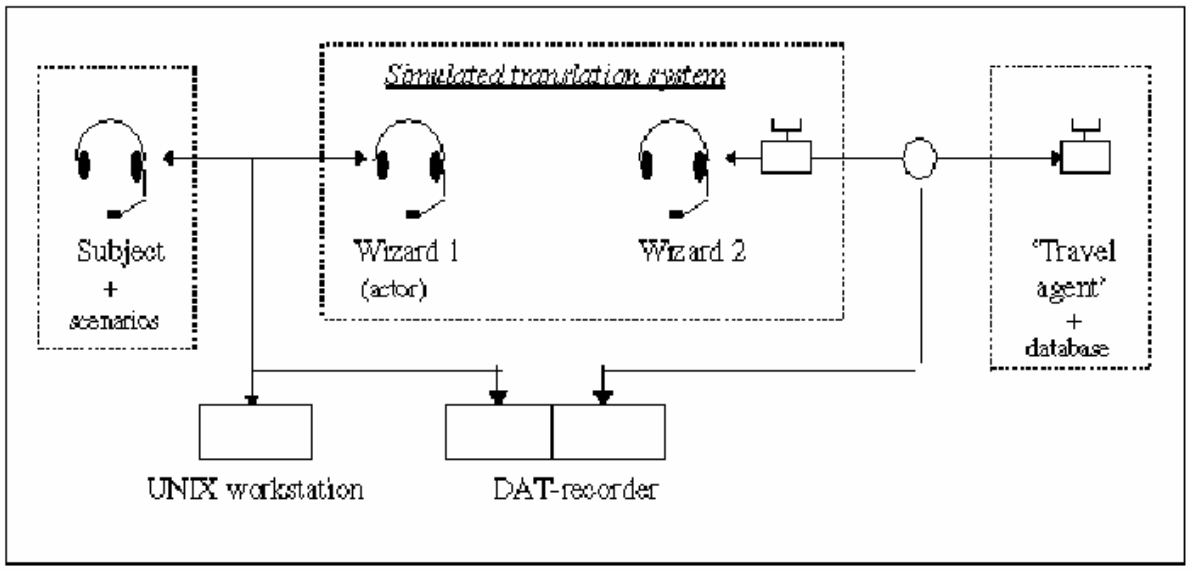


Figure 3.1. WOZ-1 set-up. The subject called a simulated travel booking system. An actor simulated speech recognition/synthesis. Wizard 1 (W1) passed the request made by the subject on to the second Wizard (W2), who called a travel agent. The travel agent collected the information required, and passed it on to W2, who provided W1 with the information, who then communicated the information to the subject. The dialogue between the subject and W1 was recorded by the SRI *Generic Recording Tool*, and was stored on disk using a Sun Sparc 5 work station. The dialogue was also recorded on a DAT recorder. The dialogue between W2 and the travel agent was also recorded by tapping the telephone line and was stored on a DAT tape.

3.3.6 Equipment

All participants were recorded on a Tascam DAT recorder, with the exception of the communication between W1 and W2. The dialogue between the subject and W1 was recorded with SRI's *Generic Recording Tool* and stored as digital sound files on a Sun Sparc 5 work station.

The subjects and the wizards wore headsets. The wizards had the possibility to disconnect the subject and the travel agent using a mute button when they wanted to talk to each other. The travel agent used a normal, landline telephone.

3.3.7 Data collected

Most subjects carried out ten tasks each (some fewer, some more than ten). Of the data collected, a total of 433 dialogues and 4022 subject utterances have been analyzed in the present work. (For a full description of the data collected, see **Appendix 1 WOZ-1**.) In time, this amounts to **212 minutes** (>3.5 hours) of speech, all between-utterance silences being excised. The shortest utterance was 0.22 seconds long, the longest 20.13 seconds long, with a mean value of 3.16 seconds.

3.4 WOZ-2 / human–"machine" (business travel)

3.4.1 Introduction

During the third phase of the SLT project (SLT-3/Database), it was decided to expand the vocabulary and domain of the project to include not only (American) ATIS data, but to cover full business travel bookings. Also, it was decided that the tasks should focus on Swedish locations (rather than American locations), flights and carriers and so forth. Thus, "new" things to be covered included hotel reservations, car rental train tickets. WOZ-2 was carried out in June 1997.

3.4.2 Goal

The goal was to collect dialogue speech data between users and a simulated database application. Once again, all data were collected using a landline telephone. A detailed description is given in MacDermid & Eklund (1997).

3.4.3 Scenario

The subjects were asked to book business trips within Sweden by calling a computerized booking service. Thee were told that the system could handle booking of flights, trains, hotels, rental cars, taxis and so on. The system did not allow barge-in by the user.

Each subject was given three tasks. In order to avoid linguistic biasing on the subjects' wordings—the so-called ***script effect*** noticed during WOZ-1, when subjects "copy" text off the task sheet and use these wordings verbatim—the tasks were given in more or less purely pictorial form, using maps with information (MacDermid & Goldstein, 1996). Another reason for providing the tasks without verbal instructions is that pictorial tasks are inherently ambiguous, which was considered an advantage here, since this was thought to create further variation in the data. Each task was presented in the form of a map of Sweden with destinations, mode of transport, dates and times indicated. The indication of times was varied so that certain task sheets indicated times as e.g. *20:32*, other task sheets gave times as *9 p.m.* while other tasks sheets showed a picture of a the face of a clock whose hands gave the time. In a similar manner, dates were varied between the formats *1997-06-06*, *6 juni*, *6/6* and a picture of a calendar. All this was done in order to obtain linguistic variation. A task sheet is shown in **Plate 3.2**.

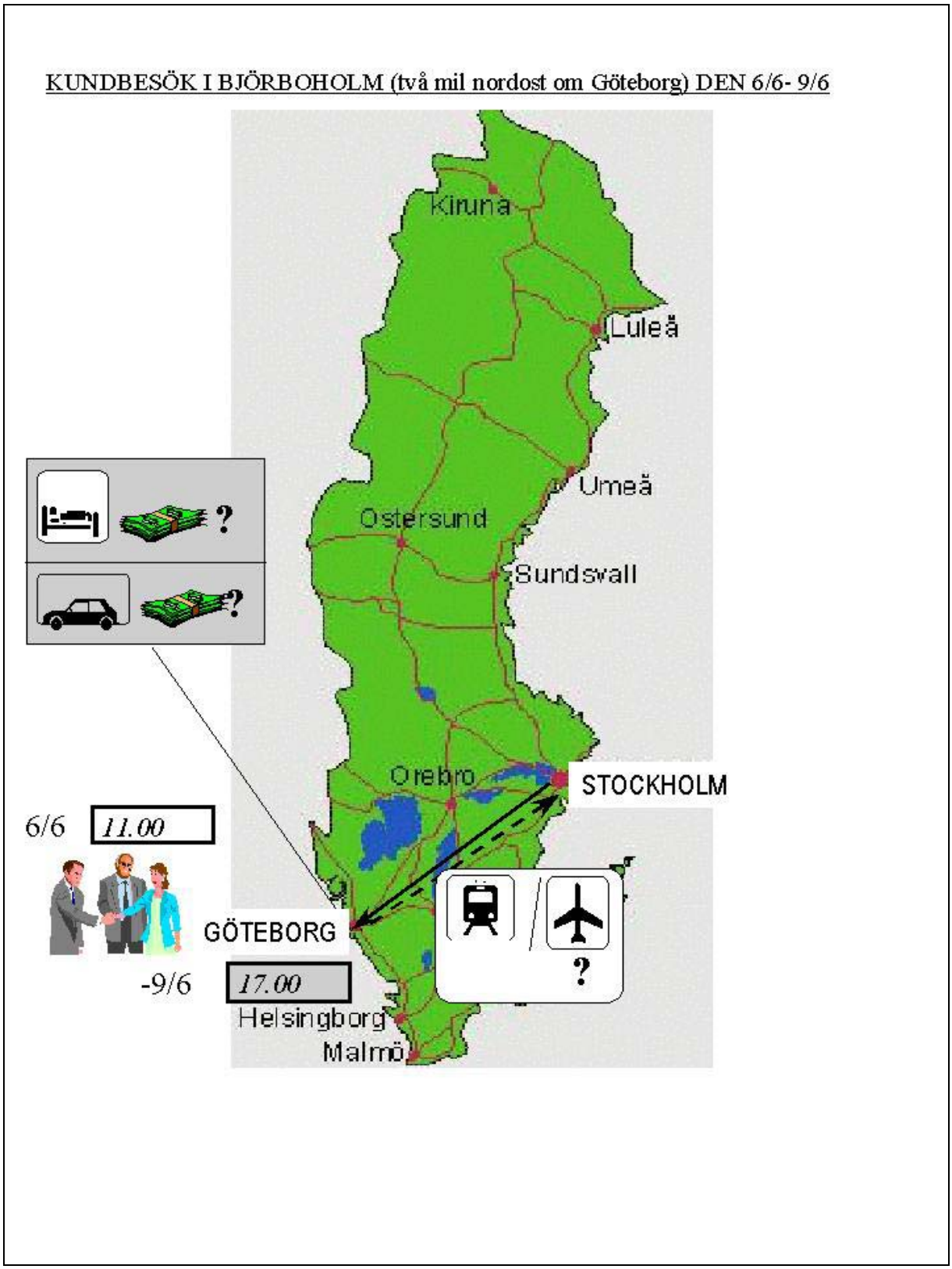


Plate 3.2. WOZ-2 task sheet number one (of three). "Client visit in Björboholm (two Swedish miles northeast of Göteborg), June 6th to June 9th". Note the icon used to indicate client meeting (three people talking). Arrows indicate departure and arrival. Icons indicate car rental. Question marks indicate that the subject should makes inquiries concerning relative prices between flights and trains, as well as car rental costs and hotel prices. Note that, on this particular task sheet, dates are given in number–slash–number format (*9/6*), and that hours are given in figure–dot–figure format (*11.00*).

In order to obtain even greater linguistic variation, the subjects were also told that they were free to book the trips according to their own preferences (small or big cars and so on). Some subjects were told that they had a "generous budget" while other subjects were told that they were on a tight budget. The tasks typically included booking of flights or trains (the choice was given to the subjects), car rental, taxi information and hotel bookings. The subjects were informed that the system (i.e., the wizard/actor) had a dialogue history, i.e., that it could remember dates, towns already mentioned by the subject and so on. The system had only little information concerning the exact location of hotels and other locations. If there was no good match to the subjects' specifications, the system used a "next-best" alternative and suggested another, suitable alternative, e.g., a departure time the preceding evening (so-called ***constraint relaxation***). In case of communication breakdown, the system was able to initiate repair sub-dialogues.

3.4.4 Subjects

Forty-nine subjects took part in the simulation. All subjects were Telia employees (Telia Data AB), and had no previous experience with speech technology. They had all booked at least one business trip before, either for themselves or for someone else. They were all native speakers of Swedish, apart from one speaker who was omitted for that reason. Two other subjects were omitted due to technical reasons (the DAT recorder did not record anything), thus leaving 46 subjects (32 male, 14 female). The data collection report (MacDermid & Eklund 1997) does not include any information on interviews with the subjects. Although the subjects were given basically the same questions as in WOZ-1 (MacDermid & Eklund 1996), the replies were not included in the report since they did not differ significantly from the corresponding replies given in WOZ-1, and consequently were not deemed central in the internal report where all potential readers were assumed to have read the WOZ-1 report.[1]

3.4.5 Set-up

The subjects called the simulated system using a landline telephone. The simulated speech recognizer/synthesis (Wizard 1) answered and asked the subjects for information. Wizard 1 then passed the information on to a second wizard (Wizard 2), who consulted a set of authentic web databases to gather the requested information. The reason for using two separate wizards was to limit the cognitive load on Wizard 1, since a previous pilot simulation at Telia Research AB had shown that it was difficult for one single person to both impersonate the computer and collect the asked-for information (Fraser & Gilbert, 1991) After having collected the relevant information, Wizard 2 then provided Wizard 1 with the information, who in turn passed the information on to the subject. Both wizards could hear the subject. The two wizards could discuss with each other, in which case mute buttons on the telephones were used to cut off the subject. The utterances of Wizard 1 were highly scripted (using fixed wordings) to strengthen the impression that the subjects were interacting with a computer. The set-up is shown in **Figure 3.2**.

[1] Catriona Chaplin, née MacDermid, personal communication.

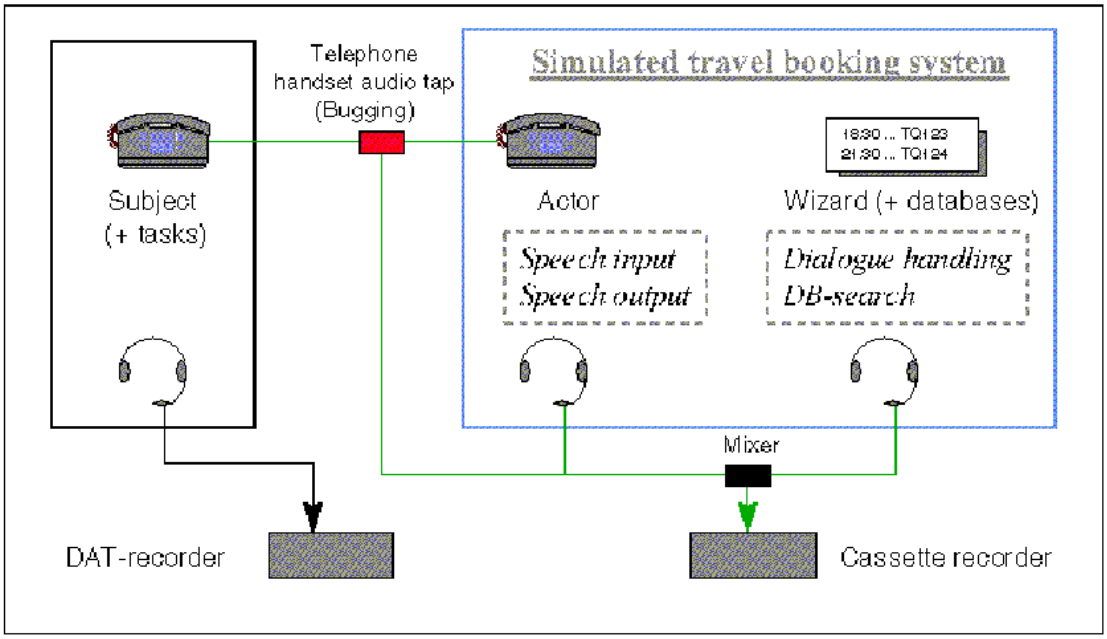


Figure 3.2. Set-up for WOZ-2. The subject called a simulated system. A professional actor (Wizard 1) acted speech recognizer/synthesis. Wizard 1 passed the subject's inquires on to a second wizard (Wizard 2) who consulted web databases to find the information asked for. The information was then passed back to Wizard 1 who communicated the information to the subject. The subject was recorded hi-fi on a DAT recorder. The entire dialogue was also recorded on an analogue cassette recorder by bugging the telephone line. A mixer was used to include the conversation between the wizards. Conversation between the wizards could not be heard by the subjects.

3.4.6 Equipment

The subjects were recorded hi-fi on a DAT recorder. The full dialogues were also recorded on an analogue cassette recorder by bugging the telephone line. A mixer was employed to include the communication between the two wizards.

3.4.7 Data collected

A total of 137 dialogues and 3436 subject utterances have been analyzed in this work. (For a full description of the data collected, see **Appendix 2 WOZ-2**.) In time, this amounts to **270 minutes** (>4.5 hours) of speech, all between-utterance silences being removed. The shortest utterance was 0.55 seconds long, the longest 45.47 seconds long, with a mean value of 4.93 seconds.

3.5 Nymans / human–human (business travel)

3.5.1 Introduction

During the project, some of the participating researchers felt that baseline data were needed, in order to see in what (potential) ways people altered their behavior when speaking with (what they believed were) a machine on the phone.[1] Consequently, a smaller corpus of human–human dialogues was collected, using the same task sheets as had been employed in

[1] One phenomenon proving that this assumption was indeed correct is the fact that the subjects all made use of ingressive speech in the human–human setting, and not at all in the human–machine setting (Eklund 2002).

WOZ-2. Also, given the wider domain within the SLT-3/DB project, there was the need to collect both speech and language data covering not only ATIS utterances, but also business trip booking utterances. To that end, it was felt that human–human dialogues could provide some insights into what kind of differences this expanded domain entailed, which perhaps was missed in the WOZ-2 corpus. The Nymans corpus was collected during the days 10–11 December 1997.

3.5.2 Goal

The goal of the human–human corpus was to collect a sufficient number of "authentic" dialogues between subjects and real travel agents.

3.5.3 Scenario

The scenarios were the same as in WOZ-2. An example is shown in **Plate 3.2**.

3.5.4 Subjects

Eight subjects participated, all of whom were Telia employees. They were all native speakers of Swedish, and had previous experience of travel bookings. In order to facilitate comparisons between human–human and human–"machine" data, all eight subjects (six male/two female) had participated in WOZ-2. Since more than six months had passed since the WOZ-2 collection, one can assume that no palpable learning effects were at play, at least not more than the later dialogues in WOZ-2 would exhibit. No post-collection interviews were carried out with the subjects since there was no need given that the conversations were all human–human.

3.5.5 Travel agents

Two professional travel agents—one male and one female—at Nyman & Schultz in Haninge, Sweden, were asked to participate, to which they agreed without any form of reimbursement. They were informed about the goal and nature of the tasks by the author (who knew them personally from his own business trip bookings) and two behavioral psychologists. The agents were instructed to behave "as natural as possible", with the exception that they were asked to deliberately misunderstand some of the utterances of the callers, in order to elicit linguistic data otherwise not obtained (on both grammatical and prosodic levels).

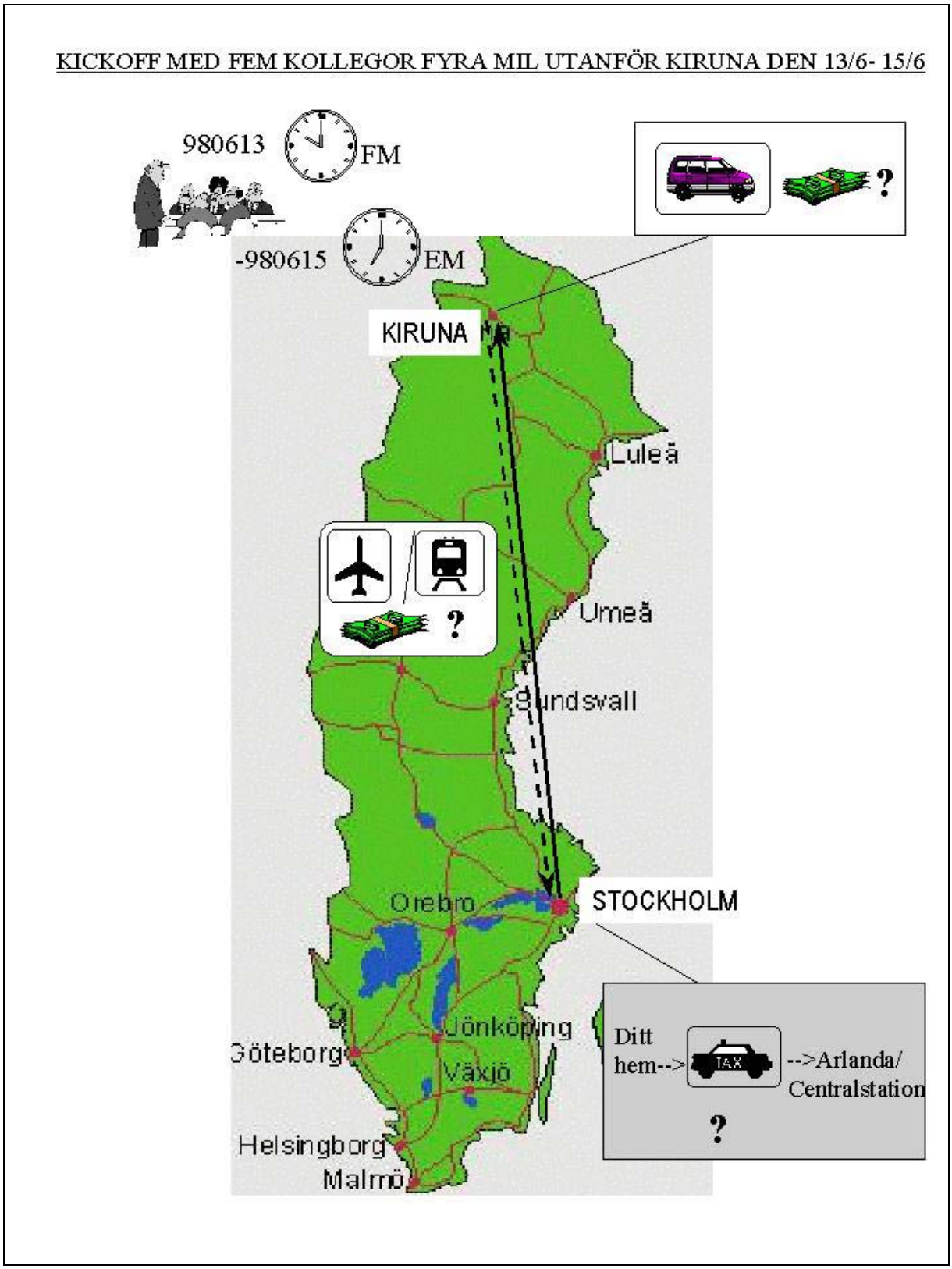


Plate 3.3. Nymans task sheet number two (of three). "Kick-off meeting four Swedish miles north of Kiruna, June 13 to June 15". Note the icon used to indicate the meeting (people around a table). Arrows indicate departure and arrival. Icons indicate car rental. Question marks indicate that the subject should makes inquiries concerning relative prices between flights and trains, as well as car rental costs and hotel prices. Note that dates are given in six-figure format (*980613*), and that hours are given as clock faces.

3.5.6 Set-up

The subjects called the travel agents on a special telephone line set aside for the collection, in order to avoid the risk of them being connected to another travel agent, who was not part of the corpus collection. (In normal cases, the same telephone number is used for all callers, who are lined up in a queue and given the first available travel agent.) The subjects and travel agents were recorded separately on DAT recorders to obtain hi-fi data needed for pitch extraction. The entire dialogues were recorded by tapping the telephone line signal with an analogue cassette recorder. The set-up is shown in **Figure 3.3**.

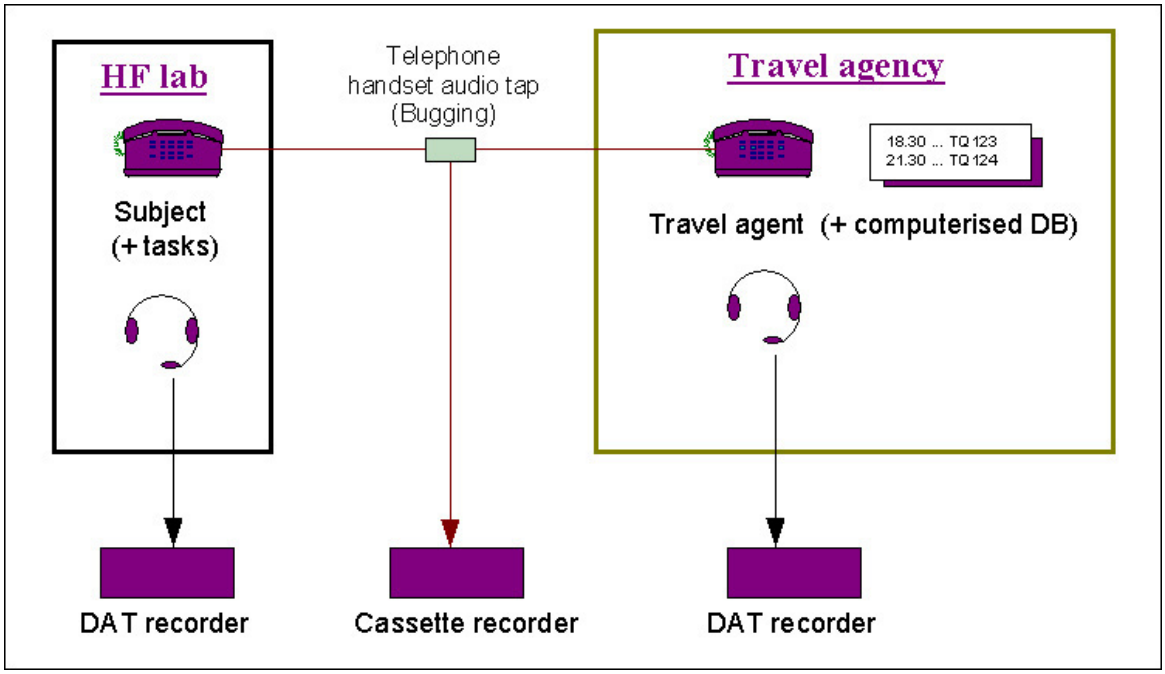


Figure 3.3. Nymans (human–human) set-up. The subjects called a travel agent on a reserved phone line. Both subjects and travel agents were independently recorded hi-fi on DAT recorders. The dialogue was recorded on a cassette recorder by bugging the phone line.

3.5.7 Equipment

Both the subjects and the agents used headset telephones. The subjects and agents were recorded separately using Tascam DAT recorders to obtain hi-fi recordings. An analogue tape recorder was used to cover the full dialogues by bugging the telephone line.

3.5.8 Data Collected

A total of 24 dialogues and 1734 subject utterances have been analyzed in this work. (For a full description of the data collected, see **Appendix 3 Nymans**.) In time, this amounts to **72 minutes** of speech, all between-utterance silences being removed. The shortest utterance was 0.26 seconds long, the longest 34.24 seconds long, with a mean value of 2.48 seconds.

3.6 Bionic / human–machine (business travel)

3.6.1 Introduction

As a result of the previous data collections, and as a general result of the SLT project, a live system was developed. In order to study authentic human–machine data, it was decided to compile a corpus of dialogues between users and the "state-of-the-art" system that was up and running at the time, christened the *bionic* corpus, denoting that authentic systems components were used to the extent that it was technically feasible (Fraser & Gilbert, 1991). However, since the Swedish recognizer at this stage did not cover some of the Swedish city names used in the tasks, a wizard (a computer scientist working at Telia Research AB) was used to simulate recognition. The Bionic corpus was collected in April 1998.

3.6.2 Goal

The goal of the data collection was to collect a sufficient number of authentic human–machine data. Another goal was also to test various dialogue management strategies, e.g. implicit confirmations and the like.

3.6.3 Scenario

The scenarios were the same as in WOZ-2 and the Nymans corpora. An example is given in **Plate 3.4**. Each subject was given four tasks. Some subjects conflated tasks while some did not complete all five tasks within the time frame. Two subjects did five tasks.

3.6.4 Subjects

Sixteen subjects participated (nine male/seven female), all of whom were Telia Employees (Telia Data AB). In this corpus, regional dialects were allowed. All subjects were used to business trip bookings and had no previous experience with speech technology. No post-collection interviews were carried out with the subjects since there was no need given that the system's voice was an authentic computer voice.

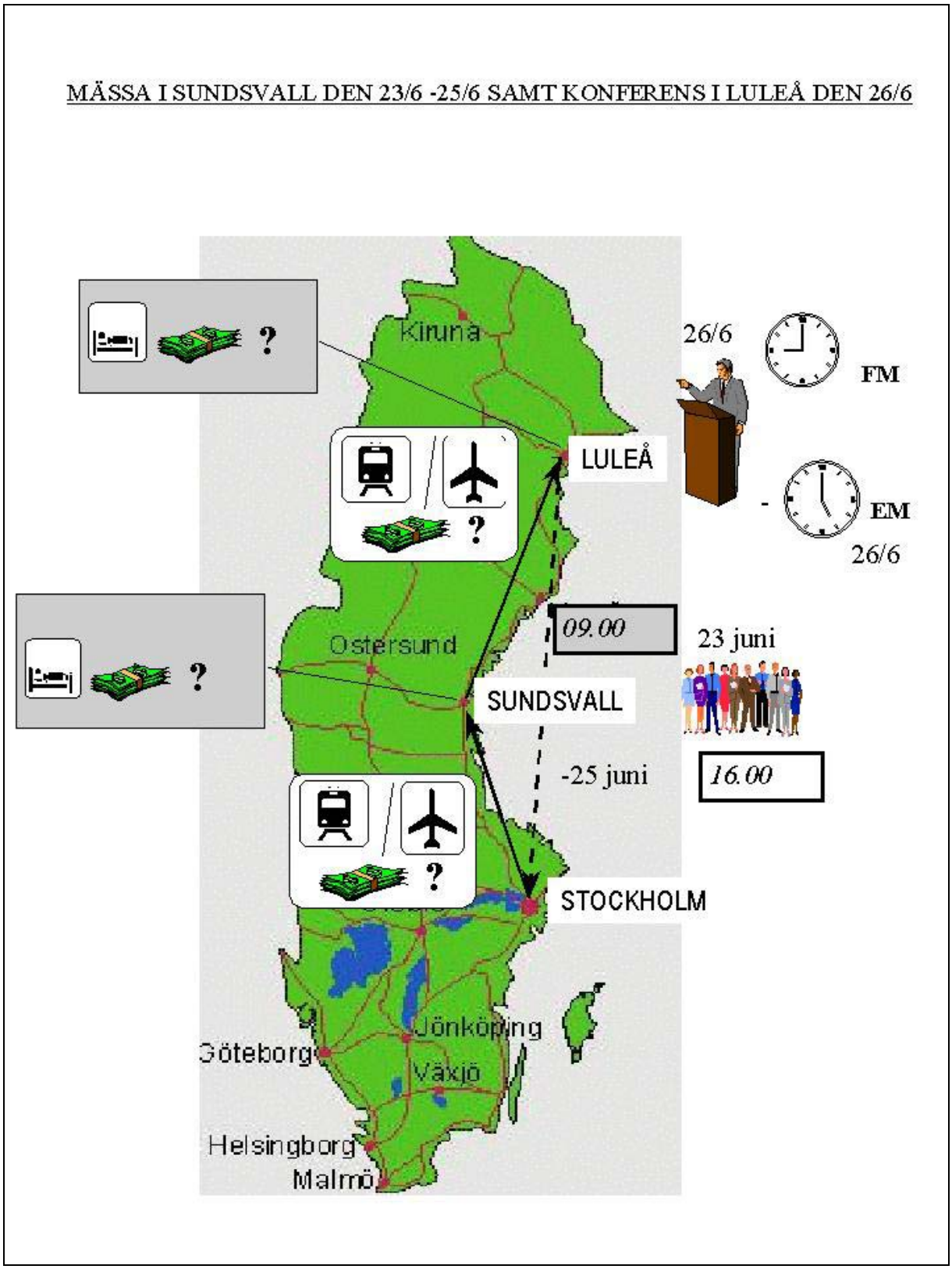


Plate 3.4. Bionic Task Sheet. Task sheet number one (of three). "Exhibition in Sundsvall, June 23 to June 25, and conference in Luleå, June 26". Note icons used to indicate exhibition (group of people) and conference (person standing behind rostrum). Arrows indicate departure and arrival. Icons indicate car rental. Question marks indicate that the subject should makes inquiries concerning relative prices between flights and trains, as well as car rental costs and hotel prices. Note that dates are given in number–slash–number format (*23/6*), and that hours are given as clock icons.

3.6.5 Set-Up

The subjects called a simulated travel booking application. The wizard simulated speech recognition, summarized the user's utterance into a "condensed form" and typed the abbreviated utterance into the (automatic) dialogue manager. For example, if the user said (in English translation):

> "Hello, my name is *N N*, and I would like to go to Sundsvall next Thursday and I would like to go there as early as possible"

… the wizard would type (in English translation):

> "from Stockholm to Sundsvall Thursday 05–10"

… or something to that effect. The dialogue manager then suggested the next move in the dialogue, which was either accepted by the wizard or rejected in favor of another alternative. The selected move was then spoken to the user using authentic speech synthesis. All utterances had been prerecorded and were stored on disk as `.raw` files. An authentic database, *TravelLink*™ (**www.travellink.se**), was used by the system to collect authentic travel data. The database properly included the *Amadeus* (**www.amadeus.net**) flight database, and also covered hotels in Sweden. A glitch in this data collection—later realized during the transcription and labeling phase—was that the wizard had not been instructed to have an upper limit concerning utterance length as to what he would accept from the subject (something which was included in the instructions to the wizards in WOZ-1 and WOZ-2). Consequently, since the system accepted whatever was communicated to it, irrespective of utterance length, the resulting corpus includes extra-ordinarily long sentences, all of which were processed by the system. The set-up is shown in **Figure 3.4**.

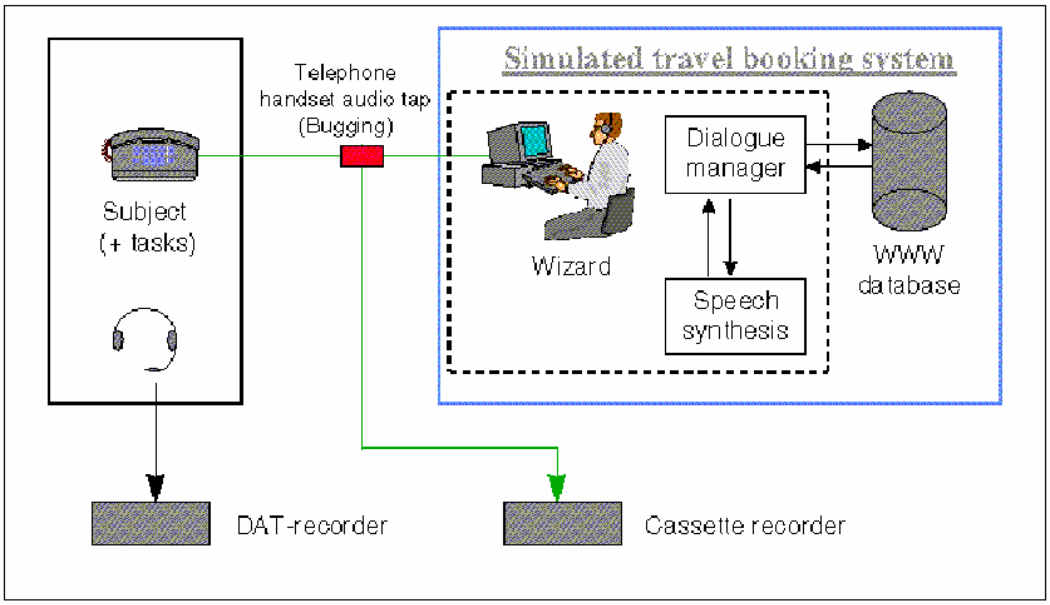


Figure 3.4. Bionic set-up. The subject called a simulated application. A wizard simulated speech recognition, and typed a filtered/abbreviated version of the subject's utterance into the dialogue manager. The dialogue manager consulted the (authentic) *TravelLink*™ web database, and suggested a move to the wizard, who either accepted the proposed move, or rejected it in favor of an alternative move. The selected move was then spoken to the subject using a real synthesizer. The synthesizer's utterances were all pre-recorded and stored on disk. The subject was recorded hi-fi on a Tascam DAT recorder. The full dialogue was recorded on an analogue tape recorder.

3.6.6 Equipment

Both the subjects and the agents used headset telephones. The subjects and agents were recorded separately using Tascam DAT recorders to obtain hi-fi recordings. An analogue tape recorder was used to cover the full dialogues by bugging the telephone line.

3.6.7 Data Collected

A total of 67 dialogues and 1985 subject utterances have been analyzed in this work. (For a full description of the data collected, see **Appendix 4 Bionic**.) In time, this amounts to **128 minutes** of speech, all between-utterance silences being removed. The shortest utterance was 0.20 seconds long, the longest 35.52 seconds long, with a mean value of 3.89 seconds.

3.7 Post-processing

All data described in the previous sections were post-processed in the following way.

3.7.1 Storage

All data were stored in NIST/Sphere format on Unix disks. The sampling rate was 16 kHz.

3.7.2 Transcription

WOZ-1, WOZ-2 and Nymans were first transcribed by a transcription bureau. However, as it became apparent that these transcriptions were not fine-grained enough, and thus sub-optimal for speech recognizer training and finer phonetic purposes, the transcriptions were all modified by a group of linguists/phoneticians who also did a lot of initial labeling of the data.

Since the aforementioned transcriptions did not suffice for the analyses carried out in this work, all data were re-transcribed and labeled separately, which is described in **chapter 4**. The bulk of the orthographic transcription work was made by the author, with additional help from a computational linguist.

3.7.2 Labeling

All corpora were labeled for disfluencies by the author, over a period of a few years, and later revised (several times) during a period of a couple of months for consistency.

3.8 Total data collected

A detailed breakdown of the data is given in **Appendices 1** through **4**. Summary statistics are shown in **Table 3.1**.

Table 3.1. Summary statistics of total data collected in the four data collections.

Corpus	No. Subjects	No. Dialogues	No. Utterances	No. Words	No. Minutes
WOZ-1	46 (25M/21F)	433	4023	27664	212
WOZ-2	46 (32M/14F)	137	3438	26261	270
Nymans	8 (6M/2F)	24	1734	9250	72
Bionic	16 (9M/7F)	67	1985	12849	128
Σ	**116 (72M/44F)**	**661**	**11180**	**76024**	**682 (11.4 hours)**

3.9 Cross-corpus subjects

As mentioned above, the eight subjects in Nymans had all participated in WOZ-2. Summary statistics for these eight subjects/two corpora are shown in **Table 3.2**.

Table 3.2. Summary statistics for subjects participating in both WOZ–2 and Nymans. During final labeling consistency control, a few corrections were made, which means that the sums given for number of utterances slightly deviate from those given in Eklund (2002), where †=625 and ‡=1,730.

Subject ID		Gender	No. Dialogues		No. Utterances		No. Words	
WOZ-2	Nymans		WOZ-2	Nymans	WOZ-2	Nymans	WOZ-2	Nymans
10	1	M	3	3	113	135	505	1356
9	2	M	3	3	89	212	531	812
13	3	F	3	3	127	213	497	970
41	4	M	3	3	86	305	1124	1796
33	5	F	3	3	48	254	283	1304
38	6	M	3	3	42	194	297	819
46	7	M	3	3	56	221	384	1169
35	8	M	3	3	57	200	827	1024
		Σ	**24**	**24**	**618†**	**1,734‡**	**4,448**	**9,250**

3.10 Chapter summary

This chapter has described the corpora used for analysis in this work, as well as outlining and discussing the method employed to collect the said data, the so-called Wizard-of-Oz method.

The following chapter will discuss the disfluency labeling of these data.

4 Transcription and annotation

4.1 Introduction

This chapter will outline how the data described in chapter 3 were transcribed and labeled.

4.1.1 Orthographic transcription

The speech was transcribed orthographically word-by-word, without any analysis applied, either to disfluencies, or according to any other linguistic parameters. The only, slightly cumbersome, linguistic decision that required some decision-making, was to decide where a given utterance begun and ended, something which will be discussed in further detail in the following.

Most of the orthographic transcription was done by the author during the period 1998–2002. Annika Asp provided orthographic transcriptions of large parts of WOZ-1 and WOZ-2 during Spring 2000.[1]

An important thing to point out is that only the subjects' utterances were transcribed. In the three human–"computer"/computer corpora (WOZ-1, WOZ-2 and Bionic), the system's utterances were of course known (since they were scripted), and were consequently not included in the data that were analyzed. In the human–human corpus (Nymans), the agents were partly orthographically transcribed (by the author), but were not labeled for disfluencies, and were consequently not included in the analysis.[2] The reason for not including the system/agent side of the transcription was partly due to limited time and budget, but also due to the fact that dialogue analysis was not the prioritized objective of the data collections (at the time).

[1] She was already familiar with corpora, having written her Bachelor's Degree thesis on speech acts on data from WOZ-2 and Bionic, cf. Asp & Decker (2001a, 2001b).

[2] The transcriptions of the agents in Nymans were done in connection with work done on ingressive speech, as presented in Eklund (2002).

4.1.2 Disfluency annotation

The speech was analyzed according to the disfluency categories described in the following. As was shown in chapter 2, there are many alternative ways to categorize disfluencies, and some of the categories described previously even overlap (e.g. *dysrhythmic phonations* vs. *prolongations* and *truncations*. I have here chosen a set of categories that would allow for easy comparison with other disfluency schemes, e.g. that of Shriberg (1994), which in turn was based on the categories used in Switchboard (for a recent summary, see Meteer et al., 1995). Moreover, although it was not intended, the disfluency annotation employed here is also not too dissimilar to Allwood et al.'s **OCM Coding Standard** (Allwood, 1988a, 1994b, 1995, 1997b; Allwood et al., 2001a, 2001b; Allwood & Hagman, 1994/1999; Nivre et al., 1999; see also Allwood, Nivre & Ahlsén, 1990, 1992; Nivre, Allwood & Ahlsén, 1999). My disfluency annotation and the OCM coding standard are compared in Allwood, Abelin & Grönqvist (1999), Allwood & Björnberg (1999) and Abelin & Allwood (1998/1999). Two things must be pointed out, however, the first being that some of the symbols that were used in 1998 were excluded at later stages, e.g. items such as *discourse markers* or *coordinating conjunction*. The other thing that must be borne in mind is that although symbol mapping between my approach and OCM is quite feasible, the research objectives are somewhat different. While Allwood et al. put more emphasis on dialogue function, the present study primarily focuses on structure and distribution. However, as Abelin & Allwood (1998/1999), comparing OCM with my analysis point out:

> Den grundläggande skillnaden är troligen att: R.E:s kodning inte gäller "own communication management", alltså en inriktning på avbrott eller ändring i syfte att reglera den egna kommunikationen, utan har som huvudsyfte att fånga "disfluenser". Trots denna skillnad blir det ofta liknande kodningsresultat i praktiken.[1] (Abelin & Allwood, 1998/1999.)

All disfluency annotation work was carried out by the author, during the period 1998–2002. During this period, categories were abolished or changed, and the data was consequently gone through several times in order to make annotation consistent.

4.1.3 Labeling consistency

As soon as labeling/annotation of more or less arbitrary character—however cleverly motivated the arbitrariness may be—plays a part in the analysis of data, it is customary to carry out some kind of consistency analysis to make sure that the data under scrutiny is reliably consistent. The standard way of doing inter- and intra-labeler consistency analysis in recent years has been the kappa coefficient (κ), also called Cohen's kappa (Cohen, 1960; Landis & Koch, 1977; Carletta et al., 1997).

Since all corpora in this thesis were transcribed by the author, no inter-labeler consistency analysis was carried out. Neither was any intra-labeler carried out, although that would have been possible, and perhaps even motivated, especially since given basic annotation principles were changed over the years. However, since the data was gone through *in toto* several times—the last time during a two-week period in October 2002, when some inconsistencies were fixed—it was decided that no intra-labeler analysis was called for, the data being carefully adapted to the same underlying, and updated, transcription principles.

[1] "The fundamental difference is probably that: R.E:s coding is not about 'own communication management', i.e., a focus on breaks or change in order to manage one's own communication, but instead has as its main objective to capture "disfluencies". Despite this difference, in practice the coding quite often is similar."

4.2 Labeling architecture: ToBI

The method applied for transcription and annotation was based on the *ToBI* labeling standard (see **References**: ToBI; Beckman & Ayers, 1993/1997; Beckman & Hirschberg, 1994; Beckman, Hirschberg & Shattuck-Hufnagel, 2004), which is a multi-tiered analysis tool, primarily created for the analysis of prosody and intonation in English (Kim Silverman et al., 1992; Pitrelli et al. 1994). It combines an auditory analysis of the speech files with a visual representation of the waveform and the F_0 contour, with the possible addition of a spectrogram. Originally developed as a labeling system for Standard American, **ToBI**, which stands for **TOnes and Break Indices**, has been used to analyze both other varieties of English, such as British English (Roach 1994), Northern Ireland English (Nolan & Grabe, 1997, although they point out some problems within the ToBI framework), and Glasgow English (Mayo, Aylett & Ladd, 1997), as well as other languages such as German (Grice et al., 1996), Japanese (Venditti, 1997) and also Australian languages, Basque, Quebec French, Pan-Mandarin and Cantonese Chinese (discussed in Beckman, Hirschberg & Shattuck-Hufnagel, 2004). Inter- and intra-labeler analyses are reported in Pitrelli et al. (1994), Grice et al. (1996) and Syrdal & McGory (2000).

As was mentioned above, ToBI is primarily created for the analysis of intonation, not disfluencies, and in the *Conventions*[1] it is stated that:

> Individual transcribers will also determine whether and how to transcribe phenomena such as filled pauses (e.g., "um", "uh") and whether to use contractions (e.g., "gotta") or not.

However, Nakatani & Shriberg (1993) presented a paper where they propose that ToBI be used for disfluency labeling, and Beckman, Hirschberg & Shattuck-Hufnagel (2004) describe how disfluencies are labeled in the miscellaneous tier, either as a "localized event" with the tag **`p`**, or using the "paired labels" **`disfl<`** and **`disfl>`**, for the beginning and end of a disfluent stretch of speech. The focus within the ToBI framework was originally not the analysis of disfluencies per se, and the disfluency labeling mentioned above was mainly carried out to make parsing of the Tones and Break Indices tiers more reliable. However, Wightman (2002) points out the future need to endow the ToBI framework with the possibility to allow for "phenomena of real speech such as disfluencies, interruptions, back-channel speech, *etc*" (Wightman, 2002, p. 26).

The basic ToBI software was downloaded from the official website, and adjusted by the author to meet the needs of the current disfluency analysis.[2]

As was mentioned above, ToBI is conceived as a tiered analysis, and this feature was kept, if mainly for visual reasons. I used a three-tiered version, containing an **orthographic tier**, a **disfluency tier**, and a **comments tier**. These will be described in detail in the following paragraphs.

The ToBI script was written in Unix shell command language, and run on Sun Sparc workstations (different versions over the years) under *Solaris*. Analysis was carried out using *ESPS/waves+™* (see **References**).

[1] **http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html**
[2] I am indebted to Gayle Ayers Elam, Mary Beckman and Colin Wightman for kind support.

An example of what the interface looked like is shown in **Plate 4.1** below.

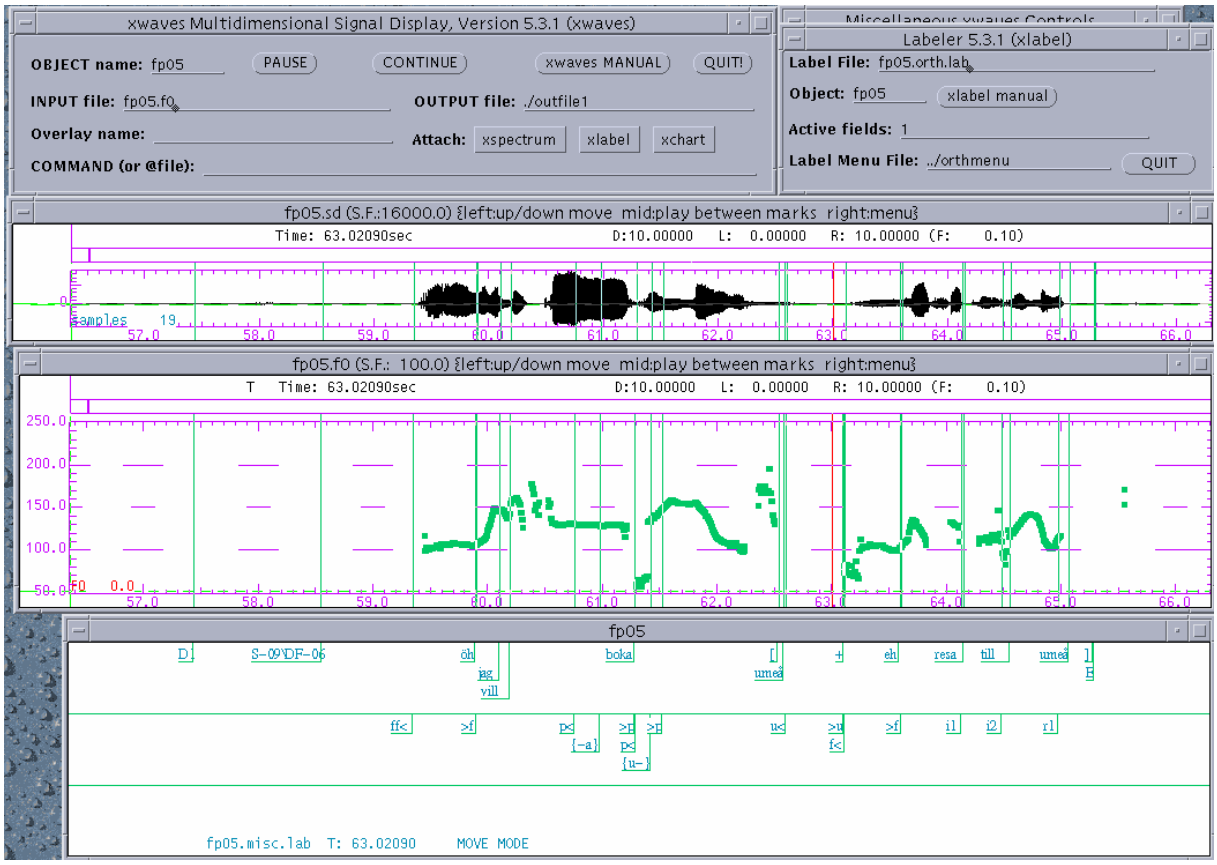


Plate 4.1. Transcription tool interface. Bionic corpus, subject no. 5, dialogue 1, first utterance. (Compare Appendix 5 Transcription Sample.) The utterance reads "öh jag vill boka umeå eh resa till umeå" ('uhm I'd like to book Umeå uhm trip to Umeå.' *ESPS/waves*™ interface, ToBI-style (xlabel). The sound wave appears in the top window, the F_0 contour in the second, and the three-tiered transcription window at the bottom. The first tier includes dialogue number (`D1`), word- and disfluency frequency in utterance (`S-09\DF-05`), the orthographic transcription and some of the disfluency labeling (not shown in this figure). The second tier includes most of the disfluency labeling, and the third tier includes additional comments (empty in this figure).

4.3 The orthographic tier

The utterances were transcribed in lexicon-lookup orthographic form. Consequently, contrary to other transcription schemes, like the **Modified Standard Orthography**, or MSO for short, (Allwood, Grönqvist, Ahlsén & Gunnarsson, 2002; Allwood, Abelin & Grönqvist, 1999; Nivre 1999), which allows for variant spellings of spoken language forms not recognized in standard orthography, like the (common) spoken form *ja* for standard *jag* ("I" or "me"), no orthographic adaptation to spoken-language forms, e.g., reductions, were made in this work. The rationale for this was mainly technical, since it is easy to go from lexical lookup form to reduced form, but harder to do vice versa. It also simplified the task, since no particular considerations had to be made as to the phonetic granularity of the transcription. Moreover, the main objective of this work is not tightly tied to reductions in spontaneous speech, which abound without necessarily indicating hesitation of any form (rather the opposite), but on disfluencies. A final reason was that it is easier to calculate vocabulary size if lookup forms are used consistently.

4.3.1 Dialogue number

Each dialogue was prefixed with the number of the dialogue, thus:

D*n*

… where ***n*** is the number of the dialogue. Compare **Plate 4.1** and **Appendix 5 Transcription Sample**.

4.3.2 Number of words / disfluencies in utterances

Each utterance was prefixed with the number of words and disfluencies of that utterance, thus:

S-*nn*\DF-*mm*

… where **S** indicates the beginning of a sentence (utterance), and **DF** denotes disfluencies in that same utterance (cf**. 4.3.2.2**).

4.3.2.1 Definition of utterance

One fairly elusive concept within linguistics is that of *utterance*, corresponding, in a loose way, to that of sentence in text. In speech, of course, things are different.[1] The definition used in this work has been a fairly pragmatic one, counting as utterances simply vocalizations made without intervening utterances from the conversational partner, i.e., the "machine" (WOZ-1 and WOZ-2) or the machine (Bionic). This means that in cases where no response was promptly provided by the system, subjects could sometimes add something, quite often a clarifying prepositional phrase, perhaps to "fill the void", in which case a long unfilled pause was included in the analysis. The problem of having the system know when to regard what the subject has uttered as a complete utterance to be acted upon is treated in Bell, Boye & Gustafson (2001). Although the Human–Human corpus (Nymans) differed from the three human–machine corpora in that it contained more overlaps between the subject and the agent, the same definition was employed for Nymans, mainly for consistency, with an awareness of its not being "ideal" in any way. Also, since the system/agent side was not transcribed or analyzed (cf. **4.1.1**), a definition of an utterance grounded in interaction or dialogue analysis was not possible.

4.3.2.2 Start-of-utterance

The beginning of each utterance was tagged:

S-*nn*\DF-*mm*

… where ***nn*** was the number of words in the utterance, and ***mm*** the number of disfluencies in the utterance. Compare **Figure 4.1** and **Appendix 5 Transcription Sample**. These figures were derived from both manual and automatic counting, according to the principles laid out in section **4.8.1**.

[1] Utterances could of course be defined syntactically, semantically, pragmatically or prosodically. I will waive a deeper discussion concerning the similarities and/or differences between these definitions here, however.

4.3.2.3 End-of-utterance

The end of each utterance was marked by the symbol:

E

The reason for doing this was mainly ease of automatic extraction. Compare **Plate 4.1** and **Appendix 5 Transcription Sample**.

4.3.3 Mispronunciations (MPs)

Words that were mispronounced were marked with a tilde:

~

An authentic[1] example would be:

jag vill ha en flygbullsbiljett~
"I would like a glound transport ticket"

4.3.4 Truncations (TRs)

Truncated words, i.e., words that were not finished, were marked with a slash sign:

/

An example would be:

jag vill ha en fly/ en biljett till Umeå
"I would like a fligh a ticket to Umeå"

If the word was continued, a hyphen was added, thus:

jag vill ha en tåg/- -/biljett
"I would like a train … ticket"

Naturally, this is sometimes hard to distinguish from a mid-word unfilled pause, and the difference was mainly made on prosodic grounds, i.e., whether or not it was obvious for prosodic reasons that the speaker was planning to continue the word s/he had begun.

4.3.5 Repairs (REPs)

The orthographic tier also included what has been called 'repairs'. In **Figure 4.1** above, one can analyze the utterance:

öh jag vill boka umeå eh en resa till umeå
"uh I would like to book Umeå er a trip to Umeå"

… so that a "cleaned-up" version, edited version would be:

[1] Most of the examples in this section are authentic, while some are made up, mainly for pedagogical reasons.

jag vill boka en resa till umeå
"I would like to book a trip to Umeå"

This implies that—besides getting rid of the filled pauses *öh* and *eh*—one breaks down the sub-part:

umeå en resa till umeå
"Umeå a trip to Umeå"

… into two parts:

[umeå + en resa till umeå]
"Umeå + a trip to Umeå"

… where the word (place-name) *Umeå* is repeated, with three "inserted" words, *en resa till* ("a trip to") in the repeated part. This has been described in the literature (e.g. Shriberg 1994) as having a **Reparandum** (*Umeå*) and a **Reparans** (*resa till Umeå*), according the general pattern:

[Reparandum + Reparans]

The plus sign in the example above denotes what has been called the **Interruption Point** (e.g. Shriberg, 1994). From a technical perspective, what is found left of the Interruption Point can be discarded, and what is found right of the Interruption Point is what is needed to interpret the utterance. While Shriberg (1994) and others use a period (full stop) to indicate the Interruption Point, a plus sign, **+**, is used here, mainly for visual reasons.

Repairs can in turn include repairs. Consider the following example (with an inserted explicit editing term *nej* ("no"), cf. **4.4.5**):

jag vill ha en resa till umeå en billig resa till umeå nej kiruna
"I would like a trip to Umeå a cheap trip to Umeå no Kiruna"

First the phrase:

en resa till umeå
"a trip to Umeå"

… is repeated/repaired with the inserted word **billig**, as follows:

[en resa till umeå + en billig resa till umeå]
"a trip to Umeå a cheap trip to Umeå"

But in the Reparans part, the place name *Umeå* is substituted, making to desired goal of the trip *Kiruna* thus:

en billig resa till [umeå + nej kiruna]
"a cheap trip to Umeå no Kiruna"

Consequently, the entire utterance could be analyzed thus:

[jag vill ha en resa till umeå en billig resa till [umeå nej kiruna]]
"I would like a trip to Umeå a cheap trip to Umeå no Kiruna"

Note how one repair is embedded in a bigger repair. The "cleaned-up" version, with the assumption that this is what the speaker intended to say (and what he or she should have put in print, being allowed to pre-edit the utterance), would then come out as:

jag vill ha en billig resa till kiruna
"I would like a cheap trip to Kiruna"

Whether or not such nested repairs do in fact correspond to what is going on in our brains during speech production is, naturally, subject to discussion, and one could e.g. analyze repairs according to a flat structure. However, a nested structure analysis has been applied throughout this study.

4.4 The disfluency tier

The second tier was used for most of the disfluency labeling.

4.4.1 Repairs (REPs)

As mentioned above, labeling of repairs was divided between two different tiers, mainly for visual reasons. (The three label files were later merged.)

4.4.1.1 Repeated items

Words in the Reparandum part are sometimes repeated in the Reparans part. This was indicated in the Disfluency tier by the tags ***rn***, where ***n*** gives the number of words repeated in the specific repair.

```
[ resa från um/ + resa till  umeå ]
                  r1   r2
```
"trip from um trip to Umeå"

4.4.1.2 Inserted items

Words can also be inserted in the Reparans part, which was indicated with the tag ***in***, where ***n*** gives the number of the inserted word.

```
jag vill ha [ resa från umeå + en  resa från umeå ]
                               i1  r1   r2   r3
```
"I would like trip from Umeå a trip from Umeå"

4.4.1.3 Deleted items

Words in the Reparandum that do no appear in the Reparans are called deleted, and are marked in the Reparandum with the tag ***dn***, where ***n*** indicates the number of the deleted words.

```
jag vill ha en [ dyr resa till umeå + en  resa från umeå ]
                 d1                   r1  r2   r3   r4
```

"I would like an expensive trip to Umeå a trip from Umeå"

4.4.1.4 Substituted items

Words in the Reparandum can be replaced in the Reparans with preferred forms, marked with the tag ***sn***, where n indicates the number of substituted items.

```
jag vill ha en resa [ till umeå + från umeå ]
                                  s1   r1
```

"I would like a trip to Umeå from Umeå"

While the tags for repeated, inserted, deleted and substituted items are taken from Shriberg (1994), a difference between her system and the system used here is that she does not index the repairs with number. The reason this is done here is mainly ease of frequency count, like what words are most likely repeated alone, what combination of two-repeated words are most frequent, what is the maximum number of repeated words in the Reparandum, and so forth.

4.4.2 Unfilled pauses (UPs)

During a stretch of speech, a speaker can turn silent for shorter or longer periods of time. Some of these are barely perceived, but they can also be very long indeed. These periods of silence are often referred to as "unfilled (or silent) pauses" in the literature.

These are marked in the disfluency tier, using the paired tag:

```
u< >u
```

… corresponding to the beginning and the end, respectively, of the silent stretch of speech. The rationale for tagging both the opening and closing of unfilled pauses is simply to make it easier to extract durational values.

It is quite often somewhat problematic to define what should count as an unfilled pause, in particular in the human–human corpus (Nymans), where there is frequent overlapping between the interacting speakers,[1] and they are quite often excluded from disfluency statistics. A discussion on unfilled pauses was given in Bell, Eklund & Gustafson (2000), where it was argued that unfilled pauses occur on a scale from sure-fire hesitation phenomena (e.g. when they occur inside words) to more dubious cases (e.g. in-between grammatically distinguishable utterances in a multi-sentence utterance). A description of this "sliding scale" is given below.

[1] As has already been mentioned, the travel agents were not transcribed, labeled or analyzed, but that there is overlapping is known, partly since it is obvious even when listening only to the subjects, partly since parts of the agent side were transcribed when analyzing the occurrence of ingressive speech (cf. Eklund, 2002), which confirmed this assumption.

4.4.2.1 Unfilled pauses inside words

The most obvious case where an unfilled pause constitutes a disfluency is when it occurs inside a lexical root. An authentic example is:

fö UP re
"bef … ore"

That this should "count" as disfluency goes without saying.

4.4.2.2 Unfilled pauses inside compounds

Another case of word-internal unfilled pauses is when they appear inside compound words, which is the case in the following example:

tåg UP stationen
"the train … station"

That the compound "train station" is in fact one word here is obvious for prosodic reasons, much the way it is easy to tell the difference between a "blackbird", a specific species, and a "black bird", e.g. a black swan, or other kind of bird. Besides Swedish (Eklund & Shriberg, 1998), word-internal UPs have been described in German (Lüngen et al., 1996; Althoff, 1997; Althoff et al., 1996), also a language with very productive compound word formation.

A more striking example, drawn from the data, is:

konferens eh UP eh lokalen
"the conference eh … eh hall"

Note that this example also includes two filled pauses.

4.4.2.3 Unfilled pauses inside phrases

Another case is when an unfilled pause occurs inside a phrase, like:

en tur och retur till UP borås
"a return ticket to … Borås"

In this case, which is a quite common location for unfilled pauses, the UP occurs just before the head of the phrase, i.e. hesitation occurs before picking out the semantic heavy item. It is argued here that this also should count as a full-fledged disfluency.

4.4.2.4 Unfilled pauses between grammatically complete forms

A much more cumbersome case is when the two utterances preceding and following the silent stretch of speech in one way or another form grammatical utterances. A common case is:

en tur och retur till borås UP på fredag
"a return ticket to Borås UP on Friday"

In this case (where *på fredag* would constitute an ellipsis), it is harder to tell whether the second part was planned at the very beginning, and was subject to some hesitation, or whether it was submitted since the system failed to respond quickly enough after the first,

grammatically complete, utterance. However, one can quite often get some clues from the prosodic realization of the first part here, i.e., whether it includes an utterance-final fall or not, and if that has been the case, such cases have been tagged with an unfilled pause.

4.4.2.5 Deliberate pauses (and clear diction)

A special case of speech which occurs mainly in the human–machine corpora, is when the subjects try to make it easier for the system (the automatic speech recognizer) by producing each word "in isolation" by inserting a short pause in-between each and every word, something which is most often accompanied by a clearer diction of the words proper. This, however, is a very obvious strategy, and rather than being a sign of hesitation, it is a sign of extreme and conscious planning, in order to adapt to the system. Consequently, such pauses have not been included in the annotation.

4.4.2.6 Final comments

It should be obvious in the previous paragraphs that the definition of what counts as an unfilled pause, irrespective of where it occurs, is not based on duration in milliseconds, an approach found in the literature. First, what is perceived as an unfilled pause should be normalized for (local) speech rate, second, it can be assumed that the perception of silences is also influenced by their position in relation to syntax and semantics. Consequently, I have labeled as unfilled pauses silences that appeared to me as (hesitation) pauses where they occurred in the speech string. This, at least, provides some kind of normalization as to speech rate and syntax/semantic in that it is the perception of a native speaker of the language. It goes without saying that this method could be criticized, but the purely durational approach is also, as mentioned above, subject to criticism.

4.4.3 Filled pauses (FPs)

The term filled pause, sometimes called filler words, refers to vocalized hesitation, including sounds as (English) *eh*, *uh*, *uhm*, *er* and the like. These do possess some kind of lexical status, and it has been shown that they are among the most common words in spoken conversation (Shillcock et al., 2001).

While the "word" proper was written into the speech stream in the orthographic tiers (like *eh* or *öh*), it was also transcribed into the disfluency tier using the paired tag:

```
f< >f
```

… to mark the beginning and end, respectively. This was done in order to facilitate durational analysis of filled pauses.

Since it has been shown over and over again in the literature (e.g. Shriberg, 1994; Eklund & Shriberg, 1998; Eklund, 1999, 2000a) that filled pauses very often begin utterances, a special opening tag was used for utterance-initial filled pauses, thus:

```
ff< >f
```

This was simply done to facilitate the distinction between utterance-initial filled pauses, and filled pauses in other positions.

4.4.4 Prolongations (PRs)

Another way to hesitate without being silent is to continue phonation by dwelling on a speech sound in the speech produced, illustrated below:

vad kostar ffffffflyg to umeå
"how much is a ffffffflight to Umeå"

Prolonged segments—once thought to be an acid test phenomenon to diagnose stutterers—were not marked in the orthographic tier, but were labeled in the disfluency tier with the paired tag:

p< >p

In order to analyze what particular segments were prolonged, as well as to clarify which segment in the corresponding word in the orthography tier the tag referred to, the sound in question was included within the tags, thus:

p< {-e} >p

In the case above, a word-final **/e/** was prolonged, the dash **-** indicating that the sound is word-final. In cases where a word-medial segment was prolonged, but normal pronunciation deleted the lexically final segment, which is the case with the word *det* ("it"), which is most often pronounced **[de]** , the following annotation was used:

p< {-e(-)} >p

A word-medial sound was surrounded by two dashes:

p< {-s-} >p

In cases where a compound word (ubiquitous in Swedish) included a prolonged segment being either final or initial in one of the joined words, this was annotated thus:

p< {-n(#-)} >p

.. where the hash sign **#** indicates a lexeme border.

Finally, the following indicates that a word-initial **/m/** is prolonged:

p< {m-} >p

Given the fact that there is no such thing as the "correct" duration of phonemes, there is a certain arbitrariness associated with this category. While there is no doubt the case that speakers may linger on segments in words pronounced in part or in full, rather than inserting a more typical filled-pause sound such as *eh*, what is *perceived* as a prolonged segment is to a certain extent depending on local speech rate. The strategy used to label prolongation used in this study was as follows:

If a certain segment sounded prolonged, the playback cursors were put so as to make the word sound "normal", listening only to the part of the word inside the cursors. The remainder of the actual word was marked as the prolonged part.

A further decision was made concerning certain words that quite often prolonged versions exist in a more or less "lexicalized" form. Such words include (confirming) *ja* ("yes"), which very often appears in a two-beat version *ja-a*. The same goes for (confirming) "*mm*" (meaning the same thing, or having more or less the same function), which also appears in a two-beat version *mm-m*, as does *nä* ("no"). These (and similar forms) were not regarded as prolonged versions, but rather two-syllable versions of the same words. They were indicated orthographically like *ja-a*, *mm-mm* and *nä-ä*, respectively.

Another phenomenon excluded from prolongation tagging was emphatically pronounced words, that typically exhibit longer segments (Heldner & Strangert, 2001).

4.4.5 Explicit editing terms (EETs)

Sometimes the subjects explicitly signal that something has gone wrong, and/or that she wants to rephrase the utterance. In English this is most often done using words like *oops*, *sorry*, "*no, wait* and so on. This was marked with the tag **eet*n*** , where **n** indicated the number of the word in the **EET**.

```
jag vill ha en flygbiljett nej  fel  en tågbiljett till umeå
                            eet1 eet2
```

"I would like a flight ticket *no wrong* a train ticket to Umeå"

EETs thus can consist of one word only, like *oops*, or several words, like *oops correct that* or even more, like *oops what am I saying, correct that*. Irrespective of the number of words, they all counted as one **EET** (cf. **4.8.1**), but the index figure made it easy to obtain statistics concerning the number of words in the occurring **EET**s.

Another trait of **EET**s is that they can either correct *form* or *content* (e.g. mispronunciations or day of departure, respectively). They can also be addressed to the agent/system, or to the speaker him/herself. These distinctions were not explicitly marked in the labeling.

4.5 The comments tier

A third tier was used for additional comments that were not part of the disfluency labeling proper. These will be briefly described in the following paragraphs.

4.5.1 General comments

Most comments in the Comments Tiers were general comments, such as:

_meta

… for utterances that were uttered outside the task proper, or:

self_directed

… for utterances that were not directed to the system, but were examples of "thinking aloud".

Comments like:

```
problematic_rep
long_pr
```

… were also added here.

All comments were written with an underscore, to be able to include or exclude from data retrieval (using e.g. the **grep** function in *Unix*).

4.5.2 Ingressive speech

As has been mentioned above, speech produced on pulmonic ingressive airstream is common in Swedish, especially on the words *ja* (yes) and *nej* (no) and similar. It was indicated in the comments tier with the tag pair:

```
_ingr< >ingr_
```

The opening/closing tags were put there to facilitate durational analysis, tantamount to the labels for filled pauses and prolongations.

4.6 Disfluency analysis files

The orthographic, disfluency and comments tiers all have their own text files according to the general ToBI (xlabel) implementation. These three files were merged (using shell scripts) into one label file, **sentences_vertical**, containing four columns, thus:

```
125.537001                      S-06\DF-02
126.085782          p<
126.217896          {-a(g)}
126.522774          >p
126.538018                      jag
126.548181          f<
126.832734                      uh
126.832734          >f
127.030904                      vill
127.223994                      åka
127.366270                      från
128.047165                      stockholm
128.357124                      E
```

The **first column** gives the location in seconds in the sound (speech) file. The figure indicates the *end* of each annotated item (in seconds from the beginning of the sound file). The **second column**, which is empty here, corresponds to the comments tier. The **third column** corresponds to the disfluency tier, which in this case contains a prolongation and a filled pause. The **fourth column** corresponds to the orthography tier, and includes the utterance tags **S-06\DF-02** (indicating that there are six words and two disfluencies in this utterance) and **E**, as well as the orthographic transcription of the sentence (this example lacking mispronunciations, truncations or repairs).

In order to facilitate per-utterance analyses, another script converted **`sentences_vertical`** into a "horizontal" file, **`sentences_vertical`**, where each sentence occupied one row, thus:

```
S-06\DF-02 p< {-a(g)} >p jag f< uh >f vill åka från stockholm E
```

For full examples of a dialogue in the vertical and the horizontal format, the reader is referred to **Appendix 5** Transcription sample.

4.7 Disfluency categories: summary

A synoptic overview of the disfluency categories employed in this thesis, with subclasses where applicable, is given in **Table 4.1**.

Table 4.1. Overview of labeling symbols.

Disfluency	Description	Symbol	Disfluency subclasses	
			Symbol	Description
UP	Unfilled Pause (Silence)	**u< >u**	(none)	(none)
FP	Filled Pause (Filler Word)	**f< >f**	**ff<**	Utterance initial
PR	Prolongation	**p< >p**	**{*x*}** **{*x*-}** **{-*x*-}** **{-*x*}** **#** **(*x*)**	Segment Word initial Word medial Word final Lexeme border Suppressed segment
EET	Explicit Editing Term (Self-correction)	**eet**	**eet1** **eet2** . . . **eet*n***	First word Second word . . . *n*th word … in **eet**
TR	Truncation	**/**	(none)	(none)
MP	Mispronunciation	**~**	(none)	(none)
REP	Repair	**[+]**	**[**	Beginning of Repair
			+	Interruption Point
			]	End of Repair
			r*n*	Repeated word *n* in *Reparans*
			d*n*	Deleted word *n* in *Reparandum*
			s*n*	Substituted word *n* in *Reparans*
			i*n*	Inserted word *n* in *Reparandum*

4.8 Obtaining the results

While general and specific disfluency rates are given in the literature, it is not always obvious exactly what has been counted (cf. the discussion in Bell, Eklund & Gustafson, 2000). For example, as noted in Fox Tree (1995), the figures given are very dependent on whether or not unfilled pauses are included in the counts. In this section, I will briefly outline exactly how disfluencies were counted, so as to render the figures in the following chapters clearer, and facilitate comparisons with other research.

4.8.1 Counting disfluencies

The figures that are reported in this thesis were obtained as follows, broken down for each category of disfluency.

4.8.1.1 Unfilled pauses (UPs)

Each incidence of **u<** was counted.

4.8.1.2 Filled pauses (FPs)

Each incidence of **f<** was counted (which captures all instances of **ff<**).

4.8.1.3 Prolongations (PRs)

Each incidence of **p<** was counted.

4.8.1.4 Explicit editing terms (EETs)

Each incidence of **eet1** was counted. This means that each **EET** was counted as one, irrespective of the number of words included in the **EET**, i.e., both "sorry" and "no, sorry", "oops, that was wrong" (and so on) were all counted as one **EET** (each).

4.8.1.5 Mispronunciations (MPs)

Each incidence of **~** was counted.

4.8.1.6 Truncations (TRs)

Each incidence of **/** was counted.

4.8.1.7 Repairs (REPs)

Each interruption point, i.e. **+**, was counted. This means that each nested **REP** was given its own count. It also means that the number of words in either the Reparandum or the Reparans did not affect the sum total of **REP**s. The number of deletions, substitutions, repetitions or deletions was calculated separately, and do not appear in the total figures.

4.8.2 Counting method

The counting was carried out by running simple *Unix* commands/scripts on the transcription files **`sentences_vertical`** and/or **`sentences_horizontal`**, most often both, in order to double check that the figures returned were correct. The returned figures were also compared to the manually calculated figures that appear in the header, **`DF-`*`nn`***, to triple check the consistency of the transcription/annotation.

4.8.3 Analyzing the figures

Besides common percentages (for which a normal calculator was most often used), statistical analyses were carried out using the statistical program package *SPSS Base/Advanced Models* (see **References**), or in some cases (e.g. test-of-proportions), using a scientific calculator.

4.9 Chapter summary

This chapter described how the data were transcribed orthographically and labeled for disfluencies. The general ToBI-style labeling architecture was described, as were the different disfluency categories, and their respective definitions. Finally, a detailed account as to how disfluency rates were obtained was provided.

The next chapter will present the results of the analyses.

5 Results and analyses

5.1 Introduction

This chapter will describe results and analyses carried out on the collected data, as described in the two previous chapters. As was mentioned in the introduction, focus will be on frequency and distribution of disfluencies throughout the corpora, beginning with the most general observations, whereupon more detailed studies will be undertaken with each disfluency type bestowed its own paragraph. Rather than separating results proper from their analysis, these will be intertwined so that analyses and comments will made immediately for each particular study/paragraph.

5.2 Summary statistics

The numbers of disfluencies—according to the typology described in the previous chapters—collected are shown in **Table 5.1**.

Table 5.1. General disfluency (DF) incidence in the corpora, broken down for type and corpus. Figures are given for unfilled pauses (UPs), filled pauses (FPs), prolongations (PRs), explicit editing terms (EETs), mispronunciations (MPs), truncations (TRs) and repairs (REPs).

Corpus	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
WOZ-1	1622	815	156	20	41	167	215	3036
WOZ-2	2179	1040	132	31	22	134	257	3795
Nymans	562	203	129	9	10	141	172	1226
Bionic	1226	543	197	28	31	146	202	2373
Σ	5589	2601	614	88	104	588	846	10430

This seems clear-cut enough, but comparing with other studies is not as straightforward as one might expect. Perusing the literature, it is often not completely clear exactly *what* is included in disfluency counts. Bell, Eklund & Gustafson (2000) concluded—in an indirect way—that unfilled pauses are not included in most counts. This conclusion was reached by comparing their own figures with comparable figures in the literature, taking into account that unfilled pauses are by far the most common type of disfluency (albeit with the associated problem of defining what counts as an unfilled pause).

Moreover, depending on the particular corpora, the incidence of one-word utterances, for instance, might affect disfluency/utterance ratios heavily, since one-word utterances rarely contain disfluencies. Another reason is that one-word utterances are very frequent in this type of task-oriented domain, where yes/no-utterances are legion.

General disfluency incidence, broken down for different kinds of counts, and with and without one-word utterances, is given in **Table 5.2**.

Table 5.2. General disfluency (DF) incidence in the corpora. Number of disfluencies (all types pooled), with and without unfilled pauses, broken down for all corpora. Numbers and percentages of utterances containing at least one disfluency are given, as are completely fluent utterances for all corpora. Numbers and percentages for all utterances, as well as numbers and percentage of numbers of completely fluent utterances with one-word utterances excluded. Percentages of disfluencies, divided by number of words are given, with and without words in one-word utterances, and with and without unfilled pauses included in the disfluency count. The row with figures surrounded by bold lines is the most likely candidate for comparison with other studies on general disfluency frequency. The overall comparison figure is given in boldface.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. DFs	3036	3795	1226	2373	10430
No. DFs –unfilled pauses (UPs)	1414	1616	664	1147	4841
No. utterances	4023	3438	1734	1985	11180
No. 1-word utts	906	351	794	643	2694
No. utts –1-word utts	3117	3087	940	1342	8466
No. words	27664	26261	9250	12849	76024
No. words –1-word utts	26758	25910	8456	12206	73330
No. disfluent utts	1268	1505	445	744	3962
% disfluent utts	31.5%	43.8%	25.7%	37.5%	35.4%
No. fluent utts	2755	1933	1289	1241	7218
% fluent utts	68.5%	56.2%	74.3%	62.5%	64.6%
No. fluent utts –1-word utts	1849	1582	495	598	4524
% fluent utts –1-word utts	46.0%	46.0%	28.5%	30.1%	40.4%
% DFs/no. words	11.0%	14.4%	13.2%	18.5%	13.6%
% DFs/no. words –1-word utts	11.3%	14.6%	14.5%	19.5%	14.1%
% DFs –UPs/no. words	5.1%	6.1%	7.2%	8.9%	**6.4%**
% DFs –UPs/no. words –1-word utts	5.3%	6.2%	7.8%	9.4%	6.6%

Bortfeld et al. (2001, p. 135) reported 5.97 disfluencies per 100 words. Shriberg (2001) summarized previous studies on disfluency, and observed that "[r]ates of disfluency per word in spontaneous English vary from under 1% for constrained human–computer dialogue, to roughly 5–10% for natural conversation" (Shriberg, 2001, p. 155). Given that the second-last line in the table above—highlighted by bold lines—probably provides the best figures for a comparison with other corpora, it would seem as if the results here repeat their observations: disfluency rates range from 5.1% to 8.9% between the corpora, and with 6.4% disfluency for all the data pooled.

The first interesting question here is of course whether there are any overall differences between the corpora. Given that there are at least four different ways to make the calculations, even if one only considers the disfluency-per-word ratio, I have chosen to present the results for all four comparisons.

Overall cross-corpus differences are shown in **Table 5.3a** through **Table 5.3d**.

Table 5.3a. Overall cross-corpus differences: The percentages of number of disfluencies divided by the total number of words (percentages and number given for each corpus). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 11.0% ; 27664	**WOZ-2** 14.4% ; 26261	**Nymans** 13.2% ; 9250	**Bionic** 18.5% ; 12849
WOZ-1	—	$p < 0.05$ (W2)	$p < 0.05$ (N)	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	$p < 0.05$ (W2)	$p < 0.05$ (B)
Nymans	$p < 0.05$ (N)	$p < 0.05$ (W2)	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	$p < 0.05$ (N)	$p < 0.05$ (B)	—

Table 5.3b. Overall cross-corpus differences: The percentages of number of disfluencies divided by the number of words excluding words in one-word utterances (percentages and number given for each corpus). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (WOZ-1, WOZ-2, N[ymans] and B[ionic].

	WOZ-1 11.3% ; 26758	**WOZ-2** 14.6% ; 25910	**Nymans** 14.5% ; 8456	**Bionic** 19.4% ; 12206
WOZ-1	—	$p < 0.05$ (W2)	$p < 0.05$ (W2)	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	n.s.	$p < 0.05$ (B)
Nymans	$p < 0.05$ (W2)	n.s.	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	$p < 0.05$ (B)	$p < 0.05$ (B)	—

Table 5.3c. Overall cross-corpus differences: The percentages of number of disfluencies excluding unfilled pauses divided by the total number of words (percentages and number given for each corpus). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (WOZ-1, WOZ-2, N[ymans] and B[ionic].

	WOZ-1 5.1% / 27664	**WOZ-2** 6.1% / 26261	**Nymans** 7.2% / 9250	**Bionic** 8.9% / 12849
WOZ-1	—	$p < 0.05$ (W2)	$p < 0.05$ (N)	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	$p < 0.05$ (N)	$p < 0.05$ (B)
Nymans	$p < 0.05$ (N)	$p < 0.05$ (N)	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	$p < 0.05$ (B)	$p < 0.05$ (B)	—

Table 5.3d. Overall cross-corpus differences: The percentages of number of disfluencies excluding unfilled pauses divided by the number of words excluding words in one-word utterances (percentages and number given for each corpus). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (WOZ-1, WOZ-2, N[ymans] and B[ionic].

	WOZ-1 5.3% / 26758	**WOZ-2** 6.2% / 25910	**Nymans** 7.8% / 8456	**Bionic** 9.4% / 12206
WOZ-1	—	$p < 0.05$ (W2)	$p < 0.05$ (N)	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	$p < 0.05$ (N)	$p < 0.05$ (B)
Nymans	$p < 0.05$ (N)	$p < 0.05$ (N)	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	$p < 0.05$ (B)	$p < 0.05$ (B)	—

As is shown in **Table 5.3a** through **Table 5.3d** is that there are significant differences between all corpus pairs, with the sole exception of WOZ-2 and Nymans, when comparing the number of disfluencies divided by the number of words excluding one-word utterances. The general trend is that WOZ-1 is the least disfluent corpus, and that Bionic is the most disfluent. WOZ-2 and Nymans are more similar in that WOZ-2 is the more disfluent corpus in one of the counts, there is no difference in the second count, and Nymans is the more disfluent corpus in the two last counts, i.e. those that most likely can be compared to the literature.

This shows that task details do indeed matter, but also that the setting is not the whole story, since WOZ-2 and Nymans should be more dissimilar than WOZ-2 and Bionic, given that the subjects thought they were communicating with a machine in both these corpora (the only difference was that they only were in Bionic), while they knew they were talking with a real

human agent in Nymans. Still, WOZ-2 and Nymans are the most similar corpora as to general and overall disfluency frequency. This is a nice feature, given that these two are the corpora used for the cross-corpus comparison later in this chapter.

Another issue when comparing the corpora is to distinguish between number of words at *token* and at *type* level. This is interesting for mainly two reasons. First, differences between the corpora can be used as indirect ways to suspect that the tasks or settings differed in some respect—despite the attempts to keep most variables fixed—or to confirm the variables one wanted to vary did indeed cause a palpable effect in the resultant data. Second, from an application-point of view, the number of word tokens is of less interest than the number of word types, both from a speech recognition and language modeling perspective. The numbers of word tokens and types are given in **Table 5.4**.

Table 5.4. Number of words at token and type levels for all corpora.

Corpus	Number of words – tokens	Number of words – types	Ratio type/token
WOZ-1	27664	1126	4.1%
WOZ-2	26261	1227	4.7%
Nymans	9250	1120	12.1%
Bionic	12849	944	7.3%
Σ	76024	4417	5.8%

As is shown in **Table 5.4**, WOZ-1 and WOZ-2 have a quasi-identical type–token ratio, while Bionic exhibits a slightly higher frequency of distinct word form types. Not surprisingly, Nymans has by far the most varied vocabulary. WOZ-1 and WOZ-2 used highly scripted dialogues for the wizards/system, as well as constraints with regard to accepted utterance length from the users. Bionic, too, used scripted system dialogue, but put no upper limit on the users' utterances, which obviously is reflected in vocabulary variation. Nymans, of course, put no constraints whatsoever on the entirely natural human–human conversations, which shows up as a three times higher ratio of unique words as compared to WOZ-1 and WOZ-2. This leads to the conclusion that the more natural a system is, the larger the vocabulary will be. Then again, that observation might sound like stating the very obvious.

Another, related question, is what the most common words are in the different corpora. The top-ten lists for all corpora are shown in **Table 5.5**.

Table 5.5. Ten most common words (type) in all corpora. For all words, absolute numbers are given as well as percentages of that word out of the total number of word tokens in the corpus. Sums of all top-ten lists are given as well as the percentages of the total number of word tokens in the corpus. The filled pause *eh* is marked with boldface.

	WOZ-1	**WOZ-2**	**Nymans**	**Bionic**
Total number words (token)	27664	26261	9250	12849
Most common word (type)	jag (“I”)	jag (“ja”)	ja (“yes”)	ja (“yes”)
Number of word (tokens)	1614	1579	630	714
% word/tot. number words (token)	5.8%	6.0%	6.8%	5.5%
2nd most common word (type)	det (“it”)	den (“it”)	det (“it”)	jag (“I”)
Number of word (tokens)	1031	1030	535	692
% word/tot. number words (token)	3.7%	3.9%	5.8%	5.4%
3rd most common word (type)	den (“it”)	**eh** (“eh”)	mm (“yes”)**	det (“it”)
Number of word (tokens)	876	914	359	541
% word/tot. number words (token)	3.2%	3.5%	3.9%	4.2%
4th most common word (type)	till (“to”)	i (“in”)	jag (“I”)	den (“it”)
Number of word (tokens)	757	768	283	464
% word/tot. number words (token)	2.7%	2.9%	3.1%	3.6%
5th most common word (type)	**eh** (“eh”)	det (“it”)	då (“well”)*	**eh** (“eh”)
Number of word (tokens)	707	713	262	433
% word/tot. number words (token)	2.6%	2.7%	2.8%	3.4%
6th most common word (type)	och (“and”)	till (“to”)	den (“it”)	vill (“want”)
Number of word (tokens)	705	657	203	320
% word/tot. number words (token)	2.5%	2.5%	2.2%	2.3%
7th most common word (type)	tack (“thanks”)	vill (“want”)	och (“and”)	maj (“May”)***
Number of word (tokens)	677	616	196	300
% word/tot. number words (token)	2.4%	2.3%	2.1%	2.3%
8th most common word (type)	är (“is/are”)	och (“and”)	är (“is/are”)	till (“to”)
Number of word (tokens)	543	577	190	290
% word/tot. number words (token)	2.0%	2.2%	2.0%	2.3%
9th most common word (type)	då (“well”)*	noll (“zero/0”)	vi (“we”)	och (“and”)
Number of word (tokens)	494	451	171	228
% word/tot. number words (token)	1.8%	1.7%	1.8%	1.8%
10th most common word (type)	från (“from”)	då (“well”)*	**eh** (“eh”)	klockan (“o’clock”)
Number of word (tokens)	461	445	168	222
% word/tot. number words (token)	1.7%	1.7%	1.8%	1.7%
Σ (tokens) ten most common words	7865	7750	2997	4204
% of total number words (tokens)	28.4%	29.5%	32.4%	32.7%

* Literally “then”, but in this context mostly used as a topic-shifting discourse marker, loosely corresponding to “well” or “ok”. See Bretan, Eklund & MacDermid (1996).

** The word form *mm* is an affirmative feedback-giving discourse marker.

*** That the month “May” appears on the top-ten list is an artifact of the specific task sheet, of course.

The first thing to note here is that the first ten word types represent around 30% of the word tokens in all corpora, which more or less is in accordance with **Zipf's Law** (Zipf, 1945). Moreover, that the words *ja* ("yes"), *jag* ("I"), *den* or *det* ("it") and so on are common is not surprising, and is probably valid for most corpora of Swedish spoken language, irrespective of domain. What is more interesting from our point of view is the filled pause *eh* makes the top-ten list in all corpora. In fact, *eh* is the third most common word form in WOZ-2 (3.5% of all word tokens), the fifth most common word in WOZ-1 and Bionic (2.6% and 3.4%, respectively) and the tenth most common word in Nymans (1.8%). This observation is parallel to that of Shillcock et al. (2001), who observed that the filled pause was one of the most frequent words in English. Another difference between the corpora which is worth pointing out is that the feedback marker *mm*—although it is used in all corpora—makes the top-ten list in Nymans only.[1] This shows that feedback is more common in a human–human setting than a human–machine setting, ceteris paribus.

5.2.1 Disfluency frequency as a function of utterance length

That one-word utterances cannot contain more than one disfluency, at most (e.g. the filled pause *eh*, a truncated word, or a word with a prolonged segment) goes without saying, but what about longer utterances? Is there a linear correlation between utterance length and disfluency frequency? And what are the observed mean values for disfluency frequencies at different utterance lengths?

5.2.1.1 Disfluency frequency at different utterance lengths

The mean number of disfluencies in utterances of lengths between 1 and 32 words are given for all corpora in **Table 5.6a** through **Table 5.6d**. Although all corpora contain longer utterances than 32 words, this number was chosen since it was the highest number with an unbroken incrementally raising number of words common to all corpora. **Table 5.6e** shows the mean number of disfluencies in utterances of lengths between 1 and 32 words when all corpora are merged into one corpus.

[1] In WOZ-1, there are 62 *mm*'s (0.2% of word tokens); in WOZ-2 there are 162 (0.6%); in Bionic 53 (0.4%).

Table 5.6a. WOZ-1 corpus. For each utterance length, the number and percentages of fluent utterances are given, as are absolute and mean numbers of disfluencies (DFs) per utterance. Standard error, variance and standard deviations are also given.

No. words in utt.	No. utts.	No. fluent utts.	% fluent utts.	No. DFs	Mean DFs per utt.	Std. error	Variance	Std. dev.
1	906	903	99.6%	3	0.003	0.002	0.003	0.06
2	199	180	90.4%	21	0.10	0.024	0.115	0.34
3	428	396	92.5%	37	0.08	0.016	0.112	0.33
4	339	287	84.7%	65	0.19	0.027	0.250	0.50
5	305	237	77.7%	87	0.28	0.034	0.362	0.60
6	256	169	66.0%	138	0.54	0.058	0.853	0.92
7	230	140	60.9%	137	0.59	0.060	0.818	0.90
8	189	108	57.1%	125	0.66	0.068	0.863	0.93
9	135	74	54.8%	117	0.87	0.115	1.773	1.33
10	158	75	47.5%	146	0.92	0.089	1.255	1.12
11	136	42	30.9%	205	1.51	0.135	2.474	1.57
12	109	43	39.4%	148	1.39	0.159	2.697	1.64
13	89	25	28.1%	165	1.85	0.211	3.967	1.99
14	81	17	21.0%	163	2.01	0.201	3.287	1.81
15	75	14	18.7%	157	2.09	0.194	2.815	1.68
16	59	14	23.7%	109	1.85	0.202	2.407	1.55
17	46	10	21.7%	105	2.28	0.322	4.785	2.19
18	36	3	8.3%	89	2.47	0.299	3.228	1.80
19	43	5	11.2%	122	2.84	0.266	3.044	1.74
20	36	2	5.5%	144	4.00	0.427	6.571	2.56
21	32	1	3.1%	105	3.28	0.443	6.273	2.50
22	25	2	8.0%	82	3.28	0.481	5.793	2.41
23	15	5	33.3%	30	2.00	0.437	2.857	1.69
24	14	0	0%	57	4.07	0.474	3.148	1.77
25	10	2	20.0%	34	3.40	0.833	6.933	2.63
26	6	0	0%	27	4.50	1.688	17.100	4.13
27	4	0	0%	28	7.00	0.913	3.333	1.82
28	15	0	0%	76	5.07	0.539	4.352	2.09
29	3	0	0%	13	4.33	1.202	4.333	2.08
30	6	0	0%	32	5.33	1.145	7.867	2.80
31	4	0	0%	11	2.75	0.629	1.583	1.26
32	4	0	0%	22	5.50	0.867	3.000	1.73

Table 5.6b. WOZ-2 corpus. For each utterance length, the number and percentages of fluent utterances are given, as are absolute and mean numbers of disfluencies (DFs) per utterance. Standard error, variance and standard deviations are also given.

No. words in utt.	No. utts.	No. fluent utts.	% fluent utts.	No. DFs	Mean DFs per utt.	Std. error	Variance	Std. dev.
1	351	346	98.6%	5	0.01	0.006	0.014	0.12
2	348	312	89.7%	50	0.14	0.025	0.227	0.48
3	396	342	86.4%	68	0.17	0.024	0.224	0.47
4	316	231	73.1%	118	0.37	0.040	0.508	0.71
5	306	193	63.1%	163	0.53	0.047	0.676	0.82
6	237	141	59.5%	148	0.62	0.059	0.837	0.91
7	215	104	48.5%	198	0.92	0.082	1.447	1.20
8	152	64	42.1%	156	1.03	0.099	1.483	1.22
9	145	42	28.9%	193	1.33	0.106	1.626	1.27
10	128	43	33.6%	174	1.36	0.125	1.996	1.41
11	120	29	24.2%	217	1.81	0.146	2.576	1.60
12	112	19	17.0%	241	2.15	0.155	2.670	1.63
13	76	21	27.6%	150	1.97	0.214	3.466	1.87
14	61	14	22.9%	128	2.10	0.246	3.690	1.92
15	79	9	11.4%	230	2.91	0.243	4.672	2.16
16	70	9	12.9%	180	2.57	0.259	4.683	2.16
17	54	7	13.0%	144	2.67	0.259	3.623	1.90
18	37	1	2.7%	102	2.76	0.283	2.967	1.72
19	36	1	2.8%	136	3.78	0.374	5.035	2.24
20	22	1	4.5%	85	3.87	0.457	4.600	2.14
21	22	0	0%	99	4.50	0.599	7.881	2.81
22	16	0	0%	52	3.25	0.487	3.800	1.95
23	19	1	5.3%	74	3.89	0.539	5.322	2.31
24	15	0	0%	68	4.53	0.639	6.124	2.48
25	16	0	0%	94	5.87	0.632	6.383	2.53
26	8	0	0%	25	3.12	0.515	2.125	1.46
27	11	1	9.1%	56	5.01	1.132	14.091	3.75
28	9	0	0%	40	4.44	0.884	7.802	2.65
29	9	0	0%	46	5.11	0.841	6.361	2.52
30	7	0	0%	43	6.14	1.932	26.143	5.11
31	2	0	0%	12	6.00	4.000	32.000	5.66
32	4	0	0%	35	8.75	3.341	44.917	6.70

Table 5.6c. Nymans corpus. For each utterance length, the number and percentages of fluent utterances are given, as are absolute and mean numbers of disfluencies (DFs) per utterance. Standard error, variance and standard deviations are also given.

No. words in utt.	No. utts.	No. fluent utts.	% fluent utts.	No. DFs	Mean DFs per utt.	Std. error	Variance	Std. dev.
1	794	789	99.4%	5	0.006	0.003	0.006	0.08
2	122	102	83.6%	24	0.20	0.044	0.242	0.49
3	111	94	84.7%	22	0.20	0.049	0.269	0.52
4	99	74	74.7%	37	0.37	0.075	0.563	0.75
5	77	47	61.0%	43	0.56	0.093	0.671	0.82
6	83	44	53.0%	58	0.70	0.098	0.798	0.89
7	53	32	60.4%	32	0.60	0.118	0.744	0.86
8	53	32	60.4%	34	0.64	0.138	1.004	1.00
9	52	18	34.6%	71	1.37	0.209	2.276	1.51
10	43	20	46.5%	46	1.07	0.211	1.924	1.39
11	30	9	30.0%	49	1.63	0.260	2.033	1.43
12	18	2	11.1%	34	1.89	0.351	2.222	1.49
13	15	3	20.0%	34	2.27	0.547	4.495	2.12
14	19	2	10.5%	60	3.16	0.491	4.585	2.14
15	19	4	21.0%	45	2.37	0.447	3.801	1.95
16	16	4	25.0%	31	1.93	0.381	2.329	1.53
17	17	5	29.4%	43	2.53	0.728	9.015	3.00
18	11	1	9.1%	41	3.73	0.776	6.618	2.57
19	8	0	0%	32	4.00	1.239	12.286	3.50
20	16	3	18.7%	49	3.06	0.854	11.663	3.41
21	7	0	0%	34	4.86	0.705	3.476	1.86
22	8	0	0%	28	3.50	0.779	4.857	2.20
23	6	2	33.3	28	4.67	1.498	13.467	3.67
24	8	0	0%	34	4.25	0.453	1.643	1.28
25	4	1	25%	16	4.00	1.581	10.000	3.16
26	6	0	0%	25	4.17	0.654	2.567	1.60
27	1	0	0%	2	2.00	.	.	.
28	5	0	0%	24	4.80	1.020	5.200	2.28
29	1	0	0%	4	4.00	.	.	.
30	4	0	0%	23	5.75	1.797	12.917	3.59
31	4	0	0%	29	7.25	1.652	10.917	3.30
32	3	1[1]	33.3%	8	2.67	1.453	6.333	2.51

[1] This, the longest completely fluent utterance in Nymans (and all corpora), is: *Ja dessutom så får vi väl fördelen att då blir vi ju lite mobila för att då har vi den med oss i varje fall om vi vill se nånting av landskapet* ("Yes, and what is more, we'll get the advantage of being somewhat mobile because we'll then have it with us in any case if we'd like to see something of the vicinities").

Table 5.6d. Bionic corpus. For each utterance length, the number and percentages of fluent utterances are given, as are absolute and mean numbers of disfluencies (DFs) per utterance. Standard error, variance and standard deviations are also given.

No. words in utt.	No. utts.	No. fluent utts.	% fluent utts.	No. DFs	Mean DFs per utt.	Std. error	Variance	Std. dev.
1	643	642	99.8%	1	0.002	0.002	0.002	0.04
2	128	105	82.0%	25	0.19	0.038	0.190	0.43
3	167	133	79.6%	40	0.24	0.040	0.268	0.52
4	150	105	70.0%	59	0.39	0.055	0.455	0.67
5	156	97	62.0%	87	0.56	0.071	0.777	0.88
6	106	47	44.3%	100	0.94	0.114	1.368	1.17
7	91	41	45.0%	93	1.02	0.123	1.377	1.17
8	58	18	31.0%	98	1.69	0.210	2.569	1.60
9	50	15	30.0%	90	1.80	0.246	3.020	1.74
10	50	5	10.0%	133	2.66	0.278	3.862	1.96
11	48	8	16.7%	132	2.75	0.353	5.979	2.44
12	41	3	7.3%	101	2.46	0.252	2.605	1.61
13	37	4	10.8%	114	3.08	0.339	4.243	2.06
14	27	3	11.1%	90	3.33	0.503	6.846	2.61
15	29	3	10.3%	109	3.76	0.541	8.475	2.91
16	25	4	16.0%	82	3.28	0.508	6.460	2.54
17	17	2	11.8%	49	2.88	0.541	4.985	2.32
18	20	0	0%	90	4.50	0.845	11.105	3.33
19	13	0	0%	50	3.84	0.783	7.974	2.82
20	17	0	0%	86	5.06	0.678	7.890	2.79
21	11	0	0%	63	5.73	0.810	7.218	2.69
22	8	2	25.0%	28	3.50	0.824	5.429	2.33
23	8	0	0%	38	4.75	0.995	7.929	2.81
24	7	0	0%	43	6.14	1.280	11.476	3.39
25	4	0	0%	21	5.25	2.097	17.583	4.19
26	8	0	0%	67	8.37	1.880	28.268	5.32
27	4	1	25.0%	15	3.75	1.315	6.917	2.63
28	5	0	0%	32	6.40	1.503	11.300	3.36
29	4	0	0%	28	7.00	1.732	12.000	3.46
30	7	0	0%	45	6.43	1.043	7.619	2.76
31	1	0	0%	5	5.00	.	.	.
32	8	0	0%	46	5.75	0.701	3.929	1.98

Table 5.6e. All corpora pooled. For each utterance length, the number and percentages of fluent utterances are given, as are absolute and mean numbers of disfluencies (DFs) per utterance. Standard error, variance and standard deviations are also given.

No. words in utt.	No. utts.	No. fluent utts.	% fluent utts.	No. DFs	Mean DFs per utt.	Std. error	Variance	Std. dev.
1	2694	2680	99.5%	14	0.005	0.001	0.005	0.72
2	797	699	87.8%	120	0.15	0.016	0.196	0.44
3	1102	965	87.6%	167	0.15	0.013	0.194	0.44
4	904	697	77.1%	279	0.31	0.021	0.415	0.64
5	844	574	68.0%	380	0.45	0.026	0.594	0.77
6	682	401	38.8%	444	0.65	0.037	0.935	0.97
7	589	317	53.8%	460	0.78	0.044	1.154	1.07
8	452	222	49.1%	413	0.91	0.056	1.414	1.19
9	382	149	39.0%	471	1.23	0.073	2.027	1.42
10	379	143	37.8%	499	1.32	0.076	2.217	1.49
11	334	88	26.3%	603	1.80	0.096	3.112	1.76
12	277	67	24.2%	524	1.88	0.100	2.798	1.67
13	217	53	24.4%	463	2.13	0.136	4.014	2.00
14	188	36	19.2%	441	2.34	0.150	4.260	2.06
15	202	30	14.9%	541	2.68	0.152	4.697	2.17
16	170	31	18.2%	402	2.36	0.155	4.115	2.02
17	134	24	17.9%	341	2.54	0.189	4.791	2.19
18	104	5	4.9%	322	3.10	0.228	5.428	2.33
19	100	6	6.0%	340	3.10	0.227	5.152	2.27
20	91	6	6.6%	364	4.00	0.084	7.333	2.71
21	72	1	1.4%	301	4.18	0.613	7.192	2.68
22	57	4	7.0%	190	3.33	0.290	4.798	2.19
23	48	8	16.7%	170	3.54	0.374	6.722	2.59
24	44	0	0.0%	202	4.59	0.348	5.317	2.30
25	34	3	8.8%	165	4.85	0.502	8.553	2.92
26	28	0	0.0%	144	5.14	0.759	16.127	4.01
27	20	2	10.0%	101	5.05	0.731	10.682	3.27
28	34	0	0.0%	172	5.06	0.418	5.936	2.44
29	17	0	0.0%	91	5.35	0.641	6.993	2.64
30	24	0	0.0%	143	5.96	0.718	12.389	3.52
31	11	0	0.0%	57	5.18	1.007	11.164	3.34
32	19	1	5.3%	111	5.84	0.852	13.807	3.71

As is seen in the **Table 5.6a** through **Table 5.6e**, the general tendency is that the longer the utterance, the smaller the percentage of completely fluent utterances. Roughly, for all corpora, for utterances about ten words in length, 50% of the utterances are fluent, and 50% contain one or more disfluencies (mean figures indicate about one disfluency per utterance, except for Bionic, which is more disfluent). At twice that length, i.e. twenty words, completely fluent utterances become very rare. However, as was shown, even very long utterances can be

completely fluent. The longest completely fluent utterances in the four corpora were 25 words in WOZ-1, 27 words in WOZ-2 and Bionic, and 32 words in Nymans. Once again, a striking similarity is shown across the corpora.

5.2.1.2 Disfluency frequency as linear regression

As we saw in the previous paragraph, the percentage of utterances that contain at least one instance of disfluency grew with increasing utterance length. This raises the question how much of disfluency incidence can be explained by utterance length alone? Or, to phrase it simply, is disfluency frequency a linear function of utterance length? **Figure 5.1a** through **Figure 5.1d** show the linear regression curves for the individual corpora, while **Figure 5.1e** shows the regression curve for all corpora pooled.

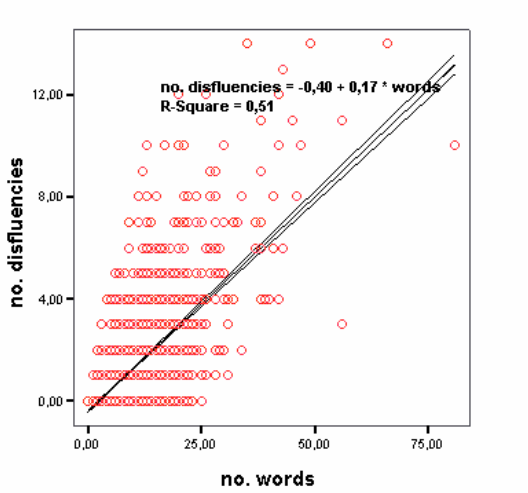


Figure 5.1a. WOZ-1. Linear regression (least-square fit) for the number of disfluencies as a function of utterance length. The exact result is $r = 0.507$. A 95% confidence interval is indicated.

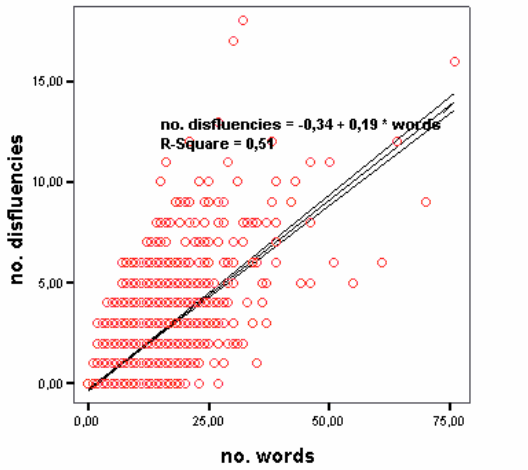


Figure 5.1b. WOZ-2. Linear regression (least-square fit) for the number of disfluencies as a function of utterance length. The exact result is $r = 0.514$. A 95% confidence interval is indicated.

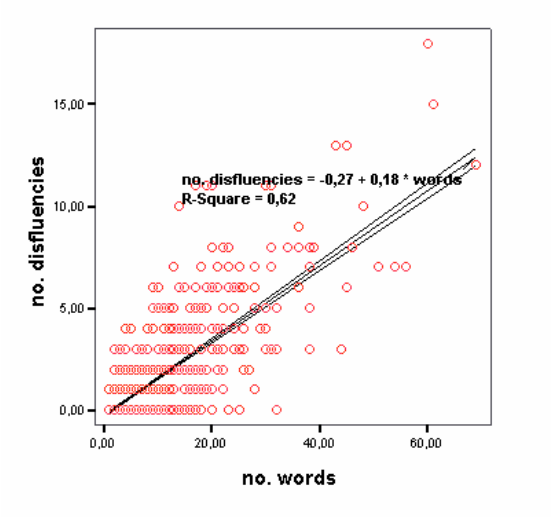


Figure 5.1c. Nymans. Linear regression (least-square fit) for the number of disfluencies as a function of utterance length. The exact result is $r = 0.619$. A 95% confidence interval is indicated.

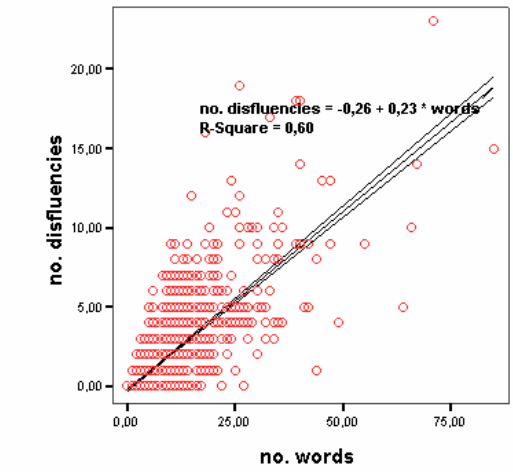


Figure 5.1d. Bionic. Linear regression (least-square fit) for the number of disfluencies as a function of utterance length. The exact result is $r = 0.600$. A 95% confidence interval is indicated.

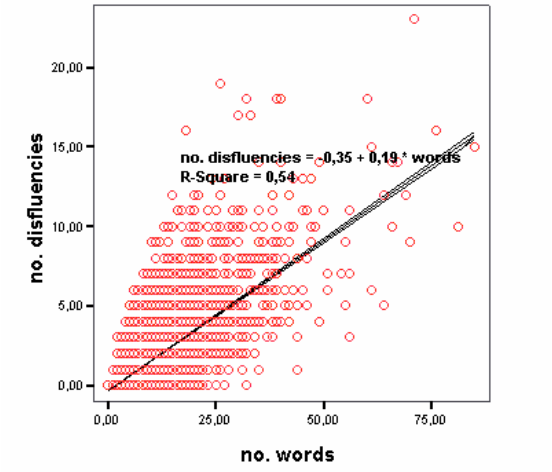


Figure 5.1e. Pooled. Linear regression (least-square fit) for the number of disfluencies as a function of utterance length. The exact result is $r = 0.540$. A 95% confidence interval is indicated.

As is seen in **Figure 5.1a** through **Figure 5.1e**, there is a strong correlation between utterance length and disfluency frequency, expressed as a linear function of utterance length. This is valid for all corpora individually, and for all corpora pooled.

5.2.2 Summary

This section has shown that the longer the utterance, the more likely it is that it will include events of disfluency. For utterances of about ten words, there is a fifty–fifty chance that it contains at least one disfluency, and for an utterance of twenty words, it will almost always contain an instance of disfluency. It was also shown that disfluency frequency to a large degree is a linear function of utterance length, with least-square *r*'s ranging from $r = 0.507$ to $r = 0.619$.

5.3 Unfilled pauses

As was shown in **section 2.18.2**, unfilled pauses do pose a problem, but that does not mean—in my view—that they should be excluded from analysis, especially since they are, by far, the most commonly occurring disfluency phenomenon. As was pointed out earlier, there is no one-to-one relationship between what is perceived by human listeners and what "is there" from an acoustic or physical point of view. To decide what is the preferable method of analysis depends on one's interests and incentives, i.e. whether or not human perception is the main focus of the study in question.

The method employed here included both human perception and an acoustic-physical analysis in that the labeler (the author) labeled pauses from a perceptual perspective, but at the same time had the acoustic signal available to him in the analysis tool. While this latter fact might have biased perception towards "hearing" silences a little too often, just because they were visually salient, at least this method eliminated perceived pauses that did not correspond to silent intervals in the physical energy.

It goes without saying that a superior method would have been to have several people—other than the author—do the labeling,[1] but this method was not available for practical reasons, and it can only be pointed out here how the analysis was done, bearing in mind all the potential flaws this brings with it.

[1] Eklund (1997) included a study where other people than the author labeled the data according to instructions. Inter- and intra-labeler agreement is reported in the poster version of the paper. Interlabeler consensus was found to vary between 100% (for the category STRESS LEVEL FOR IMPLICITLY GIVEN INFORMATION; Labeler 1 vs. Labeler 2) and 22% (for the category STRESS LEVEL FOR NEW INFORMATION; Labeler 1 vs. Labeler 2 vs. Labeler 3). The focus of this study was prosodic stress (prominence), which is known as one of the more problematic areas in linguistic labeling, and certainly more problematic than is disfluency. Trask (1996) summarizes this field: "Native speakers and phoneticians usually find it easy to determine which syllables bear stress, and even to distinguish varying degrees of stress, but the phonetic characterization of stress is exceedingly difficult: stress is variously associated with greater loudness, higher pitch and greater duration, any of which may be most important in a given case, and sometimes also with vowel quality. Earlier attempts to identify stress with greater intensity of sound are now discredited, and current thinking holds that stress is primarily a matter of greater muscular effort by the speaker, and that hearers take advantage of several types of information to identify that effort." (Trask, 1996, p. 336). Consequently, the 1997 figures cannot really be compared to the present study, but should rather serve as a reminder highlighting the fact that everything *perceptual* is subject to inter- and intraindividual variation, and that great methodological care should be taken in the analytical process before far-reaching conclusions be drawn, especially when there is no one-to-one relationship between acoustic-physical and perceptual dimensions.

5.3.1 General frequency

General frequency of unfilled pauses in the corpora is shown in **Table 5.7**.

Table 5.7. General incidence of unfilled pauses (UPs) in the corpora, as well as percentages of utterances and words that include UPs.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. unfilled pauses (UPs)	1622	2179	562	1226	5589
No. utts	4023	3438	1734	1985	11180
No. utts –1-word utts	3117	3087	940	1342	8466
No. words	27664	26261	9250	12849	76024
% UPs/utts	40.3%	63.4%	32.4%	61.8%	49.9%
% UPs/utts –1-word utts	52.0%	70.6%	59.8%	91.4%	66.0%
% UPs/words	5.9%	8.3%	6.1%	9.5%	7.3%

The main finding here is that Nymans exhibits lower overall figures, which most likely is due to more active interaction, given the human interlocutors who were there to fill the silences, when they occurred.

5.3.2 Cross-corpus differences

The next obvious and interesting question is of course whether the corpora significantly differ from each other. Statistical significance is shown in **Table 5.8**.

Table 5.8. Cross-corpus differences for unfilled pauses. Statistical significance is given relative the total number of words (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 5.9% ; 27664	**WOZ-2** 8.3% ; 26261	**Nymans** 6.1% ; 9250	**Bionic** 9.5% ; 12849
WOZ-1	—	$p < 0.05$ (W2)	n.s.	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	$p < 0.05$ (W2)	$p < 0.05$ (B)
Nymans	n.s.	$p < 0.05$ (W2)	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	$p < 0.05$ (B)	$p < 0.05$ (B)	—

As is seen, WOZ-1 is the least disfluent corpus, while Bionic is the most disfluent corpus in all corpus-pairs. The only draw is between WOZ-1 and Nymans, that are roughly equal.

5.3.3 Duration

As was mentioned above, a combination of perceptual and acoustic-physical labeling was used when labeling the data. The durational results for the corpora are given in **Table 5.9**, with and without the lower quartile excised.

Table 5.9. Durational results for unfilled pauses in all corpora given in milliseconds, with a lower cut-off duration of 250 ms, and with the lower quartile excised.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. unfilled pauses (UPs)	**1622**	**2179**	**562**	**1226**	**5589**
Longest UP (ms)	8828	7917	4045	8338	8828
Shortest UP (ms)	70	88	74	79	70
Mean duration (ms)	775	986	621	765	840
Standard deviation (ms)	653	745	511	665	692
Variance (ms)	427	555	261	443	479
No. UPs with 250 ms lower cut-off	**1447**	**2057**	**437**	**1064**	**5005**
Mean duration (ms)	846	1032	746	858	916
Standard deviation (ms)	657	740	513	670	691
Variance (ms)	432	548	263	449	478
No. UPs with lower quartile excised	**1216**	**1634**	**421**	**919**	**4192**
Shortest UP (ms)	363	491	270	348	386
Mean duration (ms)	948	1201	765	946	1031
Standard deviation (ms)	669	741	514	680	699
Variance (ms)	448	549	264	463	489

As should be obvious from the table above, the all-inclusive analysis include unfilled pauses with minimum durations of between 70–90 milliseconds, well above the lower cut-off duration of 50 ms used by Cowan & Bloch (1948) and Martin (1970), as mentioned above. Using the 250 ms lower cut-off—as employed by e.g. Goldman-Eisler (1968)—gives a mean duration just below a second, and with the lower quartile excised, mean duration rises above one second.

Of interest here is that even when using fairly hard pruning, i.e. taking off the lower 25% of the data, unfilled pauses are still by far the most common of all disfluency types (4192 instances, as compared to the runner up filled pauses, with 2601 instances). This clearly shows that any disfluency study that does not include unfilled pauses misses out on the most commonly employed type, whatever method is used (perceptual or automatic), or whatever (realistic) lower cut-off is employed.

5.3.4 Distribution: word classes

Brown (1937, 1940) was a pioneer in pointing out that events of stuttering were not randomly distributed in the speech of stutterers. As was shown in chapter two, nor are the disfluencies in the speech of nonstutterers evenly or randomly distributed. Although one of the more important factors (so I believe) is the particular speech act carried out—something which is not covered in this work—one could study how unfilled pauses are distributed as a function of word class. This is especially interesting since it has been shown, over and over again, over the past fifty years that filled pauses are not affected in the same ways as are all other types of disfluency.

The notion "word classes" is not a straight-forward one, and I will not pretend that I am using a decisive definition here. I will also refrain from diving into a deeper discussion as to the problems associated with word class analysis here. Suffice it to say that word class borders are far from clear-cut. For instance—just to give an example—the word *den* (or *det*, depending on the gender) could be a pronoun ("it"), but could also be an article ("the"). When it is encountered alone in a one-word utterance, or is the last word in a cut-off utterance, it is quite often impossible to label it correctly, or rather know whether the labeling is correct. In addition to that, many words appear in collocations or expressions that could possibly be stored (and produced) as units, rather than on a word-by-word basis, which means that the word classes of the individual words is of less interest than the word class of the entire collocation. Sometimes, it is also hard to tell whether a word is in fact part of an intended collocation or expression, or has another function. To give an example of this, when the preposition *i* ("in", in most cases) appears before a city name, things are clear and easy, but when it is part of the collocation *i alla fall* ("in all cases", "anyway"), a preposition interpretation seems more cumbersome. And so on and so forth.

Consequently, the analysis opted for and presented here must be taken for what it is, i.e. a tad "impressionistic". That being said, I do think that the results and observations—if not absolute or final—still are of interest, in that they in some way point to general *tendencies*. Also, comparisons across corpora are still of interest, since at least all corpora are bestowed with the same type of labeling, however flawed it might be. Moreover, these results become interesting when they are compared to the (corresponding) distribution of other types of disfluency, like filled pauses.

The distribution of unfilled pauses in the corpora, as compared to the following word, is shown in **Table 5.10**.

Table 5.10. Distribution of unfilled pauses (UPs) relative to word classes of the (immediately) following words. Percentages are given relative to the total number of UPs. Open and closed word classes are summarized.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
Total no. unfilled pauses (UPs)	**1622**	**2179**	**562**	**1226**	**5589**
Before open word class words	505 (31.1%)	659 (30.2%)	165 (29.4%)	411 (33.5%)	1740 (31.1%)
Nouns / names	232 (14.3%)	467 (21.4%)	82 (14.6%)	299 (24.4%)	1080 (19.3%)
Verbs	169 (10.4%)	95 (4.4%)	48 (8.5%)	62 (5.1%)	374 (6.7%)
Adjectives / adverbs†	104 (6.4%)	97 (4.4%)	35 (6.2%)	50 (4.1%)	286 (5.1%)
Before closed word class words	931 (57.4%)	1241 (57.0%)	339 (60.3%)	671 (54.7%)	3182 (56.9%)
Prepositions	238 (14.7%)	378 (17.3%)	63 (11.2%)	286 (23.3%)	965 (17.3%)
Conjunctions	281 (17.3%)	232 (10.6%)	88 (15.7%)	109 (8.9%)	710 (12.7%)
Pronouns	163 (10.0%)	152 (7.0%)	64 (11.4%)	104 (8.5%)	483 (8.6%)
Other (pooled)‡	249 (15.3 %)	479 (22.0%)	124 (22.1%)	172 (14.0%)	1024 (18.3%)
Before other disfluency	186 (11.5%)	279 (12.8%)	58 (10.3%)	144 (11.7%)	667 (11.9%)

† Question adverbs, like *hur* ("how"), were included in the "other" category.

‡ Numerals, ordinals, articles, determiners, interjections, infinitival marker, question words and so on.

Judging from the these results, it would seem that unfilled pauses do not exhibit a very strong proneness toward specific locations in the sentence. However, there is a small tendency towards unfilled pauses occurring immediately before nouns (or names) or prepositions,

which hints at some kind of 'NP-attraction' factor at play. While this does not show strongly up in WOZ-1 or Nymans, where conjunctions are the preferred follower of unfilled pauses, it is marked in WOZ-2 and very strong indeed in Bionic, where almost 50% of all unfilled pauses are found immediately preceding either a noun/name or a preposition. At the bottom end of the scale we find that adjectives/adverbs, verbs (all comparatively rare in the data) and pronouns are dispreferred locations.

However, so far we have only studied the distribution from a disfluency perspective: if we have an unfilled pause in our hands, what is likely to follow? This, of course, could be the result of the general relative frequencies of the said word classes in the data, i.e., given that unfilled pauses are completely randomly distributed, in this case meaning, completely evenly distributed in the utterances, this is exactly the kind of results we would obtain. What we would like to know is whether any of the word classes are under- or over-represented as followers to unfilled pauses. Thus, we need to know the relative frequency of word classes in the corpora. These are shown in **Table 5.11** (once again, the same provisos and caveats, are at play, but at least they should be more or less the same as in the previous table, and in that way making the comparisons at least internally consistent).

Table 5.11. Frequency distribution of word classes in the corpora given as total numbers and percentages.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
Total no. words	27664	26261	9250	12849	76024
Open word class words	18382 (66.5%)	11342 (43.2%)	3218 (34.8%)	5625 (43.8%)	38567 (50.7%)
Nouns / names	7381 (26.6%)	5851 (22.3%)	967 (10.4%)	2247 (17.5%)	16446 (21.6%)
Verbs	6219 (22.5%)	4450 (16.9%)	1636 (17.7%)	2161 (16.8%)	14466 (19.0%)
Adjectives / adverbs†	4782 (17.3%)	1041 (4.0%)	615 (6.7%)	1217 (9.5%)	7655 (10.1%)
Closed word class words	9262 (33.5%)	14919 (56.8%)	6032 (65.2%)	7224 (56.2%)	37457 (49.3%)
Prepositions	1023 (3.7%)	2723 (10.4%)	523 (5.7%)	1307 (10.2%)	5576 (7.3%)
Conjunctions	1354 (4.9%)	2008 (7.6%)	1241 (13.4%)	848 (6.6%)	5451 (7.2%)
Pronouns	1625 (5.9%)	3751 (14.3%)	1712 (18.5%)	1427 (11.1%)	8515 (11.2%)
Other (pooled)‡	5280 (19.1%)	6437 (24.5%)	2556 (27.6%)	3642 (28.3%)	17915 (23.6%)

† Question adverbs, like hur ("how"), were included in the "other" category.

‡ Numerals, ordinals, articles, determiners, interjections, infinitival marker, question words and so on. Nota bene! The filled pause was included in this category in this count.

As is seen, there is a 50–50 distribution of words belonging to open and closed word classes at a token level when all the data are pooled. There are, however, some differences between the corpora at a more detailed level. For example, the distribution of open/closed words in WOZ-1 is the reverse of the distribution in Nymans (WOZ-2 and Bionic exhibit the same pattern), and pronouns are more common in the human–human corpus, as are conjunctions. Although all such notions require a detailed syntactic analysis—which is not carried out here— this is an indication that there are linguistic differences between the corpora, and that people do structure their messages according to who they think the interlocutor is.

So, how does the general distribution affect the previous observations on unfilled pause distribution? If unfilled pauses were completely evenly distributed in the corpora, then the figures in **Table 5.10** would follow closely the figures in **Table 5.11**. Given the complications

associated with the present, not very detailed, analysis of word classes, the observations made here must be taken very cautiously, and suggestive at best. Still, I will make a few comments.

Unfilled pauses that precede nouns (19.3%) correspond well to the overall incidence of nouns in the corpora (21.6%), thus decreasing the 'NP-attraction' factor slightly. On the other hand, the percentage of unfilled pauses that precede conjunctions (12.7%) is close to double the percentage of general occurrence of conjunctions in the corpora (7.2%), thus making the case for a 'pre-sub-clause location hypothesis' stronger. What is more, when looking at unfilled pauses that precede prepositions (17.3%), as compared to the general incidence of prepositions in the corpora (7.3%), it seems as if the 'NP hypothesis' is back into consideration, only that unfilled pauses seem to precede the entire NP, rather than splitting up the preposition and the ensuing noun.[1]

It would seem, then, that unfilled pauses are fairly evenly distributed within an utterance, albeit not *completely* evenly distributed. The tendencies found are that unfilled pauses seem to precede either prepositions or nouns, making them 'NP-prone'—most notably so in WOZ-1 and Bionic—or appear immediately preceding conjunctions, which shows up in WOZ-2 and Nymans, which would hint at some kind of structuring function of unfilled pauses, in that they appear before subordinate or conjoined clauses. Whether or not this latter phenomenon is "planned" or not is a matter of speculation, of course. However, the former observation, that unfilled pauses appear before, or inside, noun phrases could well hint at some kind of hesitation when semantically heavy items are being chosen.

5.3.5 Summary

The general pattern for all disfluencies merged was more or less present when looking specifically at unfilled pauses. WOZ-1 is the least disfluent corpus, while Bionic is the most disfluent corpus. This is probably—at least partly—a function of the shorter utterances in WOZ-1 as compared to the long utterances in Bionic.

As for distribution, we found that although unfilled pauses seem to be rather equally distributed, there is a noticeable tendency for them to appear before or in noun phrases, or immediately before conjunctions, thus confirming the observation by Hawkins (1971) that pauses were located at clause boundaries. In and by themselves, these results might not be startling, but we shall see in the next paragraph that there is more to the story.

5.4 Filled pauses

As was shown in chapter two, an oft-repeated and consistent observation in the literature "is that unfilled and filled pauses are very different creatures" (Christenfeld, 1994, p. 193). As we saw, already Mahl made a distinction between filled pauses and "non-ah" disfluencies. Filled pauses have been said to serve the function of floor-holding while planning further speech, or simply stressing that second point, that they mainly signal planning problems of a more general kind, i.e. that they appear when the speaker is facing many possible ways to continue—or begin—speaking. Or, as Christenfeld put it: "[A]dding options increases filled

[1] Statistical analysis was not carried out here, since the data are derived from such a "impressionistic" and indirect labeling method. To supply statistical tests here would only potentially deceive the reader to take these figures as more than suggestive, which is what they are.

pauses" (Christenfeld, 1994, p. 197). This notion was further supported by Schachter et al. (1991), who found that filled pauses occurred as a function of number of choices present to the speaker at a given moment. This hypothesis would explain that filled pauses most frequently occur at the beginning of utterances, where no commitment has yet been made as to the contents of that particular utterance. Moreover, Cook (1971) observed that filled pauses tended to occur at the beginning of a clause, or more specifically before the first, second or third word in the entire utterance.

It has also been suggested that filled pauses occur as a function of self-monitoring, e.g. by Christenfeld (1996), who pointed out that:

> An approach to filled pauses that seems more successful is the notion that they reflect not anxiety or task difficulty, but rather speakers' concern with their speech. That is, when people are monitoring what they say, they may be more likely to say "um." (Christenfeld, 1996, p. 1233.)

So, what seems to be the case in the present data set?

5.4.1 General frequency

The general frequency of filled pauses is shown in **Table 5.12**. Two factors of interest have been considered. The first is one-word utterances, that occasionally consists of just a filled pause, but most often does not contain a disfluency. The second factor is the large number of utterance-initial filled pauses. Both these parameters are covered in the table.

Table 5.12. General incidence of filled pauses (FPs) in the corpora, as well as percentages of utterances and words that include FPs. Since FPs count as words (being vocalizations), and consequently are included in the words row, a separate row for number of words minus FPs is included. The number of utterance-initial FPs (UIFPs) is given for each corpus, as are the percentages of utterances that include UIFPs (i.e., utterances that begin with an FP).

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. filled pauses (FPs)	815	1040	203	543	2601
No. utterance-initial FPs (UI-FPs)	376	478	94	230	1178
No. utts	4023	3438	1734	1985	11180
No. utts –1-word utts	3117	3087	940	1342	8466
No. words	27664	26261	9250	12849	76024
No. words –FPs	26849	25221	9047	12306	73423
% FPs/utts	20.2%	30.2%	11.7%	27.3%	23.3%
% FPs/utts –1-word utts	26.1%	33.7%	21.6%	40.5%	30.7%
% FPs/words	2.9%	3.9%	2.19% *	4.2%	3.4%
% FPs/words –FPs	3.0%	4.1%	2.24% *	4.4%	3.5%
% UIFPs/FPs	46.1%	46.0%	46.3%	42.6%	45.3%
% UIFPs/utts	9.3%	13.9%	5.4%	11.6%	10.5%
% UIFPs/words	1.36%	1.82%	1.02%	1.79%	1.55%
% UIFPs/words –FPs	1.40%	1.89%	1.04%	1.87%	1.60%

* Given with two decimals to show the difference.

The first, striking, observation about filled pauses is their distribution: almost 50% occur in utterance-initial position, and the proportion of utterance-initial filled pauses is extremely stable across the corpora, showing that this tendency is strong.

5.4.2 Cross-corpus differences

So, are there any significant differences between the corpora from a filled pause perspective. The results are shown in **Table 5.13a** through **Table 5.13d**, broken down for utterance-initial and other filled pauses, and for total word counts including and excluding filled pauses.

Table 5.13a. Cross-corpus differences for filled pauses. Statistical significance is given relative the total number of words (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 2.9% ; 27664	**WOZ-2** 3.9% ; 26261	**Nymans** 2.19% ; 9250	**Bionic** 4.2% ; 12849
WOZ-1	—	$p < 0.05$ (W2)	$p < 0.05$ (W1)	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	$p < 0.05$ (W2)	n.s.
Nymans	$p < 0.05$ (W1)	$p < 0.05$ (W2)	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	n.s.	$p < 0.05$ (B)	—

Table 5.13b. Cross-corpus differences for filled pauses. Statistical significance is given relative the total number of words minus the number of filled pauses (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 3.0% ; 26849	**WOZ-2** 4.1% ; 25221	**Nymans** 2.24% ; 9047	**Bionic** 4.4% ; 12306
WOZ-1	—	$p < 0.05$ (W2)	$p < 0.05$ (W1)	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	$p < 0.05$ (W2)	n.s.
Nymans	$p < 0.05$ (W1)	$p < 0.05$ (W2)	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	n.s.	$p < 0.05$ (B)	—

Table 5.13c. Cross-corpus differences for utterance-initial filled pauses. Statistical significance is given relative the total number of words (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 1.36% ; 27664	**WOZ-2** 1.82% ; 26261	**Nymans** 1.02% ; 9250	**Bionic** 1.79% ; 12849
WOZ-1	—	$p < 0.05$ (W2)	$p < 0.05$ (W1)	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	$p < 0.05$ (W2)	n.s.
Nymans	$p < 0.05$ (W1)	$p < 0.05$ (W2)	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	n.s.	$p < 0.05$ (B)	—

Table 5.13d. Cross-corpus differences for utterance-initial filled pauses. Statistical significance is given relative the total number of words minus the number of filled pauses (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 1.4% ; 26849	**WOZ-2** 1.89% ; 25221	**Nymans** 1.04% ; 9047	**Bionic** 1.87% ; 12849
WOZ-1	—	$p < 0.05$ (W2)	$p < 0.05$ (W1)	$p < 0.05$ (B)
WOZ-2	$p < 0.05$ (W2)	—	$p < 0.05$ (W2)	n.s.
Nymans	$p < 0.05$ (W1)	$p < 0.05$ (W2)	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	n.s.	$p < 0.05$ (B)	—

Interestingly, a different pattern emerges here. While the general pattern for disfluency production is WOZ-1 < WOZ-2 < Nymans < Bionic, here Nymans stands out as the least disfluent corpus, for all four different counts.

So, why do the subjects use significantly fewer filled pauses in the human–human corpus? If, as it has been proposed, filled pauses serve the function of floor-holding, then their number should be higher in the human–human setting, where the risk of being interrupted is much higher than in any of the human–machine (real or faked) settings, since the automatic systems (real or staged) in no case were very verbose.

On the other hand, if filled pauses signal general speech planning problems, then their number should be higher when the subjects are having problems with the planning of their booking.

This seems to lead to the conclusion that filled pauses are not uttered in order to "keep the floor", at least not in these corpora. Moreover, since their number is lower, it could be the case that the subjects face fewer problems in the planning of their trips, which could be due to help provided by the (experienced) agent, who—presumably—would not leave the customer

hanging on, more or less confused, on the phone, without trying to suggest something to help them out of their problems.

So, the cross-corpus comparison seems to provide support for the notion of filled pauses as a signal of general speech planning problems, whereas the (more or less) alternative notion of filled pauses as a floor-holder seems to be contradicted by the present findings.

5.4.3 Duration

So, it was shown that filled pauses appeared differently in the four corpora, as compared to how unfilled pauses appeared in the data. The next issue at hand is whether or not there are any durational characteristics to distinguish filled pauses from unfilled pauses. The results are shown in **Table 5.14**.

Table 5.14. Durational results for filled pauses in all corpora in milliseconds.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. filled pauses (FPs)	815	1040	203	543	2601
Longest FP (ms)	1651	2175	1108	2037	2175
Shortest FP (ms)	97	92	68	86	68
Mean duration (ms)	465	490	490	489	483
Standard deviation (ms)	230	252	228	256	245
Variance (ms)	53	64	52	65	60

As is seen in the table, the mean values are surprisingly stable across the corpora, showing that a typical filled pause is just around 500 ms long (±250 ms). The shortest filled pauses are well above the shortest unfilled pauses, as included by Cowan & Bloch (1948).

5.4.3.1 … as compared to unfilled pauses?

So are filled pauses comparable to unfilled pauses as to duration? A *t*-test on the two full sets showed that unfilled pauses are significantly longer than filled pauses ($p<0.001$; two-tailed, equal variances assumed). Since it not entirely clear whether filled pauses and unfilled pauses are to be considered as dependent variables or not, it was decided to run nonparametric tests as well, and a Wilcoxon signed ranks test and a Mann-Whitney test were thus also performed. Both showed significance at the $p<0.001$ level. So, unfilled pauses are generally longer than filled pauses.

5.4.4 Distribution: word classes

As was discussed for unfilled pauses, disfluency distribution within an utterance does not follow random distribution. It was shown that unfilled pauses exhibited a small tendency to precede noun phrases, although this tendency was markedly stronger in WOZ-1 and Bionic than in WOZ-1 or Nymans, where the preferred location for unfilled pauses was immediately preceding conjunctions.

The question is, of course, whether filled pauses follow the same general pattern, or whether differences between the two types of disfluency may show up. The distributional results of filled pauses are shown in **Table 5.15**.

Table 5.15. Distribution of filled pauses (FPs) relative to word classes of the (immediately) following words. The results are broken down for utterance-initial FPs, as opposed to FPs in other positions. Both nominal figures and percentages are given. For both categories of FPs, the number and percentages of FPs immediately followed by other types of disfluencies are given. (Note that the sums and percentages in the first part of the table do not always add up overall totals, since not all FPs are followed by other items.)

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
Total no. filled pauses (FPs)	**815**	**1040**	**203**	**543**	**2601**
Before open word class words	206 (25.3%)	259 (24.9%)	39 (19.2%)	133 (24.5%)	637 (24.5%)
Nouns / names	90 (11.0%)	189 (18.2%)	24 (11.8%)	94 (17.3%)	397 (15.3%)
Verbs	109 (13.4%)	58 (5.6%)	9 (4.4%)	34 (6.3%)	210 (8.1%)
Adjectives / adverbs	7 (0.8%)	12 (1.2%)	6 (2.9%)	5 (0.9%)	30 (1.1%)
Before closed word class words	329 (40.4%)	481 (46.2%)	97 (47.8%)	244 (44.9%)	1151 (44.2%)
Prepositions	28 (3.4%)	72 (6.9%)	8 (3.9%)	35 (6.4%)	143 (5.5%)
Conjunctions	13 (1.6%)	9 (0.9%)	10 (4.9%)	11 (2.0%)	43 (1.6%)
Pronouns	142 (17.4%)	203 (19.5%)	34 (16.7%)	106 (19.2%)	485 (18.6%)
Other*	146 (17.9%)	197 (18.9%)	45 (22.2%)	92 (16.9%)	480 (18.4%)
Before other disfluency	256 (31.4%)	267 (55.5%)	44 (21.7%)	126 (23.2%)	693 (26.7%)
No. utterance-initial filled pauses	**376**	**487**	**94**	**230**	**1178**
Before open word class words	83 (22.1%)	83 (17.0%*)	20 (21.3%)	59 (25.6%)	245 (20.8%)
Nouns / names	21 (5.6%)	49 (10.1%)	11 (11.7%)	44 (19.1%)	125 (10.6%)
Verbs	58 (15.4%)	30 (6.2%)	7 (7.4%)	15 (6.5%)	110 (9.3%)
Adjectives / adverbs	4 (1.1%)	4 (0.8%)	2 (2.3%)	—	10 (0.8%)
Before closed word class words	196 (52.1%)	307 (63.0%*)	62 (65.9%)	146 (63.5%)	711 (60.3%)
Prepositions	6 (1.6%)	33 (6.8%)	2 (2.2%)	9 (3.9%)	50 (4.2%)
Conjunctions	2 (0.5%)	1 (0.2%)	4 (4.2%)	5 (2.2%)	12 (1.0%)
Pronouns	93 (24.7%)	174 (35.8%)	24 (25.5%)	77 (33.5%)	368 (31.2%)
Other	95 (25.3%)	99 (20.3%)	32 (34.0%)	55 (23.9%)	281 (23.8%)
Before other disfluency	97 (25.8%)	97 (19.9%*)	12 (12.8%)	25 (10.9%)	231 (19.6%)

* These figures do not add up to 100 with only one decimal point given.

The most striking observation is that for all filled pauses, the by far most common location is immediately before another *disfluency*. This observation is most apparent in WOZ-2, where more than fifty percent of the filled pauses are followed by another type of disfluency. For utterance-initial filled pauses, the most common item following the filled pause is a pronoun—most often *jag* ("I") or referential *det* or *den* ("it")—with "before other disfluency" as the runner-up. Compared to the general incidence of pronouns (11.2%, as shown in **Table 5.11**), their occurrence as followers of filled pauses (18.6% for all filled pauses; 31.2% for utterance-initial pauses) clearly shows that this is a preferred position.

The observation that filled pauses so frequently are followed by more disfluency could be taken as evidence that filled pauses signal locations in an utterance where speech planning is problematic in general. However, if this figure (19.6% for the pooled data) is compared to the general incidence of disfluency (**Table 5.2**), one has to decide exactly what general incidence figure to compare with, since the four different counts used to calculate general disfluency occurrence range from 6.4% (excluding unfilled pauses; including all words) to 14.1%

(including unfilled pauses; excluding one-word utterances). Although the figure (19.6%) in all cases is higher than the disfluency figure in general, it must be borne in mind that different ways of counting yields different "conclusions".

One can also study where filled pauses do *not* occur (which is partly the reason the table looks the way it does). For example, virtually no filled pauses appear immediately before conjunctions, 1.6% and 1.0%, as compared to their general frequency of 7.2% (as shown in **Table 5.11**), showing that this is a dispreferred location for filled pauses to appear, in contrast to what was observed for unfilled pauses. This hints at an interesting notion that while filled pauses are common when starting an utterance, they are not common when beginning a clause—conjoined or subordinate—within that utterance. Filled pauses virtually never appears immediately before adjectives/adverbs (1.1% and 0.8%), which partly is the effect of the relative dearth of adjectives in the data, but that position is still underrepresented, given a general adjective/adverb incidence of around 10%.

It goes without saying that all these observations need be taken cum grano salis, and that much more detailed analyses are needed before any far-reaching conclusions can be drawn. Above all, any distributional analysis needs to take into account the general characteristics of the linguistic material per se in the analyzed corpora, i.e. in what ways the specific vocabulary and syntax is dependent on the domain and the specific tasks given the subjects. However, the observation that filled pauses (most) frequently appear immediately before another instance of disfluency hints at the possibility that filled pauses appear at locations in the utterance where more global planning problems occur. In conclusion, then, while unfilled pauses tend to be attracted to noun phrases, filled pauses seem to be some kind of 'disfluency-disfluency', i.e. signaling higher-level speech planning problems.

As was the case with unfilled pauses, what would be of more interest would be a syntactical analysis, i.e. to see how filled pauses appear relative to syntactic constituents. Also, a functional analysis would be of interest in order to further map the distribution of filled pauses. Both these analyses would require a further analysis and labeling of the data that at present has not been carried out, however interesting that would be.

5.4.5 Summary

So, filled pauses, that so often have been shown to stand out from all other disfluency types, do not seem to let us down in this study.

First (and perhaps less interesting), is that their durations are significantly shorter than the durations of unfilled pauses. This could, of course, indicate that speakers are more inclined to pursue speaking while still sounding, whereas silence does not commit a speaker as much to continue producing speech.

Second, and more interestingly, their distribution differs from that of unfilled pauses, in that filled pauses most often seem to appear in conjunction with other disfluencies, making them some kind of 'disfluency-disfluency'—as compared to the 'subordinate-clause-disfluency' or 'NP'disfluency' characteristics of unfilled pauses—indicating major speech planning problems on behalf of the subject.

Moreover, the observation that filled pauses are significantly less common in Nymans, the human–human corpus, as compared to all human–machine (real or fake) corpora, seems to

gainsay the notion of filled pauses as a floor-holder, while at the same time seemingly lends support to the aforementioned notion of filled pauses as a general speech planning problem indicator, since it could be assumed that travel planning was less problematic in the human–human setting, where real, live agents could help out in any presumably problematic situation.

To test that latter hypothesis, it would be interesting to carry out speech analysis of the subject–agent interaction, which could easily be done, and would only require transcription of the agents' utterances, as well.

It would of course also be interesting to gauge the anxiety of the subjects (palmar sweating, galvanic skin response, etc.), but that, alas, lies beyond our present means.

5.5 Prolongations[1]

Prolongations were long included in the category *dysrhythmic phonations*, and were considered a tell-tale sign of stuttering, as illustrated by e.g. the following, relatively recent, quote in Adams, Sears & Ramig (1982): "Part-word repetitions and disrhythmic phonations in the form of audible sound prolongations were further classified as stuttering, because these behaviors have consistently been identified as the universally demonstrable features of the disorder" (Adams, Sears & Ramig, 1982, p. 24). However, later studies have shown that prolongation is wide-spread in nonstuttered speech and that it is also to a fair degree universal (Eklund 2001, 2000a; Den 2003). Rialland & Robert (2001), in their study of the intonational system of Wolof, include "loud pauses, marked by vowel lengthening" (Rialland & Robert, 2001, p. 923; see also p. 929). Allwood and colleagues (Allwood, 1998b) mention prolongation of continuants, but the studies of Eklund (2001, 2000) reveal that all types of phones are subject to prolongations, not only continuants. It is likely that the labeling of the Göteborg data has been subject to "written-language bias", and that only segments that are intuitively easy to prolong in writing, like [s], [m] or [f], have been seen as possible to prolong in speech. By this I mean that a pronunciation that includes a prolonged [s], for example, as in *Jag skulle vilja ha en bussssssssbiljett* ("I would like a bussssss ticket") is easy to represent in writing (shown here), while other segments, such as stops, do not as easily lend themselves to written representation, like [t] in *Jag skulle vilja ha en bussbilje….tt* ("I would like a bus ticke….t"). Consequently, when perusing the literature, it seems that oftentimes only continuants have been considered possible to prolong. while e.g. stops have not been considered during the labeling.[2]

[1] This section draws heavily on Eklund (2001). The major difference is that this chapter is based on more data.

[2] My guess is that probably are far more prolongations to be found in the Göteborg corpora, if further analysis would be carried out, something which Allwood agrees to (personal communication).

5.5.1 General prolongation rates

Occurrence of prolongations in the corpora is shown in **Table 5.16**.

Table 5.16. General incidence of prolongations (PRs) in the corpora, as well as percentages of utterances and words that include PRs.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. PRs	156	132	129	197	614
No. utts	4023	3438	1734	1985	11180
No. utts –1-word utts	3117	3087	940	1342	8466
No. words	27664	26261	9250	12849	76024
% PRs/utts	3.9%	3.8%	7.4%	9.9%	5.5%
% PRs/utts –1-word utts	5.0%	4.3%	13.7%	14.7%	7.2%
% PRs/words	0.6%	0.5%	1.4%	1.5%	0.8%

As can be seen, more than five percent of all utterances include prolongations, and between 0.5% and 1.5% of all words (at token levels) include prolonged segments. Comparing with Eklund (2001), which was based on a subset of the present data set, all figures are similar with the exception of WOZ-1, which had much higher figures (17.72% of all utterances and 1.81% of all word tokens) than the present figures. The explanation for this discrepancy is probably to be found in the proportion of the corpora that were fully transcribed and labeled at the time. While Nymans and Bionic were fully analyzed at the time, and while WOZ-2 had 71 of its 137 dialogues transcribed (more than half), WOZ-1 had only had 84 of its 433 dialogues transcribed at the time, so it might not come as a surprise that those figures are those that have changed the most. This is even more likely since there is a good chance that those subjects who appeared the "easiest" to transcribe were probably chosen in the beginning phases of transcription, in order to "test" the transcription tool developed for the purpose. This is also a good reminder that data that is not randomly selected might yield not-so-representative results.

Summing up, more than one utterance in twenty includes at least one prolonged sound, and roughly one percent of the words uttered contain prolongations. Moreover, instead of being an alleged tell-tale sign of stuttering, prolongation is the third most common (atomic, non-composite) type of disfluency in our data (excluding the composite category repairs, which are slightly more frequent than prolongations in all corpora and pooled).

5.5.2 Cross-corpus differences

So, in medias res, are there any significant differences between the corpora with regard to prolongation. The results are shown in **Table 5.17**.

Table 5.17. Cross-corpus differences for prolongations. Statistical significance is given relative the total number of words (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 0.6% ; 27664	**WOZ-2** 0.5% ; 26261	**Nymans** 1.4% ; 9250	**Bionic** 1.5% ; 12849
WOZ-1	—	n.s.	$p < 0.05$ (N)	$p < 0.05$ (B)
WOZ-2	n.s.	—	$p < 0.05$ (N)	$p < 0.05$ (B)
Nymans	$p < 0.05$ (N)	$p < 0.05$ (N)	—	n.s.
Bionic	$p < 0.05$ (B)	$p < 0.05$ (B)	n.s.	—

Here yet another pattern emerges in that the four corpora form two distinct groups with significant inter-group differences, but no intra-group differences. WOZ-1 and WOZ-2 are thus similar in the frequency with which prolongation occurs, while Nymans and Bionic from another group, with distinctly higher rate of prolongation.

Why this is, however, is not instantly obvious. The only obvious hypothesis would be that prolongation in some way occurs more often in more natural settings, such as human–human communication, or when speakers are allowed to speak out very long utterances, and are still understood, i.e. not noticing any "upper constraints" on their way of expressing their ideas. This, however, must remain completely conjectural for the time being.

Finally, one should not forget that the observed prolongation-per-word frequency is not overwhelmingly high, and that the results presented here might well be a fluke.

5.5.3 Duration

The mean durations for all prolongations are given in **Table 5.18**. Since durational analysis is cumbersome at lower end (when is a sound prolonged?), figures are given both for the full data set (all prolongations), with the lower-end quartile excised.

Table 5.18. Mean duration of prolonged sounds (PRs) in milliseconds, broken down for all corpora and pooled. Durations are given for both the full data set, and with the lower quartile trimmed away.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. PRs	**156**	**132**	**129**	**197**	**614**
Longest PR (ms)	1239	2454	813	1138	2454
Shortest PR (ms)	77	91	84	87	77
Mean duration (ms)	298	383	276	280	306
Standard deviation (ms)	201	316	118	171	213
Variance (ms)	40	100	14	29	45
No. PRs with lower quartile excised	**117**	**98**	**97**	**149**	**460**
Shortest PR (ms)	169	231	204	164	179
Mean duration (ms)	353	461	314	329	361
Standard deviation (ms)	203	333	111	170	219
Variance (ms)	41	111	12	29	48

As is shown, while WOZ-1, Nymans and Bionic resemble each other, WOZ-2 exhibits longer durations on the whole. It must be borne in mind, however, that this difference is barely perceivable to the human ear, and that the largest difference, that between a mean value of 461 ms for the trimmed data, as opposed to 314 ms in Nymans (trimmed mean) corresponds to roughly a seventh of a second. Moreover, this is the trimmed *mean* value—WOZ-1 also exhibits the (by far) largest standard deviation and variance—so one might instead be astonished over the fact that the other corpora display well-nigh identical values.

5.5.4 Prolongations vs. filled pauses

Prolongations and filled pauses have in common that they both constitute examples of hesitation by means of *both vocalization and duration*, setting them aside from all other disfluency types. This means that if they both serve the main purpose of floor-holding, are there any durational differences between them?

5.5.4.1 Durational differences

As was argued in Eklund (2001), although both filled pauses and prolongations are produced the same way—sustained vocalization—there is little reason to assume *a priori* that there would be any durational difference between the two types. A *t*-test on the full sets of prolongations and filled pauses revealed that filled pauses are significantly longer than prolongations ($p<0.001$; two-tailed, equal variances assumed). Since it not entirely clear whether or not filled pauses and prolongations are to be considered as dependent variables or not, a Wilcoxon signed ranks test and a Mann-Whitney test were also performed. Both tests showed significance at the $p<0.001$ level. It seems, then, that one could safely assume that filled pauses are longer than prolongations.

5.5.4.2 Individual preferences?

Since both filled pauses and prolongations signal hesitation while still "holding the acoustical floor", as it were, it is of interest to find out whether there are individual preferences as to disfluency type in this respect. Individual differences are shown in **Table 5.19**.

Table 5.19. Relative frequency of prolongation (PR) and filled pause (FP) usage. The numbers of subjects who: use more FPs than PRs; use more PRs than FPs; use an equal number of FPs and PRs; do no FPs; or do not use PRs. Note that the column sums sometimes exceed the number of subjects since the same subject appears in more than one cell when the lower figure is zero, e.g. when the number of FPs is 12 and the number of PRs is zero (WOZ-2, subject 22). The symbol ">" means "more frequent than".

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. subjects	46	46	8	16	116
Σ FPs	815	1040	203	543	2601
Σ PRs	156	132	129	197	614
FPs > PRs	40	43	7	13	103
PRs > FPs	4	2	1	3	7
FPs = PRs	2	1	—	—	3
Only PRs (no FPs)	3	1	—	—	4
Only FPs (no PRs)	8	14	—	—	22

As can be seen, the overwhelming majority of the subjects prefer filled pauses to prolongation (see **Appendix 1** through **Appendix 4** for results broken down for subjects). Moreover, a closer look at the data reveals that most subjects exhibit far more filled pauses than prolongations. Only seven subjects use more prolongations than filled pauses, and only three use an equal amount of filled pauses and prolongations. Interestingly, one subject (WOZ-1, subject 26) does not exhibit either filled pauses or prolongations. He is also among the most fluent of all subjects studied, and besides one single repair, he only uses unfilled pauses, some of which might even be planned. In any case, this shows that some speakers are extremely fluent.

5.5.5 Position within the word

It has been shown that prolongations are not evenly distributed within the word, either in Swedish (Eklund, 1999, 2000a; Eklund & Shriberg, 1998), American English (Eklund & Shriberg, 1998), Japanese (Den, 2003) or in Tok Pisin (Eklund, 1999). Also, phonological category plays a role, so that certain types of phones are more prone than other types of phones in certain positions. Prolongations, broken down for position and type, are shown in **Table 5.20**.

Table 5.20. Prolongation position and phone type for all corpora. For each corpus the number and percentages are given. For each position, number and percentages are broken down for vowels, sonorant consonants and non-sonorant consonants. Note that the number of segments is based on phonological forms—and thus marginally approximate—since phonetic reductions were not covered in the transcription.[1] Also note that the one-segment (vowel) word *i* ("in") was counted in all three categories (initial, medial, final). Consequently, sum totals exceed the number of prolongations (PRs). Percentages for the proportions initial–medial–final were calculated on the number of prolongations obtained when each instance of *i* was counted thrice.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. PRs	156	132	129	197	614
No. Phonol. segments	121,480	116,774	35,171	53,821	327,246
% PRs/Segments	0.13%	0.11%	0.37%	0.37%	0.19%
N/% Word-initial PRs	43 (25.9%)	45 (30.8%)	33 (25.4%)	63 (29.0%**)	184 (27.9%)
N/% Vowels	8 (18.6%)	9 (20.0%)	7 (21.2%*)	12 (19.0%)	36 (19.6%)
N/% C +son	2 (4.7%)	10 (22.2%)	8 (24.2%*)	17 (27.0%)	37 (20.1%)
N/% C –son	33 (76.7%)	26 (57.8%)	18 (54.5%*)	34 (54.0%)	111 (60.3%)
N/% Word-medial PRs	32 (19.3%)	34 (23.3%)	25 (19.2%)	40 (18.4%**)	131 (19.9%)
N/% Vowels	12 (37.5%)	11 (32.3%)	5 (20.0%)	14 (35.0%)	42 (32.1%)
N/% C +son	2 (6.2%)	1 (3.0%)	4 (16.0%)	1 (2.5%)	8 (6.1%)
N/% C –son	18 (56.3%)	22 (64.7%)	16 (64.0%)	25 (62.5%)	81 (61.8%)
N/% Word-final PRs	91 (54.8%)	67 (45.9%)	72 (55.4%)	114 (52.5%**)	344 (52.2%)
N/% Vowels	31 (34.1%)	18 (26.9%)	25 (34.7%)	43 (37.8%)	117 (34.0%)
N/% C +son	38 (41.7%)	42 (62.7%)	36 (50%)	60 (52.6%)	176 (51.1%)
N/% C –son	22 (24.2%)	7 (10.4%)	11 (15.3%)	11 (9.6%)	51 /14.9%)

* These figures do not add up to 100%, but moves asymptotically towards closer (21.21212121…, 24.24242424… and 54.54545454…, respectively).

** These figure do not add up to 100% with only one decimal point.

As can be seen in **Table 5.20**, the 30–20–50 ratio for initial–medial–final phones is repeated. However, as was pointed out in Eklund (2001), a detailed look will reveal that specific phones are differently prone to prolongation in specific positions. The least common combination of position in the word/phone type is sonorant consonants in word-medial position (6.1% in the pooled data). The most commonly prolonged segments are non-sonorant consonants in medial position (61.8%, all data pooled) and initial position (60.3%, all data pooled).

5.5.6 Top-five phones

Intuition tells you that certain phones should be more prone to prolongation than others, for acoustic-motoric reasons. Some phones, like continuants, are simply easier to prolong. However, as was shown in Eklund (2001), continuants were not the only phones subject to prolongation. **Table 5.21** shows the top five segments broken down for corpora.

[1] The figures given here differ from the figures given in Eklund (2001), due to the fact that the entire corpora were completely transcribed, relabeled, verified and proof-read. However, no differences were found to justify a reevaluation of the analyses made in Eklund (2001).

Table 5.21. Most commonly prolonged segments in all corpora. Actual occurrence is shown, as is total incidence of the phones in questions to normalize for general segment occurrence. For each prolonged segment, the percentage of prolonged realizations of that phone is given. Note that phone occurrence is approximate since transcription was phonological, not phonetic. The symbol # implies word border. The symbol – implies word continuation.

	WOZ-1	**WOZ-2**	**Nymans**	**Bionic**
Prolonged segment	**[–n#]**	**[–n#]**	**[–n#]**	**[–n#]**
No. of that segment prolonged	19	26	23	34
Total no. of that segment	8599	8086	2371	3398
% Segm. prol. of tot. no. that segment	0.22%	0.32%	0.97%	1.00%
Prolonged segment	**[#f–]**	**[#f–]**	**[#f–]**	**[#f–]**
No. of that segment prolonged	15	12	9	26
Total no. of that segment	2953	2170	670	1499
% Segm. prol. of tot. no. that segment	0.51%	0.55%	1.34%	1.73%
Prolonged segment	**[–l#]**	**[–l#]**	**[#s–]**	**[–l#]**
No. of that segment prolonged	15	12	9	19
Total no. of that segment	9514	9166	1521	3859
% Segm. prol. of tot. no. that segment	0.16%	0.13%	0.59%	0.49%
Prolonged segment	**[#s–]**	**[#s–]**	**[–a#]**	**[–a#]**
No. of that segment prolonged	13	11	9	16
Total no. of that segment	4542	4598	3366	4960
% Segm. prol. of tot. no. that segment	0.29%	0.24%	0.27%	0.32%
Prolonged segment	**[–t–]**	**[–t–]**	**[–t–]**	**[–t–]**
No. of that segment prolonged	11	10	7	11
Total no. of that segment	9998	10534	3016	4641
% Segm. prol. of tot. no. that segment	0.11%	0.09%	0.23%	0.24%

As is shown, "most commonly prolonged" has two answers, one at *token* and one at *type* level. For example, while [n] is by far the most commonly prolonged segment (token) in Nymans at a token level (n=23), [f] is the most commonly prolonged type (1.34% prolonged). The most striking result is of course that the same segments occupy the first two positions in both corpora, [–n#] and [#f–], respectively, which is partly—but not entirely—due to the frequently prolonged word *från* ("from").

5.5.7 Open vs. closed word classes

Eklund (2001)—on a subset of the present data—found no strong tendencies whether or not words pertaining to open or closed word classes were more subject to prolongation. The percentages of prolongations on words belonging to closed and open word classes are shown in **Table 5.22**.

Table 5.22. Percentages of prolongations (PRs) on words belonging to open and closed word classes.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. words	27664	26261	9250	12849	76024
No. PRs	156	132	129	197	614
No. / % open words	18382 (66.5%)	11342 (43.2%)	3218 (34.8%)	5625 (43.8%)	38567 (50.7%)
No. / % closed words	9262 (33.5%)	14919 (56.8%)	6032 (65.2%)	7224 (56.2%)	37457 (49.3%)
No. / % PRs on open	80 (51.3%)	72 (54.5%)	73 (56.6%)	117 (59.4%)	342 (55.7%)
No. / % PRs on closed	76 (48.7%)	60 (45.5%)	56 (43.4%)	80 (40.6%)	272 (44.3%)

As we can see, there is a slight tendency to prolong open words class words more often than closed word class words. This difference is also statistically significant, albeit only very weakly at $p = 0.487$ (Pearson chi-square). This significance also goes away if Bionic—the only corpus to reveal any major difference—is excluded, which results in $p = 0.661$ (Pearson chi-square). So, rather than elaborate on why open class words are more prone to prolongation, I prefer to say that there is no such difference worth dwelling on.

5.5.8 Phonological length

A final issue to be mentioned is phonological length, which is distinctive in Swedish, as well as mutually exclusive. This means that all VC syllables come either as V:C or VC: (or VCC). In recent work on dynamic segmental effects associated with focusing in Swedish, Heldner & Strangert (2001) showed that while focused segments were lengthened by an average 25%, short vowels are only marginally, and not distinctively, lengthened. This finding is paralleled in the present data set. While long vowels, and both long and short consonants are subject to prolongation, no instances of prolonged short vowels have been found.

5.5.9 A comparison with Tok Pisin[1]

As a private activity, I collected authentic Tok Pisin travel booking data while in Papua New Guinea in 1999/2000 (see Eklund, 2000b). These data have been used for comparison in a couple of previous studies that I will summarize here.

5.5.9.1 Introduction: Tok Pisin corpus

In order to test some of the observations made above, a comparative study was made on available Tok Pisin data. The Tok Pisin corpus (TP) consists of authentic ATIS dialogues, collected on location in Kavieng, Papua New Guinea, during the period December 1999 and January 2000 (Eklund 2000b). TP consists of 39 authentic human–human ATIS dialogues, and was labeled by the author (who is not a native speaker of Tok Pisin). Currently, a total number of 654 utterances and 3,538 words have been transcribed, with a total number of 35 prolongations.

[1] Since no new analyses of the Tok Pisin data have been carried out since the 2001 study, this section repeats and summarizes the figures already presented in Eklund (2001).

5.5.9.2 Duration

The mean duration for all prolongations was 0.347 (n = 35). The 95% confidence interval was 0.287/0.407. Standard deviation was 0.170. There was no significant difference between prolongation durations in Swedish and Tok Pisin (p = 0.055, t-test, two-tailed, equal variances assumed).

5.5.9.3 Prolongations vs. filled pauses

As was shown above for Swedish, filled pauses were significantly longer than prolongations. To check whether this holds true for Tok Pisin, the values for filled pauses in TP were explored. The mean for all filled pauses was 0.456 (n = 80). The 95% confidence interval was 0.401/0.501. Standard deviation was 0.244. Filled pauses were significantly longer than prolongations. A t-test resulted in p = 0.018 (two-tailed, equal variances assumed), and a Mann-Whitney test resulted in p = 0.008 (two-tailed).

5.5.9.4 Position within the word

The distribution of prolongations as a function of position in the word is shown in **Table 5.23**.

Table 5.23: Phone type and position of PRs.

		Tok Pisin
No.	PRs	35
No.	Segments	12,840
%	PRs / Segments	0.27%
N/%	Initial phone	6 (17.1%)
	% vowel	4 (66.8%)
	% cons +sonorant	1 (16.6%)
	% cons –sonorant	1 (16.6%)
N/%	Medial phone	—
N/%	Final phone	29 (82.9%)
	% vowel	12 (41.4%)
	% cons +sonorant	13 (44.8%)
	% cons –sonorant	4 (13.8%)

As is shown, the ratio in TP for initial/medial/final position is roughly 15–0–85, which differs from the distribution reported for Swedish and American English, mentioned above.

5.5.9.5 Top-five phones

The most commonly prolonged segments (normalized for overall segment frequency) are shown in **Table 5.24**.

Table 5.24. Most commonly prolonged segments in the Tok Pisin corpus. Actual occurrence is shown, as is total incidence of the phones in questions to normalize for general segment occurrence. For each prolonged segment, the percentage of prolonged realizations of that phone is given. Note that phone occurrence is approximate since transcription was phonological, not phonetic. The symbol # implies word border. The symbol – implies word continuation.

	Tok Pisin
Prolonged segment	**[–ŋ#]**
No. of that segment prolonged	5
Total no. of that segment	249
% Segm. prol. of tot. no. that segment	2.01%
Prolonged segment	**[–m#]**
No. of that segment prolonged	5
Total no. of that segment	778
% Segm. prol. of tot. no. that segment	0.64%
Prolonged segment	**[–s#]**
No. of that segment prolonged	4
Total no. of that segment	598
% Segm. prol. of tot. no. that segment	0.67%
Prolonged segment	**[i]**
No. of that segment prolonged	3
Total no. of that segment	1413
% Segm. prol. of tot. no. that segment	0.21%
Prolonged segment	**[–o(ŋ)#]**
No. of that segment prolonged	2
Total no. of that segment	880
% Segm. prol. of tot. no. that segment	0.23%

As is shown, the top-five list of prolonged segments in Tok Pisin differs from the corresponding list for Swedish. That other segments are prolonged more often in Swedish than in Tok Pisin is perhaps not surprising. What is more striking is that the segments seem to be prolonged for the same reason. The phones [ŋ] and [o] mainly occur in the prepositions *long* (general preposition), pronounced [loŋ] or [lo]) and *bilong* (stronger-binding preposition, genitive marker, conjunction), pronounced [biloŋ] or [blo].

5.5.9.6 Open vs. closed word classes

Rates of words belonging to open and closed word classes and PR rates are shown in **Table 5.25**.

Table 5.25. Ratio open/closed word classes and PR rates in TP.

	Tok Pisin
No. open / % total no. words	1,592 (45.0%)
No. closed word tokens / % total no. words	1,946 (55.0%)
No. closed unique words / % total no. closed words	39 (2.0%)
No. PRs open / % total no. PRs	6 (17.1%)
No. PRs closed / % total no. PRs	29 (82.9%)

Unlike Swedish, the tendency to prolong words belonging to closed word classes is more marked in TP than in the Swedish data. Out of 35 prolongations, 29 occur either in prepositions (*long*, *bilong*) or in grammatical markers such as *i* (predicate marker), *bai* (future marker) or *ol* (plural marker). Moreover, three of the six prolonged words belonging to open word classes are from the domain. *fe* ("fare"), *ples* ("place") and *tri* ("three"), and two instances of a prolonged transitive suffix *-im* in the words *salim* ("send") and *sekim* ("check"), and the two latter could arguably be analyzed as grammatical prolongations.

5.5.9.7 Swedish–Tok Pisin discussion

As was shown in Eklund (2000a), there are no great general differences with regard to disfluency in Swedish and Tok Pisin. The only marked differences between Swedish and Tok Pisin, as reported in Eklund (2000a, 2001), occur for prolongations. While there are no significant differences as to prolongation durations proper, there are differences with regard to segment type and distribution of word-position and segment type. While Swedish exhibited a 30–20–50 ratio for initial, medial and final word-position, respectively, Tok Pisin showed a 15–0–85 ratio. One *possible* explanation for these differences could be the underlying morphotactic constraints of the two languages. Whereas Swedish allows C^3VC^8 syllables, syllable structure in Tok Pisin allows only C^2VC^1 syllables, and even such initial clusters are often split in two by the insertion of epenthetic vowels.

The 'morphology matters' hypothesis possibly receives some support from Eklund & Shriberg (1998) who reported a 30–20–50 ratio for American English, while Den (2003), reports a 10–5–85 ratio reported for Japanese (Den, 2003) and Lee et al. (submitted for publication) who report a 1–4–95 ratio for Mandarin. Comparing these languages, Swedish and English, of course, are very similar as to phonotactics and morphology, while Japanese and Mandarin more resemble Tok Pisin with their more constrained morphology.

Finally, Tok Pisin top five list was different as compared to Swedish, but it seems as if the phones seemingly had the same "source", i.e. phones from words in corresponding loci in the utterances, which happened to affect different phones in the two languages.

5.5.10 Summary

The first thing to point out with regard to prolongations, given its historical baggage, is that not only are they common in the speech of nonstutterers (as observed in several languages), they are also among the most common types.

Second, which might appear as somewhat unintuitive, we can establish that *all segment types* might be prolonged, although there is a tendency towards prolonging continuants.

Looking at phonological length, it is striking to find that no short vowels are prolonged in our data. This observation supports the hypothesis that phonology puts constraints on the production of PRs, which receives further support from the observations reported by Heldner & Strangert (2001), who observed that while segments in stressed syllables in Swedish are longer in duration than the same segments in unstressed syllables, this effect did not occur for short vowels.

As regards duration, prolongations are significantly shorter than filled pauses, despite their physiological, acoustic and functional similarities.

From a morphological point of view, the favored position for segment prolongation is word-final, in both Swedish and Tok Pisin. However, the observation that the ratio initial–medial–final position differs between Swedish and Tok Pisin could suggest that prolongation production could be language-specific, a notion which receives further support from the figures reported for Japanese and Mandarin.

Stepping up to full words, there is a small tendency for words belonging to open word classes in Swedish, while closed words are preferred in Tok Pisin. I will, however waive any further discussion as to the underlying reasons for these observations since the difference is slight in Swedish, and the Tok Pisin data set is fairly small, which makes any speculations or conclusions tentative at best.

In conclusion, the prototypical Swedish PR would be the final segment—preferably a continuant—of a preposition or article, or appear in a domain-dependent word which signals crucial information with regard to the task at hand. The comparison with Tok Pisin suggests that these observations probably do not hold for all languages, and that more cross-linguistic studies of prolongations need be done in order to gain deeper insights with regard to the role and function of segment prolongation in human speech production.

5.6 Durational disfluencies: final comments

As we have seen in the previous sections on the three "durational" disfluencies unfilled pauses, filled pauses and prolongations, there are differences between them, not only regarding basic characteristics, but also concerning frequency and distribution.

Although differences as regards duration proper were studied between the pairs unfilled and filled pauses, and filled pauses and prolongations, it should not be taken as self-evident that mean values are the best way to compare the data, since we cannot assume a Gaussian distribution of the data (even if we assume equal variances).

To compare all three durational disfluency types, they are plotted against each other in the following figures. **Figure 5.2a** shows the numbers of unfilled pauses, filled pauses and prolongations in different duration intervals.

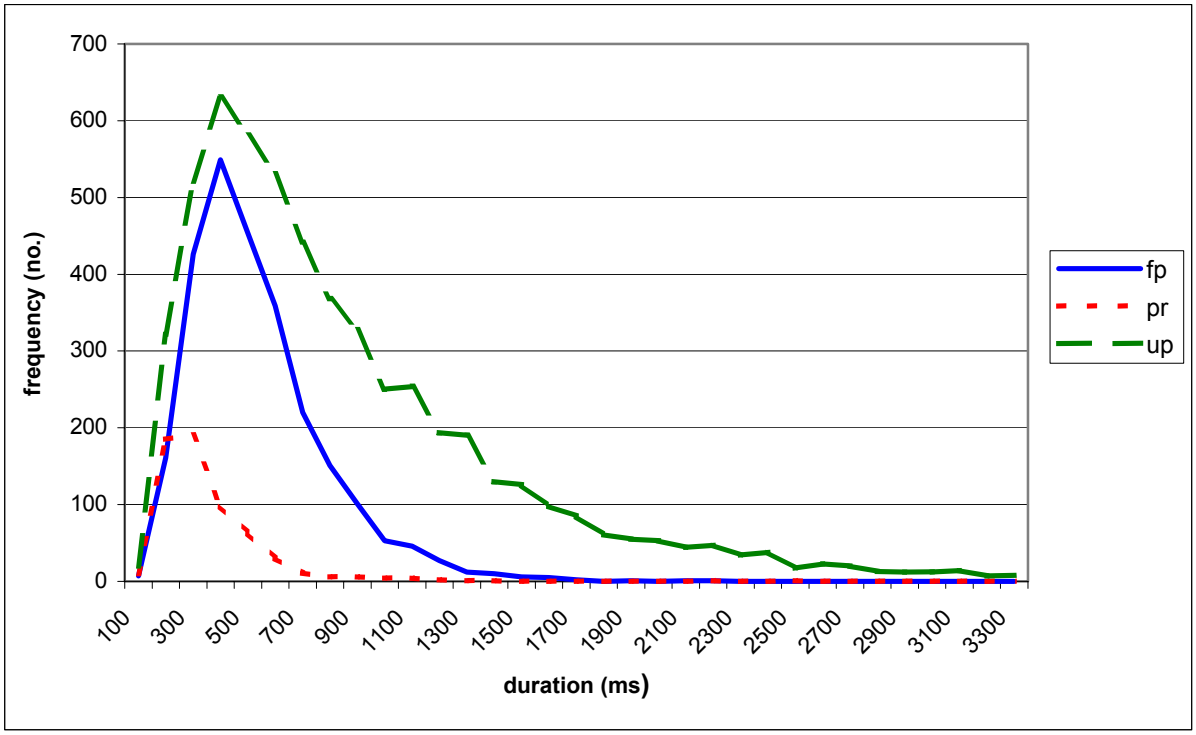


Figure 5.2a. Comparison of pooled numbers of unfilled pauses, filled pauses and prolongations in different duration intervals.

The first thing to note in **Figure 5.2a** is of course that the general distribution is not Gaussian (which was not to be expected). The second thing to observe is that unfilled pauses fall off slower than do filled pauses, which in turn fall off slower than do prolongations.

In **Figure 5.2b**, the cumulative frequency of unfilled pauses, filled pauses and prolongation is shown.

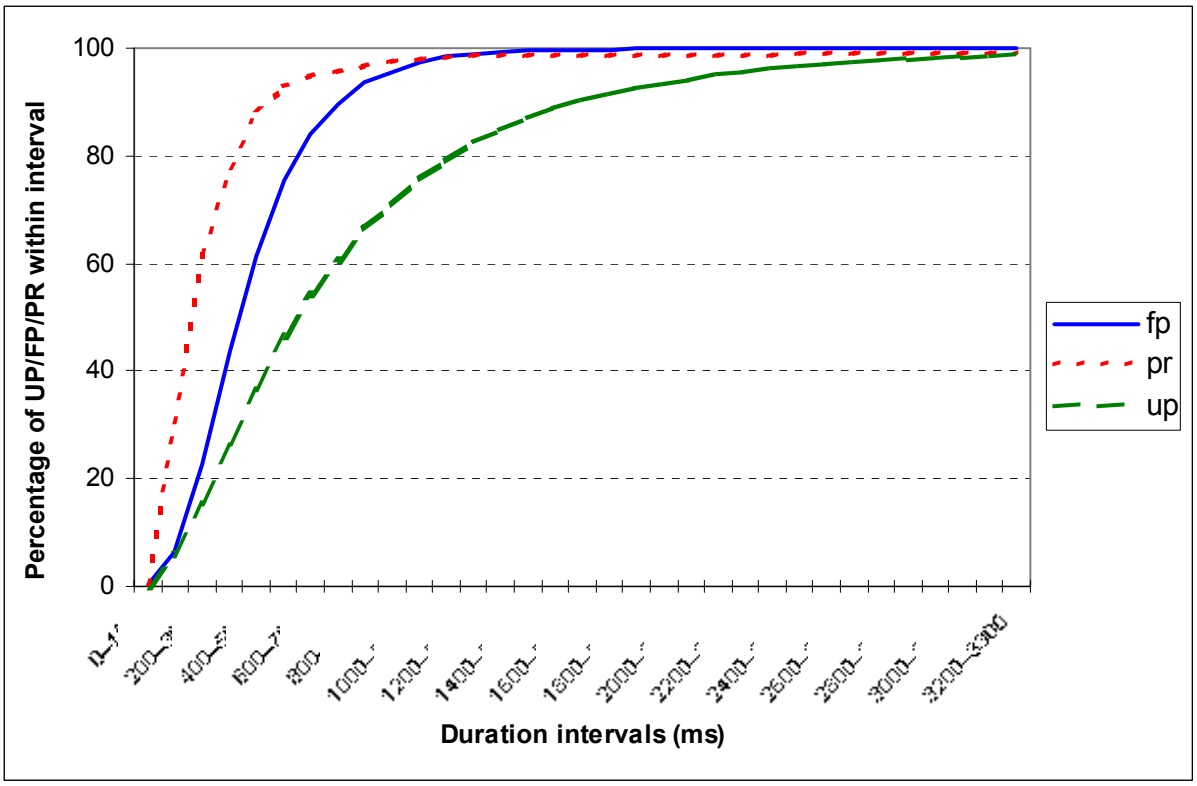


Figure 5.2b. Cumulative percentages of pooled numbers of unfilled pauses, filled pauses and prolongations within different duration intervals.

Once again, we see that while 80% of prolongations are shorter than 500 ms, we need to include unfilled pauses up to 1 second of duration to include 80% of all instances. This shows that not only are the frequency and mean durations of unfilled pauses, filled pauses and prolongations different, so is their proportional distribution, with a steeper curve for prolongations, than for filled pauses, which in turn fall off quicker than do unfilled pauses.

5.7 Explicit editing terms

Hindle (1983) based his model for disfluency detection on the occurrence of explicit editing terms, such as *oops* or *sorry*. Since then, it has been shown over and over again that such overt marking of speech errors is rare indeed.

5.7.1 General explicit editing rates

Occurrence of explicit editing terms in the corpora is shown in **Table 5.26**.

Table 5.26. General incidence of explicit editing terms (EETs) in the corpora, as well as percentages of utterances and words that include EETs.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. Explicit editing terms (EETs)	20	31	9	28	88
No. utts	4023	3438	1734	1985	11180
No. utts –1-word utts	3117	3087	940	1342	8466
No. words	27664	26261	9250	12849	76024
% EETs/utts	0.5%	0.9%	0.5%	1.4%	0.8%
% EETs/utts –1-word utts	0.6%	1.0%	0.9%	2.1%	1.0%
% EETs/words	0.07%	0.12%	0.09%	0.22%	0.12%

As is shown, explicit editing is extremely rare in all corpora, averaging about than 1% of utterances containing explicit editing terms, and about 0.1% editing terms per word. It has been pointed out in the literature that explicit editing terms do not provide substantial help in automatic detection of disfluency, as was suggested by Hindle (1983), and this is confirmed in the present data set.

5.7.2 Cross-corpus differences

Differences between the corpora are shown in **Table 5.27**.

Table 5.27. Cross-corpus differences for explicit editing terms. Statistical significance is given relative the total number of words (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 0.07% ; 27664	**WOZ-2** 0.12% ; 26261	**Nymans** 0.09% ; 9250	**Bionic** 0.22% ; 12849
WOZ-1	—	n.s.	n.s.	$p < 0.05$ (B)
WOZ-2	n.s.	—	n.s.	$p < 0.05$ (B)
Nymans	n.s.	n.s.	—	$p < 0.05$ (B)
Bionic	$p < 0.05$ (B)	$p < 0.05$ (B)	$p < 0.05$ (B)	—

As is shown, very few significant differences occur between the corpora with regard to explicit editing, and the only corpus that stands out is that Bionic is significantly more disfluent than all other corpora, while WOZ-1, WOZ-2 and Nymans do not differ from each other with regard to explicit editing.

First, it must be pointed out that EETs are so rare that one could easily question whether our data provides a sufficient basis for statistical comparison. Second, however, *if* the present data do indeed provide such a basis, then it is obvious that explicit editing is not too consciously applied, since then it should be more common in the human–human setting, where the interlocutor could actually interpret the "sorries" provided by the speakers.

5.7.3 Summary

Given the rareness of EETs, not much can be said, other than joining the post-Hindle literature, and once again point out that EETs are indeed rare, and that they do no seem to be subject to setting manipulation to any larger degree.

5.8 Mispronunciations

The category mispronunciation here is more or less equivalent to what has been referred to as *slips-of-the-tongue* in the literature, or is at least included in that category. One difference is that many slips just reorganize the position of words (which would constitute a repair here), mispronunciations in the present work always lead to *non*-words (also included in the definition of slips). As was noted earlier, slips/mispronunciations are so rare in spontaneous speech that much of the research devoted to them has been carried out on *elicited* slips.

5.8.1 General mispronunciation rates

The number of mispronunciations is given in **Table 5.28**.

Table 5.28. General incidence of mispronunciations (MPs) in the corpora, as well as percentages of utterances and words that include MPs.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. mispronunciations (MPs)	41	22	10	31	104
No. utts	4023	3438	1734	1985	11180
No. utts –1-word utts	3117	3087	940	1342	8466
No. words	27664	26261	9250	12849	76024
% MPs/utts	1.0%	0.6%	0.6%	1.6%	0.9%
% MPs/utts –1-word utts	1.3%	0.7%	1.1%	2.3%	1.2%
% MPs/words	0.15%	0.08%	0.1%	0.24%	0.14%

As is shown, mispronunciations are rare indeed. In fact, only one word in around 730 is mispronounced, and not even one percent of utterances include a mispronounced word. In a way, this could be taken as evidence for the notion that disfluency is not due to motor problems, at least not at a fine-grained, detailed, level. Most words come out with the phones in their intended and appropriate order.

5.8.2 Cross-corpus differences

Differences between the corpora are shown in **Table 5.29**.

Table 5.29. Cross-corpus differences for mispronunciations. Statistical significance is given relative the total number of words (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 0.15% ; 27664	**WOZ-2** 0.08% ; 26261	**Nymans** 0.1% ; 9250	**Bionic** 0.24% ; 12849
WOZ-1	—	$p < 0.05$ (W1)	n.s.	n.s.
WOZ-2	$p < 0.05$ (W1)	—	n.s.	$p < 0.05$ (B)
Nymans	n.s.	n.s.	—	$p < 0.05$ (B)
Bionic	n.s.	$p < 0.05$ (B)	$p < 0.05$ (B)	—

The sparse amount of data makes any conclusion tentative with regard to mispronunciations. If, however, the differences presented in **Table 5.29** do indeed represent differences in the real world, then one can at least point out that mispronunciations, presumably being a more "motoric" disfluency, still follow the general pattern of appearing in a higher number in a more disfluent corpus, in that Bionic exhibits significantly more mispronunciations than at least two other corpora.

Once again, it must be pointed out that this observation is based on very little data.

5.8.3 Repair or not?

Another issue is whether mispronunciations are attended to and consequently repaired, or whether they are left as they are, with no correction being provided by the speaker. Basically, this boils down to the question whether or not a mispronunciation is also part of a repair, or whether it should be considered as a "stand-alone" disfluency. A couple of illuminating examples are given below:[1]

Example 1: mispronunciation without a repair:

> *Jag vill boka flygbiljett **fnån** Stockholm to Helsingborg den femtonde augusti.*
> "I would like to book a flight **fnom** Stockholm to Helsingborg on the 15th of August."

Example 2: mispronunciation with a repair:

> *Okej, då skulle jag gärna vilja ha **sapan… sammanfattning** på min bokning.*
> "OK, then I'd like to have a **suppa… summary** of my booking."

[1] All examples are authentic, taken from WOZ-2, subject number 38.

In compounds, occasionally only the mispronounced part of the word is repaired, as in the following example:

Example 3: mispronunciation of a compound, with the mispronounced part repaired:

*/…/ vi ska vara där den trettonde i sjätte på **förmig… middagen** klockan tio.*
"/…/ we'll have to be there on the 13th of June on the **forenoog… noon** at ten o'clock."[1]

Here only the second part—middag ("noon")—is repaired, since the first part was OK. This pattern occurs several times in the corpora, and has been included in the category "repaired mispronunciation".

Once again, it should be pointed out that it is not always clear whether a mispronounced word is in fact repaired, since sometimes it is not possible to know for sure what the target word for the mispronounced item actually was, making it impossible to know whether the following word is a repaired version of the *intended* word, or just *another* word. This being said, the figures for repaired mispronunciations are given in **Table 5.30**.

Table 5.30. Numbers and percentages of repaired mispronunciations in the corpora.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. mispronunciations (MPs)	41	22	10	31	104
No. repaired MPs	20	12	5	8	45
% repaired MPs	48.8%	54.5%	50.0%	25.8%	43.3%

About half of the mispronunciations are corrected by the speakers, indicating that the subjects are in fact aware of their mistakes, an issue of interest since whether or not speakers do hear their own mistakes has been discussed in the literature. That Bionic exhibits a lower percentage of corrected mispronunciations is hard to explain, but it must be borne in mind that the data set, for all corpora, is very small, so any observations must be considered with some caution, lest too far-reaching conclusions be drawn.

5.8.4 Summary

Not much can be said about mispronunciations, given how rare they are in spontaneous speech. Perhaps the only interesting observation is that not even half of the mispronunciations are repaired, but left as they are. Whether this shows that the remaining mispronunciations are not noticed (heard or "felt" by the speech motor system) by the speaker or not—thus indicating attentive self-monitoring—or whether they are simply not noticed by the speaker, is not revealed in our present data set, and must be left unanswered in this study.

[1] Sorry about the low-frequency word "forenoon", but *förmiddag* is as common as *eftermiddag* ("afternoon") in Swedish, so I opted for its counterpart in English for pedagogical reasons.

5.9 Truncations

Truncations were, like prolongations, part of the category dysrhythmic phonations. Although these two could be signals of the same speech production phenomena in monologue, or arguably belong to the same category in stuttering, in *dialogue* many truncations are caused by interruptions by the interlocutor(s), something that rarely, if ever, leads to segment prolongation. Since only one side of the dialogues was analyzed here, truncations put the labeler on a slippery slope when deciding which truncations are genuine speech disfluencies, and which are external interruptions, caused by the other party of the dialogue. A first solution to this problem was to exclude all utterance-final interruptions, since these beyond any doubt, in most cases, were caused by the agent.

5.9.1 General truncation rates

Occurrence of truncations in the corpora is shown in **Table 5.31**.

Table 5.31. General incidence of truncations (TRs) in the corpora, as well as percentages of utterances and words that include TRs.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. truncations (TRs)	167	134	141	146	588
No. utts	4023	3438	1734	1985	11180
No. utts –1-word utts	3117	3087	940	1342	8466
No. words	27664	26261	9250	12849	76024
% TRs/utts	4.1%	3.9%	8.1%	7.3%	5.3%
% TRs/utts –1-word utts	5.4%	4.3%	15.0%	10.9%	6.9%
% TRs/words	0.6%	0.5%	1.5%	1.1%	0.8%

As is shown, when all data are pooled, around 5% of the utterances contain truncations. That Nymans contains more truncations than the other corpora is surely due to interruptions from the (human) agent, an effect which is still there, despite the fact that all utterance-final truncations were omitted from labeling and analysis. Again, this shows that disfluency in monologue is not equal to disfluency in dialogue.

The fact that the rates are higher in Bionic than in WOZ-1 and WOZ-2 might be attributable to the previously mentioned fact that no constraints were put on utterance length in Bionic—as opposed to WOZ-1 and WOZ-2—where the wizards were instructed to not accept too long utterances. Thus, no matter how long the utterances were in Bionic, the system accepted the input, which led to longer utterances, which in turn led to more disfluency in the form of self-truncation. This stresses the importance of keeping the instructions the same across different data collections.

5.9.2 Cross-corpus differences

Occurrence of truncations in the corpora is shown in **Table 5.32**.

Table 5.32. Cross-corpus differences for truncations. Statistical significance is given relative the total number of words (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 0.6% ; 27664	**WOZ-2** 0.5% ; 26261	**Nymans** 1.5% ; 9250	**Bionic** 1.1% ; 12849
WOZ-1	—	n.s.	$p <= 0.05$ (N)	$p <= 0.05$ (B)
WOZ-2	n.s.	—	$p <= 0.05$ (N)	$p <= 0.05$ (B)
Nymans	$p <= 0.05$ (N)	$p <= 0.05$ (N)	—	$p <= 0.05$ (N)
Bionic	$p <= 0.05$ (B)	$p <= 0.05$ (B)	$p <= 0.05$ (N)	—

The obvious observation in **Table 5.32** is that Nymans is significantly more disfluent than all other corpora with regard to truncations, and it should be remembered that this is still the case although all obvious utterance-final truncations were not included in the truncation count.

The main conclusion to be drawn from this observation is, of course, that when interrupted by an interlocutor, speakers do not feel obliged to finish an entire word before halting their speech, but instead voluntarily stop short mid-word, which means that the integrity of the word is not absolute, but instead not too strong.

5.9.3 Summary

Summing up our observation of truncations, the main conclusion to be drawn from this study is that truncations increase when a speaker is interrupted, indicating that speakers do not feel forced to finish the word they have begun speaking. An entailing conclusion to be drawn from this is that speech production must work with smaller chunks than full words before being forwarded to the motoric system for emission.

5.10 Repairs

Repairs constitute a different kind of disfluencies than all the others in that they work at larger levels than word levels, and that they might include all other kinds of disfluencies. Although they may be analyzed from a wide variety of perspectives, I will begin by presenting the general frequencies, and the differences between the corpora.

5.10.1 General repair rates

General repair frequencies are shown in **Table 5.33**.

Table 5.33. General incidence of repairs (REPs) in the corpora, as well as percentages of utterances and words that include REPs.

	WOZ-1	WOZ-2	Nymans	Bionic	Pooled
No. repairs (REPs)	215	257	172	202	846
No. utts	4023	3438	1734	1985	11180
No. utts –1-word utts	3117	3087	940	1342	8466
No. words	27664	26261	9250	12849	76024
% REPs/utts	5.3%	7.5%	9.9%	10.2%	7.6%
% REPs/utts –1-word utts	6.9%	8.3%	18.3%	15.0%	9.9%
% REPs/words	0.8%	1.0%	1.9%	1.6%	1.1%

Concerning repairs, the corpora seem to be divided into two groups, with WOZ-1 and WOZ-2 showing very low figures, and Nymans and Bionic showing higher figures.

5.10.2 Cross-corpus differences

Differences between the corpora with regard to repairs are shown in **Table 5.34**.

Table 5.34. Cross-corpus differences for repairs. Statistical significance is given relative the total number of words (test-of-proportion, two-tailed, 0.05 level). If significant, the more disfluent corpus is indicated (W1=WOZ-1; W2=WOZ-2; N=Nymans; B=Bionic).

	WOZ-1 0.8% ; 27664	**WOZ-2** 1.0% ; 26261	**Nymans** 1.9% ; 9250	**Bionic** 1.6% ; 12849
WOZ-1	—	p < =0.05 (W2)	p < =0.05 (N)	p < =0.05 (B)
WOZ-2	p < =0.05 (W2)	—	p < =0.05 (N)	p < =0.05 (B)
Nymans	p < =0.05 (-N)	p < =0.05 (N)	—	n.s.
Bionic	p < =0.05 (B)	p < =0.05 (B)	n.s.	—

As is seen, repairs occur more frequently in Nymans than in the other corpora, although the difference vis-à-vis Bionic is not significant (at a .05 level). A cautious conclusion here would be that speakers feel more motivated to repair their speech when speaking to humans than with speaking with computers, although the latter presumably would benefit more from repairs, given their limited capacity of understanding what was intended, despite overt errors in the speech string. This would mean that repairs are more of a "social tool" than a logical tool that is used to enhance communication from an understanding point of view proper. This, however, must remain a mere conjecture for the time being, pending a more detailed analysis of the entire discourses in all corpora.

5.10.3 General patterns

As was previously mentioned, unlike the other disfluency types included in this analysis, repairs are "structured", in that they both might include several items, but also that they can include all other types of disfluencies.

The basic notion is that a repair is two-fold, and that it contains something that went wrong, and consequently (at least occasionally) calls for a repair. The structure thus becomes:

… [*something-wrong-to-be repaired* | *the-repaired-substitution*] …

The left-hand part is often called *Reparandum* in the literature (e.g. Shriberg, 1994), while the right-hand side is referred to as *Reparans*. Both have structures in that the Reparandum can include items that are deleted in the Reparans, while the Reparans might simply repeat words already said in the Reparandum, insert new words not uttered in the Reparandum (for greater specificity, for example), or substitute a word in the Reparandum with a new word in the Reparans (e.g. a mispronounced word with a correctly pronounced version).

As was pointed out earlier, about 50% of mispronounced words were repaired, like in the following (authentic) example:

Vilken flyt flygplats landar planet på?
"What arr airport does the plane land on?"[1]

A word might also be truncated in the Reparandum, and substituted with another word in the reparans, as in the following example:

Skulle vilja by/ boka en eh flygresa mellan Chicago och Stockholm.
"Would like to cha[nge] book a flight between Chicago and Stockholm."

Here the word *byta* ("change") is stopped halfway through, and substituted with the intended word *boka* ("book").

As you can see, the general point here is that repairs can include filled pauses, mispronunciations, truncations and so on, while at the same time include repetitions of perfectly fluent words, with or without substitutions and so on.

5.10.3.1 What's in a repair?

The basic research question here is whether there are any detectable patterns in the data. However, preliminary analysis reveals no such trends, and pending a more detailed study, I can only conclude—for the time being—that it would seem as if repairs might consist of just about anything.

How repairs should be regarded depends on the research angle, but it goes without saying that they are of interest, primarily to developers of speech production models.

[1] The translations here are not supposed to be good English, but rather reflect the Swedish original.

5.10.3.2 Covert repairs, or ∅ reparandum / reparans?

Just to follow up on the speech production theme mentioned in the previous paragraph, I would just like to briefly mention the notion of "covert" repairs, as touched upon in chapter two. If indeed covert errors exist, then there should be utterances that are being repaired before anything erroneous has been uttered.

This would lead to structured repairs with a **null reparandum**, thus:

I would like a ticket to [Ø + Göteborg]

… where the subject first intended, and was on the verge of, saying e.g. *Malmö*. It could be argued that there is evidence in favor of such instances. For example, there are cases where one could just make out the beginning of the articulation of an /m/ after the preposition, or in the final stages of the articulation of that preposition, but where the /m/ in question is never begun as a phone "in its own right", as it were. The speaker in this case must have detected the imminent error prior to phonation and stopped it before being executed. This clearly speaks in favor of inner loop monitoring, like e.g. efference-copy monitoring

Although such evidence relies on much more vague argumentation than does evidence based on observations of explicit data, this is necessarily so. Covert repairs are by definition *covert*, and the fact that one can hear anticipatory coarticulation of a planned phone is as close as one can get to actual data.

Moreover, one could perhaps hypothesize that there are cases with null reparans, where something erroneous is not being repaired, e.g.

I would like a ticket to [Göteborg + Ø]

… where *Göteborg* is wrong, but is not repaired, for some reason. Whether or not this latter case involves any covert part or not is of course difficult to assess (except asking the speaker?), but an automatic system working according to an expected

[Reparandum | Reparans]

… structure would have to be able to handle both the aforementioned cases.

In conclusion, while there is no need to argue that there are overt repairs in the utterances produced by humans, sometimes even marked with explicit editing such as *Oops, that was wrong!*, covert editing is harder to prove, as it were. It is, however, my firm conviction that covert editing is in fact evidenced by such phenomena as anticipatory coarticulation, and that such examples occur in the present data set. However, given the limited scope of this study, further study of covert editing will have to wait for the time being.

5.10.4 Back-tracking (a.k.a. retracing)

During repairs, the speaker has the option of back-tracking to an earlier point in the utterance, and restart from where it went wrong, repeating some of the words that were in fact correct. i.e. words before the interruption point that are repeated in the reparans. Eklund & Shriberg (1998) showed that this back-tracking exhibited similar characteristics in American English

and Swedish. There is a specific typology involved here, since words can be repeated with different degrees of modification or amendment.

The simplest case is when back-tracking involves only one word:

Då då *vill jag att ni bokar mig på den*
"Then then please book that [flight] for me."

However, backtracking might also include more words, as in the following (authentic) three-word example:

Eh nej ***det är för det är för*** *tidigt*
"Uh no it is too it is too early."

Back-tracking might or might not include other words, inserted, substituted or repaired, but what about verbatim back-tracking, as in the two examples above?

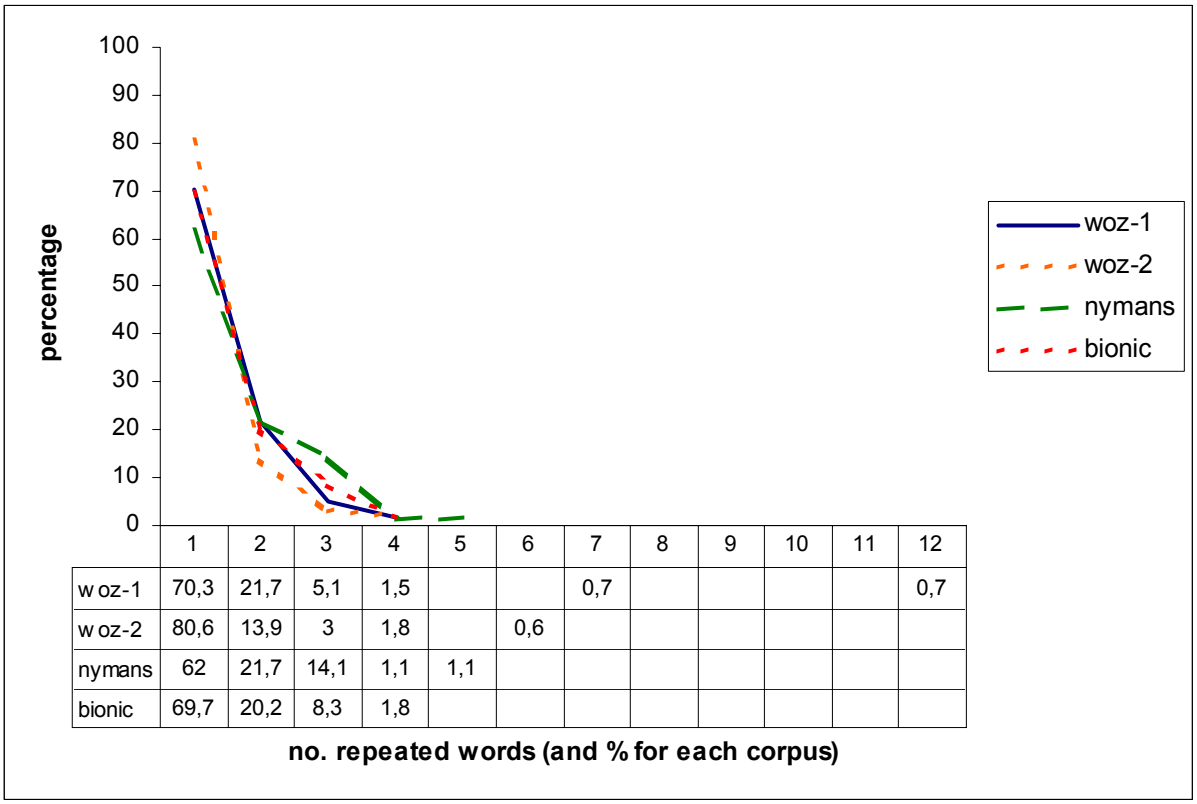
5.10.4.1 Verbatim back-tracking

So, when a speaker stops dead, then retraces and starts all over again, how many words does he or she back up in the utterance? Are there any limits, or does the frequency follow a specific curve as to the distribution of frequencies at different numbers of words? The incidence of verbatim retraced words is shown in **Table 5.35**.

Table 5.35. Incidence of verbatim retraced words in the corpora. For all corpora, the number of retraces for different retracing lengths is shown. Relative percentages for each retrace length within each corpus are also given within parentheses.

No. of verbatim retraced words	WOZ-1	WOZ-2	Nymans	Bionic
One word	97 (70.3%)	133 (80.6%)	57 (62.0%)	76 (69.7%)
Two words	30 (21.7%)	23 (13.9%)	20 (21.7%)	22 (20.2%)
Three words	7 (5.1%)	5 (3.0%)	13 (14.1%)	9 (8.3%)
Four words	2 (1.5%)	3 (1.8%)	1 (1.1%)	2 (1.8%)
Five words	—	—	1 (1.1%)	—
Six words	—	1 (0.6%)	—	—
Seven words	1 (0.7%)	—	—	—
Eight words	—	—	—	—
Nine words	—	—	—	—
Ten words	—	—	—	—
Eleven words	—	—	—	—
Twelve words	1 (0.7%)	—	—	—
Σ	138	165	92	109

As is shown, verbatim repeats of up to four words in length exist in all corpora, with a couple of odd examples of longer retraces. The question is whether retrace lengths are comparable between the different corpora. Eklund & Shriberg (1998) showed that Swedish and American English exhibited similar proportions as to retrace lengths, so the question could be asked whether different Swedish corpora are similar in that respect. The proportions of retrace lengths for all corpora are shown in **Figure 5.3**.



	1	2	3	4	5	6	7	8	9	10	11	12
woz-1	70,3	21,7	5,1	1,5			0,7					0,7
woz-2	80,6	13,9	3	1,8		0,6						
nymans	62	21,7	14,1	1,1	1,1							
bionic	69,7	20,2	8,3	1,8								

Figure 5.3. Retrace length percentages. For each corpus, the relative percentage for each verbatim word retrace length is given. The data table specifies the exact percentage for each retrace length for each corpus, i.e. and e.g. for WOZ-1, 70.3% of all verbatim retraces are one-word repetitions, 21.7% are two-word repetitions and so on.[1]

As is shown, the relative proportions of retrace lengths in the different corpora are very similar (which was also expected given e.g. Eklund & Shriberg, 1998). However, there is a larger number of verbatim three-word retraces in Nymans than in the other corpora. This difference is significant as compared to WOZ-1 and WOZ-2 ($p < 0.05$, test-of-proportion, two-tailed), but not as compared to Bionic. If taken as a real differences, it could be an effect of human–human interaction, although the reason why this should be escapes explanation for the moment and which will have to be an hypothesis awaiting further elaboration, where a first, necessary step would be a full analysis of the dialogues in the corpus, i.e. including the agents' utterances. Also, the data set is not huge, and the fact that the difference is not significant compared to Bionic makes me leave this without further speculation.

The observation that twelve words are repeated verbatim in WOZ-1 is almost to regard as a freak of nature, and could arguably be an exception to Miller's famous claim that the number

[1] That commas are used as delimiters instead of full stops/periods is due to software idiosyncrasies.

seven, "plus or minus two", is the upper limit for our capacity for processing information (Miller, 1956).[1]

5.10.5 Summary

As was pointed out before, repairs differ from the other types of disfluencies, and in a way they merit—and require—a study of their own. The preliminary study undertaken here does not reveal any strong characteristics as regards what a repair might look like, but this does not mean that such trend could exist, and could show up given a more careful study. Suffice it to say that repair require prior understanding of the other disfluency types, since these are "the stuff that repairs are made of", plus the additional complications associated with the longer stretches of speech involved.

5.11 Gender differences

Gender has often been studied in regard to disfluency, but sources vary as to whether or not they find any significant differences between the sexes as to disfluency behavior. That men are more disfluent than women (in one way or another) has been shown by Feldstein, Brenner & Jaffee (1963), Edelsky (1981), Lickley (1994), Shriberg (1994), Bortfeld et al. (1999), Branigan, Lickley & MacKelvie (1999) and others, while e.g. Christenfeld (1995) and Bell, Eklund & Gustafson (2000) found no gender differences in their studies.

Gender differences are shown in **Table 5.36a** through **Table 5.36f**.

Table 5.36a. Gender differences in WOZ-1. Frequencies and percentages (relative to number of words) are given both broken down for type, and for all data pooled. Sum totals are given including and excluding unfilled pauses (UPs). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the gender of the more disfluent group is indicated (M=male; F=female).

	Male ***n*** **= 25**	**Female** ***n*** **= 21**	**Statistical Significance .05 level**
No. words	14597	13067	—
UP	874 (6.0%)	748 (5.7%)	n.s.
FP	441 (3.0%)	374 (2.8%)	n.s.
PR	93 (0.6%)	63 (0.5%)	n.s.
EET	9 (0.06%)	11 (0.08%)	n.s.
MP	16 (0.1%)	25 (0.2%)	n.s.
TR	70 (0.5%)	97 (0.7%)	$p < 0.05$ (F)
REP	98 (0.7%)	117 (0.9%)	$p < 0.05$ (F)
Σ (all types)	1601 (10.97%)	1435 (10.98%)	n.s.
Σ – unfilled pauses	727 (5.0%)	687 (5.3%)	n.s.

[1] The utterance in question is: *Jag undrar om det går ett flyg från Boston den tionde maj ... jag undrar om det går ett flyg till Boston den tionde maj* ("I would like to know whether there is a flight from Boston on May 10 … I would like to know whether there is a flight from Boston on May 10").

As we can see in **Table 5.36a**, there are only two significant differences, in that women evince more truncations and repairs. In all other respects the (slight) differences are not significant. Since the number of subjects is sufficiently large (M=25/F=21), one would surmise that there are no important gender differences as to disfluency production in our data. But things are not that clear-cut, as the **Table 5.36b** will tell us.

Table 5.36b. Gender differences in WOZ-2. Frequencies and percentages (relative to number of words) are given both broken down for type, and for all data pooled. Sum totals are given including and excluding unfilled pauses (UPs). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the gender of the more disfluent group is indicated (M=male; F=female).

	Male ***n*** **= 32**	**Female** ***n*** **= 14**	**Statistical Significance .05 level**
No. words	17384	8877	—
UP	1513 (8.7%)	666 (7.5%)	$p < 0.05$ (M)
FP	772 (4.4%)	268 (3.0%)	$p < 0.05$ (M)
PR	97 (0.6%)	35 (0.4%)	n.s.*
EET	25 (0.1%)	6 (0.07%)	$p < 0.05$ (M)
MP	14 (0.08%)	8 (0.09%)	n.s.
TR	102 (0.6%)	32 (0.4%)	$p < 0.05$ (M)
REP	187 (1.1%)	70 (0.8%)	$p < 0.05$ (M)
Σ (all types)	2710 (15.6%)	1085 (12.2%)	$p < 0.05$ (M)
Σ – unfilled pauses	1197 (6.9%)	419 (4.7%)	$p < 0.05$ (M)

* This is such a close shave that I might as well have put significant here.

In WOZ-2, as shown in **Table 5.36b**, men produce significantly more disfluency in all but two categories—and regarding EETs, it is so close to significant at .05 that one might as well include that, too. Like WOZ-1, this corpus includes a sufficient number of both subjects and words so as to be taken as a serious basis for far-reaching conclusions whether men are more, or less, disfluent than women in our data (or perhaps even Swedish?). This discrepancy is further "elaborated" when we turn to Nymans, in **Table 5.36.c**.

Table 5.36c. Gender differences in Nymans. Frequencies and percentages (relative to number of words) are given both broken down for type, and for all data pooled. Sum totals are given including and excluding unfilled pauses (UPs). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the gender of the more disfluent group is indicated (M=male; F=female).

	Male ***n*** **= 6**	**Female** ***n*** **= 2**	**Statistical Significance .05 level**
No. words	6635	2615	—
UP	413 (6.2%)	149 (5.7%)	n.s.
FP	135 (2.0%)	68 (2.6%)	n.s.
PR	97 (1.5%)	32 (1.2%)	n.s.
EET	5 (0.07%)	4 (0.1%)	n.s.
MP	9 (0.1%)	1 (0.04%)	n.s.
TR	117 (1.8%)	24 (0.09%)	$p < 0.05$ (M)
REP	125 (1.9%)	47 (1.8%)	n.s.
Σ (all types)	901 (13.6%)	325 (12.4%)	n.s.
Σ – unfilled pauses	488 (7.3%)	176 (6.7%)	n.s.

In Nymans—the human–human corpus—there is only one statistical difference in that men produce more truncations than women do. As has been pointed out before, many of the truncations in Nymans are probably due to agent interruption, but there is of course a possibility that truncations also are self-interruptions, rather than other-interruptions. In any case, this would seem to lead to the conclusion that if indeed there is a difference between men and women in that men evince more TRs, then either men allow themselves to be interrupted more often than women do, or men are interrupted more often than women are interrupted, or interrupt themselves more often than women do. So, the conclusion would then seem be that if men interrupt speakers more often than women do, then these speakers seem to be themselves.

There is, however, one other factor that needs to be considered: the gender of the agent. The data from Nymans, broken down for agent, are shown in **Table 5.36d**.

Table 5.36d. Agent gender in Nymans. Frequencies and percentages (relative to number of words) are given both broken down for type, and for all data pooled. Sum totals are given including and excluding unfilled pauses (UPs). For both the male and female agent, three subjects were male, and one was female. Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the gender of the agent is indicated (MA=male agent; FA=female agent).

All subjects pooled WOZ-2 / Nymans	**Male agent Subjects 1M/2M/3M/4F**	**Female agent Subjects 5M/6F/7M/8M**	**Statistical Significance .05 level**
No. words	4934	4316	—
UP	291 (5.9%)	271 (6.3%)	n.s.
FP	148 (3.0%)	55 (1.3%)	$p < 0.05$ (MA)
PR	103 (2.1%)	26 (0.6%)	$p < 0.05$ (MA)
EET	2 (0.04%)	7 (0.2%)	n.s.
MP	5 (0.10%)	5 (0.12%)	n.s.
TR	92 (1.9%)	49 (1.1%)	$p < 0.05$ (MA)
REP	102 (2.1%)	70 (1.6%)	n.s.
Σ (all types)	743 (15.1%)	483 (11.2%)	$p < 0.05$ (MA)
Σ – unfilled pauses	452 (9.1%)	212 (4.9%)	$p < 0.05$ (MA)

Here another picture emerges. For three of our disfluency categories (FPs, PRs and TRs), the subjects are significantly more disfluent when interacting with the male agent, and more disfluent in yet another category (REPs) if not significantly so. For two categories (EETs and MPs), subjects are more disfluent when interacting with the female agent, but not significantly. This would lead one to assume that whatever the gender of the subject, that person will be more disfluent when interacting with a man. This, however, is of course much too hasty a conclusion to draw from the present data set. First, while both agents were the most experienced at the travel agency, they also had different personalities that surely played a role, irrespective of the obvious caution one should take before making one person per gender represent woman- and manhood, respectively. Having interacted with both travel agents myself (with real-life tasks), I would feel more inclined to ascribed any potential agent-effect to individual traits, rather than gender traits. This, however, needs not really be said, since the agent data set is far too small to allow any definite conclusions. Suffice it to say here that the behavior or characteristics of the agent also seem to play a role, and that differentiation *within* the human group—as compared to (alleged) machines—might have

palpable effects on the communicative behavior of the subjects, similar to such possible effects as a function of whether or not the agent is a machine or not.

So, let us turn to our last corpus, the Bionic one, which is shown in **Table 5.36e**.

Table 5.36e. Gender differences in Bionic. Frequencies and percentages (relative to number of words) are given both broken down for type, and for all data pooled. Sum totals are given including and excluding unfilled pauses (UPs). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the gender of the more disfluent group is indicated (M=male; F=female).

	Male ***n*** **= 9**	**Female** ***n*** **= 7**	**Statistical Significance .05 level**
No. words	5876	6973	—
UP	560 (9.5%)	666 (9.5%)	n.s.
FP	225 (3.8%)	318 (4.6%)	n.s.
PR	89 (1.5%)	108 (1.5%)	n.s.
EET	21 (0.3%)	7 (0.1%)	$p < 0.05$ (M)
MP	18 (0.3%)	13 (0.2%)	n.s.
TR	63 (1.1%)	83 (1.2%)	n.s.
REP	89 (1.5%)	113 (1.6%)	n.s.
Σ (all types)	1065 (18.1%)	1308 (18.8%)	n.s.
Σ – unfilled pauses	505 (8.6%)	642 (9.2%)	n.s.

Once again, there is only one significant difference, in that men this time explicitly signal their errors, and produce more EETs. However, the data set here is relatively small, and one should not make too much out of it, perhaps. So, basically, once again we face a tie between the two genders when it comes to disfluency production.

What, then, happens if we merge our four corpora? In three of the corpora, only small differences were found, where women produced significantly more TRs and REPs in WOZ-1, and men produced significantly more TRs in Nymans and significantly more EETs in Bionic. In the remaining corpus, WOZ-2, men were more disfluent in all but two (one) category.

The pooled data are shown on **Table 5.36f**.

Table 5.36f. Gender differences for all corpora merged. Frequencies and percentages (relative to number of words) are given both broken down for type, and for all data pooled. Sum totals are given including and excluding unfilled pauses (UPs). Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level). If significant, the gender of the more disfluent group is indicated (M=male; F=female).

	Male ***n*** **= 72**	**Female** ***n*** **= 44**	**Statistical Significance .05 level**
No. words	44492	31532	—
UP	3360 (7.5%)	2229 (7.1%)	$p < 0.05$ (M)
FP	1573 (3.5%)	1028 (3.3%)	$p < 0.05$ (M)
PR	376 (0.8%)	238 (0.7%)	n.s.
EET	60 (0.1%)	28 (0.09%)	n.s.
MP	57 (0.13%)	47 (0.15%)	n.s.
TR	352 (0.8%)	236 (0.7%)	n.s.
REP	499 (1.1%)	347 (1.1%)	n.s.
Σ (all types)	6277 (14.1%)	4153 (13.2%)	$p < 0.05$ (M)
Σ – unfilled pauses	2917 (6.6%)	1924 (6.1)	$p < 0.05$ (M)

As can be seen, the WOZ-2 results carry over to the merged data, given that the frequency distribution was so even in the other three corpora. Men are more disfluent with regard to UPs and FPs, which shows up in both total counts (including and excluding UPs).

So, basically, there are no overall differences between the corpora, except that in WOZ-2, men are significantly more disfluent in five (almost six) of the seven categories, which transfers over and shows up in the merged data. This could be simply summarized in pointing out that men are more disfluent than women with regard to the two most frequent categories UPs and FPs, which also make them overall more disfluent than women, since there are no differences with regard to any other type (or the results go both ways). This, however, is not a satisfactory way out, in my view. In three of the corpora, with three different "interlocutors" (a "machine", a human, and a machine), there are no differences, and in one corpus, with the exact same tasks as in two of the other corpora, there are significant differences in almost all categories: So, one wonders, *why this thusness?* What is different in WOZ-2, as compared to Nymans and Bionic, with identical task sheets, and what is different between WOZ-2 and WOZ-1, where the much larger number of subjects is roughly the same?

We have already established that the task sheets were exactly the same as in Nymans and Bionic, so the tasks per se could not provide an explanation. Of course, in Nymans, the subjects were talking with real human beings, and in Bionic the "recognizer" (which was staged) accepted any input, however long it was, whereas in WOZ-2, any longer utterance would not be understood. So subjects could utter longer stretches of speech in Bionic than in WOZ-2, which should simply lead to more disfluency, not to gender differences? In WOZ-1, like in WOZ-2, longer utterances were not understood, but the task sheets were different.

So, one explanation *could* be that when not interrupted and allowed to speak out very long utterances, men get more disfluent than woman, so there is a kind of utterance-length/gender effect at work here. Another, more trivial, explanation is that the particular set of subjects in

WOZ-2 was not representative of the entire population, or indeed that the men in WOZ-1, Nymans or Bionic were not representative of the population.

In any case, no obvious solution to these observations is readily available, and perhaps a more detailed study of system/agent–subject interaction could provide the (obvious?) answer to this conundrum. As for now, we would have to settle with the somewhat unsatisfactory conclusion that our results are equivocal, and could point to greater disfluency production in men (with regard to UPs and FPs), but also that that need not necessarily be the case.

5.12 Cross-corpus observations

As was mentioned, eight subjects participated in two corpora, viz. WOZ-2 and Nymans. They carried the same exact tasks, and a sufficient amount of time had passed between the two data collection so as to rule out any major likelihood of learning effects. Since the subjects and the tasks were exactly the same, it is of interest to see whether there are any differences between the two corpora, since such differences most likely should be due to the interaction—or interlocutor—proper, given that all other variables are kept constant.

The results are shown in **Table 5.37a** through **Table 5.37c**.

Table 5.37a. Number of words and disfluencies for all subjects who participated in both WOZ-2 and Nymans, broken down for corpus and disfluency type. Actual numbers are given, as well as percentages for each disfluency type relative to the number of words in that corpus. Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level).

Subject number WOZ-2 / Nymans	Disfluency type	WOZ-2	Nymans	Significance (0.05 level)
10 / 1	No. words	505	1356	—
	UP	27 (5.3%)	65 (4.8%)	n.s.
	FP	24 (4.7%)	37 (2.7%)	n.s.
	PR	2 (0.4%)	18 (1.3%)	$p < 0.05$
	EET	0	0	—
	MP	0	3 (0.2%)	—
	TR	5 (0.9%)	39 (2.9%)	$p < 0.05$
	REP	9 (1.8%)	41 (3.0%)	n.s.
	Σ	67 (13.3%)	203 (15.0%)	$p < 0.05$
9 / 2	No. words	531	812	—
	UP	45 (8.5%)	64 (7.9%)	n.s.
	FP	13 (2.4%)	57 (7.0%)	$p < 0.05$
	PR	1 (0.2%)	46 (5.7%)	$p < 0.05$
	EET	3 (0.6%)	0	—
	MP	0	0	—
	TR	5 (0.9%)	1 (0.1%)	n.s.
	REP	6 (1.1%)	4 (0.5%)	n.s.
	Σ	73 (13.7%)	172 (21.2%)	$p < 0.05$
13 / 3	No. words	497	970	—
	UP	33 (6.6%)	63 (6.5%)	n.s.
	FP	24 (4.8%)	22 (2.3%)	$p < 0.05$
	PR	5 (1.0%)	14 (1.4%)	n.s.
	EET	1 (0.2%)	0	—
	MP	0	1 (0.1%)	—
	TR	1 (0.2%)	32 (3.3%)	$p < 0.05$
	REP	4 (0.8%)	19 (1.9%)	n.s.
	Σ	68 (13.7%)	151 (15.6%)	n.s.
41 / 4	No. words	1124	1796	—
	UP	114 (10.1%)	99 (5.5%)	$p < 0.05$
	FP	57 (5.1%)	32 (1.8%)	$p < 0.05$
	PR	5 (0.4%)	25 (1.4%)	$p < 0.05$
	EET	1 (0.08%)	2 (0.1%)	n.s.*
	MP	0	1 (0.05%)	—
	TR	4 (0.3%)	20 (1.1%)	$p < 0.05$
	REP	7 (0.6%)	38 (2.1%)	$p < 0.05$
	Σ	188 (16.7%)	217 (12.1%)	$p < 0.05$

* It could be argued that "n.s." in this case stands for "not sensible [calculation]" given that we are comparing a data set which is really not sufficiently big so as to provide a basis for statistical analysis. This is also the case in a few other instances in **Table 5.37a** through **Table 5.37c**.

Table 5.37b. Number of words and disfluencies for all subjects who participated in both WOZ-2 and Nymans, broken down for corpus and disfluency type. Actual numbers are given, as well as percentages for each disfluency type relative to the number of words in that corpus. Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level).

Subject number WOZ-2 / Nymans	Disfluency type	WOZ-2	Nymans	Significance (0.05 level)
33 / 5	No. words	283	1304	—
	UP	18 (6.4%)	65 (5.0%)	n.s.
	FP	0	9 (0.7%)	—
	PR	1 (0.3%)	3 (0.2%)	n.s.
	EET	0	1 (0.08%)	—
	MP	0	0	—
	TR	1 (0.3%)	26 (2.0%)	$p < 0.05$
	REP	2 (0.7%)	31 (2.4%)	$p < 0.05$
	Σ	22 (7.8%)	135 (10.3%)	n.s.
38 / 6	No. words	297	819	—
	UP	9 (3.0%)	50 (6.1%)	$p < 0.05$
	FP	3 (1.0%)	36 (4.4%)	$p < 0.05$
	PR	0	7 (0.8%)	—
	EET	0	2 (0.2%)	—
	MP	2 (0.7%)	0	—
	TR	0	4 (0.5%)	—
	REP	1 (0.3%)	9 (1.1%)	n.s.
	Σ	15 (5.0%)	108 (13.2%)	$p < 0.05$
46 / 7	No. words	384	1169	—
	UP	17 (4.4%)	50 (4.3%)	n.s.
	FP	10 (2.6%)	5 (0.4%)	$p < 0.05$
	PR	0	2 (0.2%)	—
	EET	0	1 (0.08%)	—
	MP	0	3 (0.3%)	—
	TR	0	8 (0.7%)	—
	REP	1 (0.3%)	10 (0.8%)	n.s.
	Σ	28 (7.3%)	79 (6.8%)	n.s.
35 / 8	No. words	827	1024	—
	UP	85 (10.3%)	106 (10.4%)	n.s.
	FP	16 (1.9%)	5 (0.5%)	$p < 0.05$
	PR	2 (0.3%)	14 (1.4%)	$p < 0.05$
	EET	0	3 (0.3%)	—
	MP	3 (0.4%)	2 (0.2%)	n.s.
	TR	8 (1.0%)	11 (1.1%)	n.s.
	REP	9 (1.1%)	20 (1.9%)	n.s.
	Σ	123 (14.9%)	161 (15.7%)	n.s.

Table 5.37c. Pooled number of words and disfluencies for all subjects who participated in both WOZ-2 and Nymans, broken down for corpus and disfluency type. Actual numbers are given, as well as percentages for each disfluency type relative to the number of words in that corpus. Statistical significance is given with the number of words weighed into the analysis (test-of-proportion, two-tailed, 0.05 level).

All subjects pooled WOZ-2 / Nymans	Disfluency type	WOZ-2	Nymans	Significance (0.05 level)
10 / 1 + 9 / 2 + 13 / 3 + 41 / 4 + 33 / 5 + 38 / 6 + 46 / 7 + 35 / 8	No. words	4448	9250	—
	UP	348 (7.8%)	562 (6.1%)	$p < 0.05$
	FP	147 (3.3%)	203 (2.2%)	$p < 0.05$
	PR	16 (0.4%)	129 (1.4%)	$p < 0.05$
	EET	5 (0.1%)	9 (0.1%)	n.s.
	MP	5 (0.1%)	10 (0.1%)	n.s.
	TR	24 (0.5%)	141 (1.5%)	$p < 0.05$
	REP	39 (0.9%)	172 (1.9%)	$p < 0.05$
	Σ	584 (13.1%)	1226 (13.2%)	n.s.

The first, more than obvious, observation that can be made from **Tables 5.37a** through **Table 5.37c** is, of course, that the subjects were very much more verbose in Nymans. As a matter of fact, to solve the exact same tasks, the subjects (pooled) used more than twice the amount of words as compared to WOZ-2, and one subject used over four times the number of words in Nymans than in WOZ-2 (subject 33/5).

As is shown in **Table 5.37c**, three subjects are significantly more disfluent in one corpus than the other—subjects 9/2 and 38/6 are more disfluent in Nymans, while subject 41/4 is more disfluent in WOZ-2—but the observed higher incidence of disfluency in Nymans (13.5%) is not significantly higher than the incidence in WOZ-2 (13.1%), so for all the data pooled one cannot say that there is a difference between the two corpora.

Regarding whether or not there were any differences as to specific disfluency rates between the corpora, the first thing to point out is that although there are significant differences here and there, it is hard to disinter any kind of tendency with regard to these differences—which by the way often are very small—and they tend to work "both ways" in that one person evinces more of a specific type of disfluency in WOZ-2, while another person produces more of that disfluency in Nymans.

As regards EETs and MPs, the incidence is too low to permit any meaningful statistical analysis. However, looking at **Table 5.37c**, we can see that for both corpora, both EETs and MPs occur roughly once per one thousand words, which is both extremely seldom, but also strikingly stable across the two corpora, irrespective of the lack of statistical verification.

The next obvious difference is that TRs are more common in Nymans. As we saw earlier men produced more TRs than women in Nymans, but it was also shown that this most likely is due to agent interruption. Moreover, as we saw earlier in **Table 5.36e**, disfluency production also differed significantly as a function of what particular agent the subjects were interacting with. The subjects who interacted with the male agent produced significantly more TRs than those who interacted with the female agent. As was pointed out earlier, the only relatively safe conclusion one can draw from this observation is that there are finer distinctions to consider than whether or not the agent is a machine (real or fake) or a human being. Whether this is

due to the gender or personality of the agent remains to be studied, but my guess would be that gender *might* play a role, while personality is *very likely* to play a role.

As for UPs and FPs, both types occur significantly more often in WOZ-2, while PRs and REPs occur significantly more often in Nymans. While far-reaching conclusions regarding *why* this is observed must await a more detailed analysis of the data, at least *one* interesting hypothesis may be forwarded: according to the "floor-holding" hypothesis FPs serve the purpose of preventing interruption from an interlocutor, and the observation that FPs occur more often in the human–machine corpus seems to be at odds with that particular interpretation, since the risk of being interrupted in WOZ-2 was far less marked than in the human–human corpus (Nymans). On the contrary, according to the "many-options/major planning" hypothesis FPs occur where major planning decisions are made by the speakers, and many possible routes are available (e.g. at the beginning of utterances). That subjects produced more FPs in the human–machine corpus could be attributed to the hypothesis that while the subjects received help from the agents in Nymans, no such help was available in the human–machine corpus, and the subjects were consequently "on their own" in making decisions about their journeys.

So, although significant differences are observed, these differences are not overwhelming, and any discussion as to why these differences occur must be a cautious one, pending a detailed analysis of the interaction. The most striking observation is perhaps that the differences are *not* marked, as one would perhaps have expected. The only really palpable difference is the number of words used to solve the tasks, where subjects were much more verbose in Nymans than in WOZ-2.

To sum up, the bottom line of the cross-corpus comparison is that there are no statistically significant overall differences, and the differences that occur—significant or not—are too slight so as to permit any far-reaching conclusions.

5.13 Other observations

Besides specific analyses of overall and specific disfluency rates, there are a number of other, sundry, observations that can be made from the data, and I will summarize some of these in this section.

5.13.1 Individual differences

It has been noted in the literature (e.g. Oviatt, 1995; Bell, Eklund & Gustafson, 2000) that disfluency production is subject to marked individual differences. For instance, Bell, Eklund & Gustafson (2000) found a ratio of 1–15 between the least and most disfluent subject in their study. As was also shown, at least one subject approached well-nigh complete fluency. Oviatt (1995) observed a variation "of two to 11-fold" (Oviatt, 1995, p. 33) and suggested that "categories of spoken language should be studied individually" (ibid., loc. cit.).

The percentages of the least and most fluent subject in all corpora, including and excluding unfilled pauses, are shown in **Table 5.38**.

Table 5.38. Least and most disfluent subject in all corpora, including and excluding unfilled pauses, given as percentages of total number of words.

Corpus	Least disfluent subject		Most disfluent subject		Ratio most/least	
	Incl. UPs	Excl. UPs	Incl. UPs	Excl. UPs	Incl. UPs	Excl. UPs
WOZ-1	2.5%	0.2%	23.9%	12.7%	9.6	63.6
WOZ-2	5.0%	0.6%	31.5%	21.6%	6.3	36
Nymans	6.7%	2.9%	21.2%	13.3%	3.2	4.6
Bionic	4.3%	2.8%	31.8%	16.8%	7.4	6
Mean	4.6%	1.6%	27.1%	16.1%	5.9	10

So, differences between least and most disfluent subjects in the corpora range from slightly more than three times more disfluent (Nymans, including unfilled pauses) to a staggering 63.5 times more disfluent (WOZ-1, excluding unfilled pauses). For all data pooled, the figures range from 5.9 (including unfilled pauses) to 10 (excluding unfilled pauses), so generally speaking, there is a magnitude of order of difference between speakers' rates disfluency (given that unfilled pauses most often are not included in counts).

As regards the most fluent speakers, who exhibit 0.2% and 0.6% disfluency, it should be pointed out that these speakers are not isolated cases. In fact, several speakers approach nigh-total fluency, especially when one excludes unfilled pauses. In WOZ-1, speaker 9 (female), exhibits only 1.8% percent disfluency, speaker 25 (female) 1.6%, speaker 26 (male), 0.3% (in fact, he produces only *one* disfluency except unfilled pauses, which are few), speaker 34 (male) 1.5%, and speaker 50 (female) exhibits only 0.2% disfluency (also only *one* non-unfilled pause: a prolongation). In WOZ-2, speaker 6 (male) produces 1.7% disfluency, speaker 17 (female) 1.8%, speaker 21 (male) 1.1%, speaker 33 (male) 1.4%, and speaker 48 (female) only 0.6%. In Nymans and Bionic, no such exceptional fluency is observed.

From this we can deduce that fluency is quite possible, if not common, and that some speakers approach almost total fluency in a constrained and familiar domain such as travel booking, despite the unusual circumstances.

5.13.2 Meta-comments

As was pointed out in **section 3.2.2**, WOZ simulations only take us so far, and it must be borne in mind that subjects are not actually performing real-life actions when in the laboratory, however life-like the researcher tries to make it look. Hopefully, some of the behavior we (the researchers) want to model is beyond the conscious control of the subjects, and thus, once again, hopefully (but far from certainly) less prone to coloring from the discrepancy between carrying out a real need in a real-life situation, and carrying out a fake need in the laboratory. One such example could be the observation that all subjects exhibited ingressive phonation in the human–human corpus, but no one in the human–"machine" corpus, as has been mentioned before.

However, there are a few explicit examples that clearly indicate that the subject is aware of "playing a game". For instance, subject 5 (first dialogue) in Nymans (human–human), includes the following utterance, directed to the agent:

Ja du vet det här du vet är fingerat va <fniss> men nu står det att det är den sjätte juni
"Well you know this you know is simulated <giggle> but the way it's presented it says the sixth of June"

In this example, there was evidently a mismatch between the date (June 6th) and the weekday. This, of course, could happen also in real life, but the reaction here shows an awareness of the simulatedness of the situation.

There are also a number—albeit small—of self-addressed comments that border on the "meta", mostly commenting on vague or opaque instructions. (Remember that the task sheets were deliberately made unclear so as to yield more variety in the linguistic output.) Such examples include:

Subject 8 in Nyman (human–human), dialogue 1:

Vad står det här?
"What does it say here?"

This is accompanied by the rustling of the task sheet.

Although the number of such overt indications of subject awareness of the tasks' detachment from real life is slight, examples as the ones given above clearly indicate that it cannot be taken for granted that the elicited, or exhibited, behavior stands in a one-to-one relationship with the same persons' actual behavior in a similar (not identical) situation in real life. As has been mentioned before, there are good reasons to assume that WOZ simulations result in much more authentic-looking data than a lot of alternative methods, but authentic-*looking* is still not equivalent to *authentic*.

5.13.3 Overlapping communication in human–human setting

Much of the difference between human–human and human–machine corpora are most likely due to more frequent overlapping conversation in the former setting. However, given that only subjects have been digitized (for reasons given earlier), it has been hard to pinpoint such occasions with any great certainty. In some cases it is more or less obvious that the subject has been interrupted by the agent, e.g. when the final word of an utterance is suddenly stopped short. In order to avoid any major influence from such occasions, utterance-final truncations have not been included in the labeling. That overlapping speech makes it harder to define what is an "utterance" is clear, even if both interlocutors are available for analysis. Sometimes the subject begins an utterance, stops short for what appears to be an interruption from the agent, and then continues as if the silent interval had not existed in the first place. Indeed, if you cut out the silent part of such instances, and paste together the two bits, they result in a perfectly natural-sounding, unbroken utterance. So, it could be argued that "an utterance" can be separated in time by utterances made by an interlocutor. However, there is no possibility to provide here more than these more or less impressionistic accounts of what is clearly the case, but beyond the present means of detailed analysis.

5.14 Main findings

This chapter has provided preliminary observations on disfluency incidence in the studied corpora. It goes without saying that the results presented represent only a minuscule amount of the studies that could—and should—be carried out, but due to limited space and scope of this work, we will have to settle with what has been presented here.

So, given these observations, and seen against the backdrop provided in chapter two, what results seem to be confirmed in our observations on Swedish, and what seems to run counter to the previously published studies on (mainly) other languages? I will here try to briefly summarize the most important findings.

5.14.1 General frequency

The general frequency of disfluency more or less confirms previous reports on disfluency frequency, and the overall figure of 6.4% is comparable to what is found in the literature. However, since it is often not clear exactly what has been counted or included in previous studies, there are certain problems associated with comparing across studies, and the notion that the overall disfluency figure is comparable to other studies is to a certain degree circular: we assume that we have included the same disfluencies in the count *because* the figure is more or less the same as in previous, rather than knowing beforehand what we are counting, and *then* compare the figures.

It was also found that the corpora were significantly different from each other, which was partly a function of settings and modes, but most likely also partly a function of task details and methodological set-up. That speakers in Bionic faced no constraints on utterance length resulted in higher rates of disfluency than in the other corpora.

5.14.2 General distribution of disfluencies

There is a correlation between utterance length and disfluency incidence, as evidenced by regression analysis. This has been shown several times in the literature (e.g. Oviatt, 1995, p. 32; Oviatt, 2000, p. 880), and cannot be seen as a revolutionary observation. Roughly speaking, for utterances of ten words' length, about half are completely fluent, while almost no twenty-word utterances are entirely fluent. Seen in the light of the aforementioned higher disfluency rates in Bionic, when designing a human–machine interface, one way of keeping disfluency rates low would be to, in one way or another, try to keep subject utterances short.

5.14.3 Unfilled pauses

Unfilled pauses are by far the most common of all disfluent types, but are also the most problematic in that they span from sure-fire disfluency to planned structure strategies from the speakers. However, even if one cuts away the lower 25% of all unfilled pauses, or employs a 250 ms lower cut-off, unfilled pauses still exceed in quantity any other disfluency type. Unfilled pauses are also significantly longer than filled pauses.

As for distribution, unfilled pauses seem to have a rather even distribution, with a slight tendency to appear immediately prior to nominal phrases or subclauses, making them 'NP-prone', or "subclause-prone' The previously reported observations that unfilled pauses occur at clause boundaries thus receives some support in this study given the slight tendency to

appear prior to conjunctions. However, a more detailed study of syntax needs be made before any far-reaching conclusions may be drawn.

5.14.4 Filled pauses

Filled pauses are the second most common type in all corpora, but exhibit a very different distribution, as compared to unfilled pauses. The two most common positions are utterance-initially and immediately prior to another type of disfluency, making filled pauses some kind of 'disfluency-disfluency', which might possibly signal speech planning problems at a higher level. Contrary to unfilled pauses, they do not occur prior to conjunctions.

The suggested hypothesis that filled pauses occur where many options are available to the speaker seems to be confirmed in that almost 50% of filled pauses occur utterance-initially, where no commitment has yet been made by the speaker, while the alternative floor-holding hypothesis seems to be contradicted by the observation that filled pauses occur significantly less in the human–human corpus than in the three human–machine corpora, an observation that could lend even further support to the 'many-options' hypothesis, in that planning problems are less in the human–human setting where the agents could provide help as soon as options and/or problems appeared, thus decreasing the number of problematic situations.

5.14.5 Prolongations

Contrary to the old 'tell-tale sign of stuttering notion', prolongations have been shown to be very common in the speech of nonstutterers. In fact, for all corpora, prolongations are the third most common type, by a large margin.

The durations of prolongations are significantly lower than the duration of filled pauses, but of interest is the observation that all kinds of segments are subject to prolongation, not only continuants, which perhaps would be the intuitive notion.

Cross-linguistic studies from other languages, such as American English, Tok Pisin, Japanese and Mandarin, point to language-specific distribution within the word, which could possibly be a function of the morphological complexity of the language in question.

5.14.6 Floor-holding revisited

Filled pauses have been suggested to serve as a floor-holder when a speaker needs some additional time to plan the utterance, but do not want to cede the floor to the interlocutor(s). Filled pauses also share with prolongations the trait of being both durational and vocalized, which is what makes it hard to interrupt a "filled pausing" speaker, as opposed to an "unfilled pausing" speaker. Consequently, if filled pauses indeed serve the function of keeping the "conversational ball" (Maclay & Osgood, 1959, p. 41), then prolongation should fill that role equally well, if not better, since prolongations occur on a word, rather than between words, as proposed by Streeck (1996).[1]

[1] While unfilled pauses and prolongations are found inside roots in several languages, no such example of a filled pause is mentioned in the literature, to the best of my knowledge.

Streeck (1996) discusses "stretched-out sounds" (prolongations) in Ilokano—an agglutinating Philippine language—in detail. He also states that:

> [S]ound stretches /… / are thus comparable to the familiar items *uh*, *uhm*, etc. and their cross-cultural variants. Such fillers, however, are almost totally absent from Ilokano conversations. /… / Perhaps this is all there is to it: where speakers of English produce a 'filler', Ilokano speakers simply stretch out the last vowel before the trouble source. The effect is the same. (Streeck, 1996, p. 195.)

That Ilokano should lack filler words (or almost so) is challenged by Rubino (1996), who presents disfluency data that include both unfilled pauses and filled pauses. However, Streeck and Rubino agree with regard to the analysis of the important role played by prolongation in Ilokano. They also agree on the phonological constraints prolongations are subject to, and that they almost exclusively appear on word- or prefix-final vowels.

The notion that prolongations should work even better than filled pauses as floor-holders is also suggested by Streeck:

> Ilokano speakers not only continue to vocalize, but also to *speak*: they never cease to say words. (Streeck, 1996, p. 195; italics in original.)

Whether or not prolongations do indeed serve that function cannot be concluded in the present data set unless a detailed analysis of the entire discourse is carried out, i.e. both interlocutors. Consequently, this hypothesis will have to await further studies.

5.14.7 Durations: unfilled pauses vs. filled pauses vs. prolongations

The observation that filled pauses generally have longer duration than prolongations—despite their sharing of the two traits durational and vocalized—could imply that filled pauses have a different status in speakers' minds, and are viewed as words in their own right. Also, that prolongations, unlike filled pauses, are observed in word-medial position is another trait that implies that prolongations and filled pauses do not have the same status in speech production.

Finally, it was shown that not only are the frequencies and mean durations of unfilled pauses, filled pauses and prolongations different, so are their distributions, with a steeper fall-off for prolongations than for filled pauses, which in turn fall off faster than do unfilled pauses.

5.14.8 Explicit editing terms

While Shames & Sherrick (1963) stated that speakers frequently edit their speech, and Hindle (1983) proposed that disfluency detection makes use of explicit editing, the present study confirms the results of e.g. Eklund & Shriberg (1998) that explicit editing is extremely rare. Contrary to what could have been expected, the subjects did not provide more EETs in the human–human setting, but the general pattern that Bionic was the most disfluent corpus was observed. However, the dearth of data makes all conclusions tentative at best. Suffice it to restate that explicit editing is a rare phenomenon.

5.14.9 Mispronunciations

Mispronunciations, or slips-of-the-tongue, are also rare, pointing to a certain fluency in spontaneous speech: speakers are indeed proficient at phone levels of speech production.

Only around half of the mispronunciations are repaired, while half are left "as is" by the speaker. Whether or not this is due to the fact that the speakers detect their own mispronunciations is not known in the present study.

5.14.10 Truncations

Truncations appear more often in the human–human setting, more than likely being caused by interlocutor interruptions. This by itself may not be a breathtaking observation, but it is still of interest from a speech production point of view in that words obviously are not produced as complete entities before being spoken out, but are produced and spoken out on the fly, and that it is possible to stop speaking mid-word. However, the observation that Bionic exhibits more disfluency than WOZ-1 and WOZ-2 also reveals that speakers tend to interrupt themselves if they are given the opportunity, that is speaking out long utterances.

5.14.11 Repairs

Repairs differ from all other disfluencies in that they are structured, complex and might include any of the other types. The preliminary analysis carried out here revealed no *obvious* patterns as to what either the Reparandum or the Reparans part includes, which does not mean that such tendencies do not exist, if bestowed a more careful study.

Verbatim back-tracking (or retracing) had similar frequency distribution in all corpora, and it would seem as if the maximum number of verbatim repeated words is four, in all corpora, excepting a couple of outliers, including the exceptional case with 12 verbatim repeated words (not including any other type of disfluency or change). The entailing conclusion, of course, is that speakers in general do not back up too far when starting over.

5.14.12 Gender differences

As was shown, it is hard either to corroborate or rebut previous reports in the literature, in that the results here are equivocal. In three of the corpora (WOZ-1, Nymans and Bionic) there were no gender differences—except that both women and men produced significantly more disfluencies with regard to specific categories in some instances—while men were significantly more disfluent in one corpus (WOZ-2), which carried over to the two total counts in the merged data set. Regrettably, there is no easy way out to explain this observation in that WOZ-2 shares too many traits with other corpora to make readily available explanations cumbersome. Either there is an intricate interaction between several different parameters, or the subjects in either WOZ-2—or the other three corpora—were not "representative", whatever that might mean. In any case, an explanation will have to await a more detailed study of the data, and for now we can only observe that men are more disfluent in one corpus, and that there are no gender differences in the other three.

5.14.13 Cross-corpus observations

Contrary to what might perhaps have been expected, there were no clear differences between WOZ-2 and Nymans as regards overall disfluency rates. Unfilled pauses and filled pauses were produced more frequently in WOZ-2—the latter observation lending more support to the "many-options" hypothesis as opposed to the "floor-holding" hypothesis—while truncations and repairs were more common in Nymans—the former observation more than likely due to agent interruption. Despite these differences (which were statistically significant, albeit

weakly), there was no overall statistical difference between the two corpora as to disfluency production. The only marked difference was that subjects were much more verbose in the human–human setting, which could be ascribed to the fact that more roundabout chit-chatting occurs, as do more affirmations (compare Eklund, 2002, on ingressive speech, which commonly occurs on the third repetition of an affirmation).

5.14.14 Exceptional fluency

While the typical subject exhibits around 6% disfluency, it was also discovered that it is indeed possible to be (almost) entirely fluent. Subject 26 in WOZ-1 uttered 364 words in 63 utterances, and besides one repair, he only produced unfilled pauses, some of which might have had a structuring function. Subject 50 in WOZ-1 also produced only one disfluency besides unfilled pauses, and several other speakers in both WOZ-1 and WOZ-2 produced very few instances of disfluency overall. This shows that it is indeed possible to speak more or less completely fluently, even if it is marked behavior. This entails that there is no inherent *necessity* to be disfluent, only that it is much more common than being fluent.

5.14.15 WOZ limitations

It must be borne in mind that a simulation is a simulation is a simulation. Although WOZ simulations have proven to provide high-quality data, and that they by far are superior to most alternatives, they still are not the "real thing", and the fact that the participating subjects are not carrying out real-life tasks of concern to their own lives probably shows up in their language. That subjects are aware of the simulated situation is shown by the occurrence of meta-comment, although the number of such comments is small.

5.15 Final comments

This chapter could have dived into the data in much more detail, but given its limited scope had to be restricted to general observations. Still, seen in the light of previous research, findings and hypotheses, the observations reported here both point to similarities and confirmations, as well as to differences as compared to results reported in the literature over the past fifty years.

The present study is, admittedly, a tad lopsided in that prolongations are given relatively much study, while repairs are somewhat neglected. This is partly so for historical reasons (work carried out over the years), but also due to the different problems that repairs pose from an analysis point of view. One could also claim that the lack of detailed study of repairs is tantamount to the (relative) lack of study devoted to the potential interaction of all other disfluency types. "Nothing can be about everything", and so on. This, of course, does not mean that I do not find such studies interesting, and I am convinced that there are things to discover, both concerning interaction in general, and concerning the structure of repairs in particular. Pending future studies, however, we will have to settle with this for now.

Another thing I want to point out is that some of the studies one would like to carry out require a more detailed study of the *entire discourse* in general—including both the subjects' and the system's or agents' speech acts—as well as a study of *what* is being said. And this is where we go full circle from the studies that were initiated by Mahl and colleagues, who started to disregard a contents analysis of what was being said in the psychological interview,

and instead focused on *how* it was said. This, above all (methinks), points to the fact that content and form are intertwining, inter-dependent phenomena, and that one can only reach so far by looking at the one without considering the other.

In the final chapter, I will try to summarize what I think are the most interesting aspects of this field, and point to future studies I deem of interest to further our knowledge about disfluency production in human spontaneous speech.

6 Conclusions and future research

6.1 Introduction

I will now try to sum up what findings I deem to be the most interesting and will also try to point out in what way they relate to previously reported research, albeit in a very (overly) succinct way (which means that the reader should not assume an exhaustive account here). Let me begin with a quote from Chafe (1980) that might as well have introduced this book, but serves its purpose as well here, and is worth considering, wherever it appears.

> There is a natural tendency, when some interesting phenomenon is being explored, to want to treat it as something which can be studied in and of itself, without regard for its interrelationships with other phenomena. The entire field of linguistics has to some extent suffered from this tendency, in that a great deal of research has attempted to deal with language apart from its psychological, social, and cultural settings. It is a healthy development that fields like psycholinguistics, sociolinguistics, and ethnolinguistics have begun to bring a broader perspective to linguistic studies. On a different level it has seemed to me that there has been the same tendency in research on hesitations, or pausology, or whatever it may be called, to look at the phenomenon in isolation. But in the long run I am sure we are going to find that such a specialization of effort is futile; that hesitational phenomena can be understood only as natural consequences of the processes which occur during the production of speech. Viewed in that way, they can be seen as contributing important clues to the nature of these processes. (Chafe, 1980, p. 169.)

However, the entire blame is not put on linguistics. From a stuttering perspective, Wingate (1987) argued for greater openness vis-à-vis findings within general linguistics:

> At approximately the same time, two lines of research have studied disfluencies from different orientations—one in stuttering and the other in normal speech. In certain important respects the findings of these separate lines differ. Resolution of these differences, which is particularly important for the understanding of stuttering in its relation to disfluency and fluency, has been precluded because the two research areas have remained essentially isolated from each other. (Wingate, 1987, p. 79. [1])

[1] Wingate included the work carried out within psychotherapy by George F. Mahl and colleagues in the category "normal disfluency", as spearheaded by Frieda Goldman-Eisler, but pointed out that "Goldman-Eisler /.../ and Mahl /.../ pursued their investigations independently, and evidently unaware, of the other" (Wingate, 1987, p. 86).

In the unlikely event of a reader who has not already suspected that this is the case, the above quotes could well serve as my own program: to view disfluency ("pausology" in Chafe's wording) in a broader perspective. Of course, there is most likely no possible way to collect a data set of speech which will enable a researcher to study *all* of the aspects included in chapter two of this book—and most likely not even desirable to attempt it—but whatever analyses can be carried out on a specific data set, like the one studied here, it is my contention that the larger picture should be borne in mind.

6.2 Most important findings

I will now try to sum up what findings I deem to be the most interesting, and also try to point out what future studies might be carried out on the present data set, both as it stands with the present labeling, and what could be done with additional labeling. I will also spend some time on what I think would be interesting future studies on disfluency in general, but lie beyond what is possible to do on the present data set, mainly given the way data were collected (or rather, were *not* collected).

So, what are the main findings of this work? That, of course, depends on your perspective, and different people will most likely find different parts of this work more or less interesting. However, I will try to summarize what I find interesting, seen against the backdrop of previously conducted research and general assumptions that circulate in the literature.

6.2.1 General frequency

The first observation is that the current study repeats the reported overall disfluency rates, i.e. somewhere around 6%. It is, of course, of interest to note the stability of this figure, given the range of different settings and languages studied, and although there are different ways to count disfluencies (as discussed previously) and although there is a great deal of individual variation as to disfluency rates, the present analysis can only confirm that human speakers are disfluent on every 20th word on average. Also, it was shown that disfluency production to a large extent is a linear function of utterance length which has been pointed out previously in the literature.

6.2.2 General distribution

Although the distributional study undertaken in this work must be taken cum grano salis, it seems safe to conclude that unfilled pauses and filled pauses have different distributions in the data. First of all, almost half of all filled pauses occur utterance-initially. Second, while unfilled pauses exhibit a slight tendency to occur prior to noun-phrases (or rather, prepositions) and a slightly stronger tendency to occur immediately prior to conjunctions (i.e. subordinate clauses) filled pauses are over-represented immediately before other types of disfluency, thus signaling speech planning problems in general.

6.2.3 Unfilled pauses

Although often excluded from disfluency studies, possibly due to their problematic "status", they are by far the most common way to be disfluent. Even with the lower quartile excised, they are still more common than the second most common type, filled pauses, so the original studies by Goldman-Eisler on hesitations started with the "right" phenomenon, as it were, since the most common way to be disfluent is to be silent. The present study seems to suggest

that unfilled pauses are 'subordinate clause-prone', although the analysis on which this is based is somewhat indirect in that it only looked at the word immediately following the unfilled pause. However, a more detailed study of the present data could easily verify whether this hypothesis holds given a more rigorous study.

6.2.4 Filled pauses

Besides being the second most common type, the most obvious observation concerning filled pauses is that almost half of them occur utterance-initially. This seems to support the notion that filled pauses signal that there are many options available to the speaker, and that he or she has yet to commit to one of the available alternatives. That filled pauses are associated with general planning problems could be supported by the additional information that they are less common in the human–human setting where it could be assumed that any potential problem for the subjects was acted upon by the human agents, which reduced the number of planning-associated hesitations. The alternative hypothesis—that filled pauses serve the function of keeping the conversational floor—does not receive support in the present study given that filled pauses are less frequent in the human–human setting, where interlocutor interruption is far more imminent—the wizards were all fairly "taciturn".

6.2.5 Prolongations

It can be safely stated that the old view that prolongation was a bona fide acid test of stuttering is not the entire story. Of course, detailed differences as to how and when young children exhibit prolongation might reveal an imminent danger of developing stuttering, but it is clear that prolongations are legion in the speech of nonstutterers as evidenced in the present data set. Another interesting observation is that any type of segment is subject to prolongation, not only continuants, which perhaps would have been expected. Perhaps the most interesting observation is that there seem to be language-specific traits at play here. While word-position ratio in Swedish and English is 30–20–50 ratio for initial, medial and final segments, respectively, morphologically less complex languages like Tok Pisin, Japanese and Mandarin exhibit figures that are very different from that found in Swedish or English. This shows not only that prolongation occurs in typologically different languages, but also that these typological differences might play a role in how prolongations occur in a given language. Regrettably, the studies done on Tagalog and Ilokano are not detailed enough so as to allow a fine-grained comparison.

6.2.6 Floor-holding

It has been suggested that prolongations are even stronger floor-holders than are filled pauses since the speaker does not cease to utter real words. Although no strict floor-holding study was done at the present data set—since that would require a study of who is not allowed to enter the floor, i.e., the interlocutor(s)—it was shown that prolongations are significantly shorter than filled pauses, in addition to being less frequent. However, there is some variation at the individual level with regard to production of filled pauses vis-à-vis prolongations.

6.2.7 Explicit editing terms

The most obvious thing about explicit editing is how rare it is. In fact, it is so rare that this is more or less the only thing one can say about it, since even in a study of this size, their number is not great enough to allow statistical analysis. However, contrary to what might

have been expected, the subjects did not evince more explicit editing when speaking to humans as compared to machines, but as I said before, the number is really not big enough to allow any final conclusions other than their scarcity.

6.2.8 Mispronunciations

Like explicit editing, mispronunciation is rare, which of course has been acknowledged from the beginning of speech production studies where slips most often were elicited in order to obtain linguistic data. This shows that whatever the reason for disfluency, it is rarely a motor problem, while at the same time it is interesting that it is so easy to elicit slips, given how rare they are in spontaneous speech. In the present data set, about half the mispronunciations are repaired, but whether this means that the other half went unnoticed by the speaker is not known. Perhaps the speaker just did not care to repair them, or perhaps the speaker did not notice having made the mistake.

6.2.9 Truncations

Despite the fact that no utterance-final truncations were included in the analysis, the human–human corpus exhibits a higher number of truncations, most likely due to interlocutor interruptions. As pointed out before, a study of interlocutor speech is needed to verify this notion.

6.2.10 Repairs

As was pointed out previously, repairs differ in some respects from the previous set of "atomic" disfluencies, and are mainly a mixture of all other types of disfluency—for example 50% of mispronunciations lead to a repair—plus some additional phenomena like insertions, deletions, substitutions and repetitions. Of all the studies in this work, repairs are the most neglected and in most need of a more detailed study, possibly in order to disinter tendencies with regard to characteristics special to either the Reparandum or the Reparans. However, verbatim back-tracking revealed that speakers exhibited the same patterns across the corpora, with a practical observed upper limit of four verbatim repeated words.

6.2.11 Gender differences

In a way, the present study repeats the literature in that we both find support for the notion that men are more disfluent than women *and* also find no gender differences. In three of the corpora, there are no differences, or the few significant differences that are found work in both ways. In one corpus, however, men are significantly more disfluent than women. In conclusion, then, the present study produces equivocal results and we can only restate what is said above: in three of the corpora, there are no differences, while in one corpus there are.

6.2.12 Cross-corpus differences

The eight subjects that participated both in WOZ-2 and in Nymans behaved more or less the same in the two settings. While somewhat surprising, this could be taken as good news for developers of automatic systems since the speech data obtained in the WOZ collection were representative of the speech produced in a human–human setting. The only major difference between the two corpora was that the subjects used many more words to make the exact same bookings.

6.2.13 Fluency is possible

As a final observation it should be pointed out that some of the subjects in WOZ-1 and WOZ-2 approached total fluency if unfilled pauses are excluded. This is perhaps most interesting from a psychological or philosophical perspective, but it is still interesting that extemporized (more or less) spontaneous speech can be fluent, instead of being a reserved domain for professional actors who have rehearsed their lines.

6.3 Future work

So, how do I envision future work? To begin, there is the need to divide this paragraph into different sections. First, what kind of future work could be carried out on the present data set, the way it stands now, labeled as it is. Second, what kind of extensions to the present data could be done enabling other additional studies to be performed? Third, what kinds of future studies would not be possible to carry out on the present data set, irrespective or whether I, or anyone else, would find such studies of importance or interesting? I will try to answer all those questions in the following sections.

6.3.1 Possible work, the way things are now

I will here describe work that could be done on the present data set the way the data are transcribed and labeled right now.

6.3.1.1 More of the same

As far as the present analyses go, they are in no way exhausted. There are innumerable studies concerning frequency, distribution, statistical differences between groups and types and group-type that could be carried out, but which given the present lack of time and space were not done in this thesis. The interaction between any given type of disfluency and any other category—be it another type of disfluency, another corpora, another type of utterance, task, locus within that utterance, speaker gender and so on and so forth—is in no way even begun to be studied here.

6.3.1.2 Speech production model testing

One of my main points when discussing speech production models in chapter two was that the intricate timing relationships of different events have been overlooked to some extent in the speech production literature. Given that the present data-set is labeled in detail for durations between correctly executed words and phrases as well as pauses, repairs, prolongations, editing comments, truncations and so on, one could surely use the data as a test bench for speech production models. If repair (or disfluency) *X* appears at time *t1*, and is repaired at time *t2*, when must detection have occurred, given the latency of motor execution (of the speech organs)? Given the results by Libet et al., what is the role of consciousness, i.e., *conscious* monitoring, as claimed within certain (but not all) speech production models? How far ahead of *X* would it have to be detected (by an inner loop, for example), in order to enable the speech organs to actually move at *t2*? Also, as we have seen, work done on the arena of neuroscience has revealed neural correlates of linguistic phenomena in general (that semantics seemingly precedes syntax in perception, for example) and on pauses in particular. Work has also been done on how speech production shows up in brain potentials of various sorts (ERP, CNV, N400 and so on). It would be of great interest to see whether spontaneously produced

disfluencies would leave specific traces in the neurological arena, and if so, whether the different role filled pauses (as opposed to other disfluencies) seems to play from a number various perspectives would have a distinct neurological correlate.

6.3.1.3 Crosslinguistic comparison

Although I have already begun such work, far more could be done in comparing frequencies, distribution and categorization of disfluencies across languages. To enable other researchers to do such work, I have included detailed tables of my corpora in **Appendices 1–5**. It must be noted, however, that distributional information in the said appendices is limited to a distinction between utterance-initial filled pauses and filled pauses at other locations.

6.3.1.4 Effects of disfluency

Given that the data are now labeled, and that we consequently know where disfluencies occur in the recorded speech, it would be easy to play back the material to listeners in order to study the particular type and to what degree disfluencies are noticed (or not) by listeners. Thus, knowledge could be gained as to the particular type of disfluency, as well as its particular location in an utterance as a function of whether it is either noticed or missed in perception tests.

6.3.2 Possible work, with extended labeling of the data

As was mentioned before, the present data-set is in no way exhausted as regards analyses, not only from the perspective of what is already "there for the grabbing", but also what could be done given that additional labeling be carried out. I will here describe work that *could* be carried out on the present data-set given extra time and effort.

6.3.2.1 Speech act analysis

The present data set could easily be labeled for speech acts—an undertaking already commenced by Ask & Decker (2001a, 2001b) on a subset of the data. That speech acts play an important role in verbal communication, be it human–human or human–machine has been shown over and over again in the literature and there is no doubt that such an analysis would shed further light on how future automatic dialogue systems could, or should, be designed.

6.3.2.2 Prosodic analysis

The present data, with already generated F_0 contours and already carefully analyzed as to durations (at least at the word level), would easily lend itself to prosodic analyses, as well. In fact, the main reason why this was not carried out in this study was the notion that prosody should preferably be labeled by someone other than the analyzing researcher which was not feasible at the present stage.[1]

[1] During the initial stages, transcription and labeling was not carried out by me, and did at that time include prosodic labeling, as reported in Eklund (1997). This activity was terminated in 1998, however, whereupon labeling and analysis was reduced to disfluencies.

6.3.2.3 Syntactic analysis

The present data is not tagged or parsed for word classes, linguistic function or phrasal categories. If such work were carried out, manually or automatically (or a combination of these two methods), further distributional analyses could be carried out. The observations made in this thesis as to word classes were all carried out by post-labeling manual inspection, not on pre-tagged data (meaning that the label files—see **Appendix 5 Transcription Sample**—do not include that information).

6.3.3 Not possible work on the present data set—but still of interest

There are many things that are not covered in the present data set—in fact, more things than are covered. Some of these concern knowledge about the subjects and how the data could be labeled, keeping the data the way they appear. Some are beyond reach simply because the data collections were carried out the way they were. I will here mention some of the studies that cannot be done on the present material, however interesting it would be.

6.3.3.1 General

The first limitation includes things like lack of e.g., video recordings of the subjects, which makes it impossible to study **gestures** or **facial expressions** during the sessions (which is known to occur even during telephone conversations). Moreover, we did not measure **brain potentials** or **galvanic skin responses**, nor did we measure **hemispheric lateralization** or **palmar sweat** of the subjects. This precludes any such analysis of stress levels in the subjects, however interesting that would have been. Along the same line, no subjects were asked to complete any **personality tests** as employed within stuttering and psychology research, which puts all analyses as to the correlation between personal traits and disfluency production beyond reach. Along similar lines (although more anecdotal), we did not measure **alcohol levels** in the subjects (which were, however, presumably zero), or ask the female subjects if they were **ovulating** at the time they carried out the tasks. We did not interview the subjects' parents about their view on **upbringing**, we did not administer nembutal or benzodrine or any other **drugs**, and we did not gauge the level of **sense of humor** in our subjects. And so on and so forth ad infinitum (asymptotically).

This does not preclude me from envisioning future research that, in my view, should be carried out anyway, even if the present data set cannot be used as a primary source. In a way, we are still only in the beginning phases of understanding how disfluencies occur in human communication and how they reflect human behavior in general. Some of this research is of commercial interest, whereas some is of more profound, scientific or philosophical interest. Whether your are a designer of a cutting-edge automatic dialogue system, or a neurophilosopher wanting to comprehend what our "mind stuff" is, the basic question that is being posed when looking at disfluencies is *why* are they produced. For a software engineer/system designer, a speech act answer may be enough to enhance the performance to hitherto unheard of success rates, and there is no need to delve deeper into the metaphysical, or neurological, ocean. For the philosopher or neurologist, such an answer would probably still be unsatisfactory and even "superficial". Consequently, one could argue that the basic question asked within all (or at least most) the previously mentioned research fields—stuttering, psychology, philosophy, linguistics, system design—is basically the same: *why do we produce disfluencies?* The main difference between fields would be the *depth* of the answer which is deemed satisfactory.

6.3.3.2 Multimodality

That gestures are interrelated with speech production in general, and speech production in particular, has been shown over and over again, as was shown in chapter two. Future systems might well employ computer vision systems that could make use of the subject's/user's gestural (arms, hands, torso, head, eyebrows etc.) behavior to arrive at more robust speech recognition. Moreover, multimodal assistants (PDAs) already include on-screen pens for simultaneous speech–pen input, and disfluency detection in writing could also provide a system for supporting hypothesis in the task of speech recognition.

6.3.3.3 Speech recognition and children

There are to date very few, and still rather rudimentary, automatic speech recognizers that handle children's speech well. It is general knowledge that children are notoriously disfluent and also that their disfluency is not exactly comparable to disfluency in adults. With the development of better recognizers, studies of children's disfluency are beginning to appear. Not only have all different age groups been studied and compared, several studies have also been devoted to different speaking situations. Granted, no studies from the '30s, '40s, '50s, '60s or even '70s have examined computer-directed speech, but by looking at what is known about children's "baseline" disfluency at various ages would assumedly not only make it easier to create hypotheses on what kind of dialogue interaction will be difficult, but also to pinpoint in a safer way exactly what disfluency phenomena are due to the system under scrutiny.

6.3.3.4 Disfluency and consciousness

While references to neurology seem to be completely absent within linguistics and references to philosophy are scant, the opposite is not true: neurologists and philosophers do refer to linguistics, and even mention speech errors and disfluency, when discussing consciousness, timing events and motor execution. I, for one, would like to see a Libet/Haggard/Frith-style experiment carried out in the field of speech production in general, and disfluency in particular. If I were given a wish, that would be *The Book* (or article) for which I would pay almost anything to read. So while speech production models probably to some extent could be tested on the present data, the inner workings of the brain are beyond reach, and as was shown, ERP studies of language pose formidable problems given the complex characteristics of the motor action associated with speech production. It should also be noted that perception and production studies present the experimenter with more or less opposite problems vis-à-vis control of either the stimuli or the end product. However, as was also shown, these problems are not insurmountable as is evidenced by the large number of ERP studies of phonology, syntax, semantics, prosody and so on. Disfluency should be just around the corner (methinks!).

6.4 Final comments

A friend of mine once said that "no one is ever the first to do anything".[1] Although in some sense obviously wrong, I would argue that there is some truth to it. Throughout this thesis, it has been a more or less implicit tenet of mine that there is much to learn from unexpected sources. Sometimes this is acknowledged. For example, it is striking that so much work

[1] "Ingen är nånsin först med nåt." Thanks, Sunk.

within computationally motivated research draws heavily on William James's observations from the 19th century, well before the computer was conceived in any detail. However, sometimes it is my impression that most of related research is not known simply because it has been carried out within another field: Also, although one might think that the difference in disfluency production in different settings, like face-to-face vs. telephone was "discovered" within application-motivated interface design, such differences were noted much earlier within psychotherapy studies (at least) already in 1965 (Kasl & Mahl)—well before any automatization of commercial services was the motivation, or even *could* have been, given current technological levels.

E. F. Schumacher (1977) differentiated between what he labeled as convergent versus divergent problems. **Convergent problems** are problems that tend to result in similar solutions, i.e., the more people who spend time in trying to solve the problem, and the more time they spend on the problem in question, the more the suggested solutions will resemble each other. This is typical within the natural sciences, e.g., how to create a heat-resistant compound. **Divergent problems** are, as you might have guessed, the opposite. The more people working on and the more time spent on a problem, the greater the number of different, even opposite, solutions will be suggested. This is more typical of the humanities, e.g., the interpretation of a character in one of Shakespeare's plays.

So, is disfluency a convergent or divergent problem? It would seem from the previous that the more people working on the problem, the larger the number of different explanations, models, theories and so on and so forth. On the other hand, despite all the different rationales, and the sundry backgrounds of the researchers, it is striking how similar a set of disfluencies they have come up with, with only minor differences in categories. This is even more remarkable given that much of the early work was obviously done with a total lack of knowledge of parallel work within other disciplines. I will waive a final verdict here, but content myself by pointing out that disfluency is indeed a multi-faceted phenomenon, ranging over a huge number of different disciplines and areas of interest. This is, in my view, what makes it so fascinating.

Although the main goal of this thesis has been to provide a detailed, structural, analysis of disfluencies in spoken Swedish (with a lot of constraints on the task, mode, channel, labeling and so on), there has been the implicit, larger, goal of broadening the horizons of the *phenomenon* of disfluency above and beyond what has been analyzed and discussed in detail here. Disfluency is so much more, and my (more or less implicit) mission has also been to try to raise the awareness of the multifaceted character of disfluency, whether this be your term of choice or not. Also, my own personal outlook on scientific method is basically zetetic, or perhaps, to use the words of Fodor (1983), more or less *isotropic*. I'll let Fodor himself provide the definition:

> By saying that [scientific] confirmation is isotropic, I mean that the facts relevant to the confirmation of a scientific hypothesis may be drawn from anywhere in the field of previously established empirical (or, of course, demonstrative) truths. Crudely: everything that the scientist knows is, in principle, relevant to determining what else he ought to believe. In principle, our botany constrains our astronomy, if only we could think of ways to make them connect. /…/ [T]hat is because of a profound conviction—partly metaphysical, partly epistemological—to which scientists implicitly subscribe: the world is a connected causal system *and we don't know how the connections are arranged*. (Fodor, 1983, p. 105; italics in original.)

Perhaps it is needless to say that I subscribe to the same view, and that I in this work have tried to provide as wide a horizon as possible from the disfluency viewpoint. It is also my view that everything in chapter two is indeed "connected" and that one of the outstanding quests within this field is to learn and understand exactly what these connections look like, something which is far from clear at the moment. If I have succeeded in drawing the attention of anyone engaged in disfluency research to a hitherto unknown field of research—for ideas, insights or simply inspiration—than at least half the work is done.

6.5 Signing off

Let me conclude with yet another passage taken from Chafe (1980):

> [H]esitation phenomena /…/ provide good evidence that speaking is not a matter of regurgitating material already stored in the mind in linguistic form, but that it is a creative art, relating two media, thought and language, which are not isomorphic but require adjustments and readjustments to each other. A speaker does not follow a clear, well traveled path, but must find his way through territory not traversed before, where pauses, changes of direction, and retracing of steps are quite to be expected. The fundamental reason for hesitating is that speech production is an act of creation. (Chafe, 1980, p. 170.)

Hear, hear!

Finis

References

Editorial remarks

I have tried to meet two different main criteria:

1. To provide as much information as possible, including such things as *full* first names of authors and editors, volume and page numbers, locations and full dates of conference proceedings, locations and publishers of books and so on.

2. Maintain consistency.

This has proven impossible to completely live up to. Below, I have listed some of the principles used in this set of references.

Names

I have tried to list full names of authors, instead of providing only initials. However, some people's first names *are* obviously *initials*, and even journals with a full-first-name policy give their names as initials only, even when co-authors of the same articles are given with full names. In some cases where only initials appear in one article, the full names are known from other work. In most such cases I have added the full name within square brackets. Also, authors' names sometimes appear in different forms in different works, e.g. with or without an extra initial, and so on. I have tried to make the names look the same in all included work. Also, there is at least one example of an author who has used two different spellings ("Lallgee" vs. "Lalljee").

Some authors have double surnames, e.g. "Bernstein Ratner", and it not always clear which of the names is the look-up name. In such cases I have included entries for both names, with cross-referencing to the name where I have listed the listed works. I extend my excuses to authors who would have preferred to be listed under the alternative name. Moreover, I have listed all names beginning with "van" under <v>, since there is not standard usage in this respect, and I am not familiar with individual preferences.

Titles

As regards title capitalization (initial word only, all words belonging to open word classes, or all words), I have tried to provide the titles as they appear in the original source. If original titles are given in capitals only, I have only capitalized the first—and obvious)—words.

Years

Some articles appear in identical form in two sources, e.g. as reprints of previously published articles in books. In those cases where the years are different, the sources are given with two years, e.g. "1971/1988", with both sources specified.

Some articles are "open peer" articles—notably from *Brain and Behavioral Sciences*—where the main author(s) is (are) given an article, which is then replied to by a number, which in turn is then replied to by the author(s), leading to a two-fold article. I have referred to those articles as e.g. "1986a/1986b, making a distinction between the opening and the closing article, mainly for the sake of clarity.

Page numbers

I have not been able to track down (original) page numbers for some of the articles, most often when I have downloaded the article in question from e.g. the author's homepage, where the bibliographical information oftentimes is not complete.

Editorial corrections

Some (silent) editorial corrections have been made when the original title contains typographical errors. Thus, *human-computer* (which should mean "cyborg"?) has been changed into *human–computer* (which is what is intended), and so on. Also, inch and foot signs have been changed into single and double quotes.

Abelin, Åsa & Jens Allwood. 1998/1999. *Jämförelse mellan OCM-kodningsstandard och Robert Eklunds disfluenskodningsstandard*. Project report, Department of Linguistics, Göteborg University, September 1, 1998. Also in: *Swedish Dialogue Systems (SDS). HSRF/NUTEK. A Platform for Multimodal Spoken Language Corpora*, Dec[ember] 31, 1999, Department of Linguistics, Göteborg University, Sweden. [No page numbers.]

Adams, Martin R. 1987. Voice onsets and segment durations of normal speakers and beginning stutterers. *Journal of Fluency Disorders*, vol. 12, pp. 133–139.

Adams, Martin R. 1982. Fluency, nonfluency, and stuttering in children. *Journal of Fluency Disorders*, vol. 7, pp. 171–185.

Adams, Martin R., Rosa Lee Sears & Peter R. Ramig. 1982. Vocal Changes in Stutterers and Nonstutterers During Monotoned Speech. *Journal of Fluency Disorders*, vol. 7, pp. 21–35.

Adams, Martin R. & Charles M. Runyan. 1981. Stuttering and fluency: Exclusive Events or Points on a Continuum. *Journal of Fluency Disorders*, vol. 6, pp. 197–218.

Adams, Martin R. & Peter R. Ramig. 1980. Vocal characteristics of normal speakers and stutterers during choral reading. *Journal of Speech and Hearing Research*, vol. 23, pp. 457–469.

Adams, Martin R. & Paul Hayden. 1976. The ability of stutterers and nonstutterers to initiate and terminate phonation during production of an isolated vowel. *Journal of Speech and Hearing Research*, vol. 19, pp. 290–296.

Adams, Martin R., Charles M. Runyan & A. R. Mallard. 1974. Airflow characteristics of the speech of stutterers and nonstutterers. *Journal of Fluency Disorders*, vol. 1, no. 2, pp. 4–12.

Adams, Martin R., Jeffrey I. Lewis & Thomas E. Besozzi. 1973. The Effect of Reduced Reading Rate on Stuttering Frequency. *Journal of Speech and Hearing Research*, vol. 16, pp. 671–675.

Agnäs, Marie-Susanne, Hiyan Alshawi, Ivan Bretan, David Carter, Ken Ceder, Michael Collins, Richard Crouch, Vassilios Digalakis, Barbro Ekholm, Björn Gambäck, Jaan Kaja, Jussi Karlgren, Bertil Lyberg, Patti Price, Stephen Pulman, Manny Rayner, Christer Samuelsson & Tomas Svensson. 1993. *Spoken Language Translator: First-Year Report*. Project report, November 1993, Telia Research AB and *Technical Report SRI Technical Report CRC-043*, SRI, Cambridge, England.

Aitchison, Jean. 1976/1993 (third edition). *The Articulate Mammal*. London: Routledge.

Akelaitis, Andrew J. 1944. A Study of Gnosis, Praxis and Language Following Section of the Corpus Callossum and Anterior Commissure. *Journal of Neurosurgery*, vol. 1, pp. 94–102.

Akins, Kathleen A. & Daniel C. Dennett. 1986. Who may I say is calling? *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 517–518.

Allwood, Jens. 1997a. Notes on Dialog and Cooperation. In: Kristiina Jokinen, David Sadek & David Traum (eds.), *Proceedings of the IJCAI '97 workshop "Collaboration, Cooperation and Conflict in Dialogue Systems"*, 25 August 1997, Nagoya, Japan, pp. 9–21.

Allwood, Jens. 1997b. Dialog as Collective Thinking. In: Paavo Pylkkänen & Pauli Pylkkö (eds.), *New Directions in Cognitive Science. Publications of the Finnish Artificial Intelligence Society. International Conferences*, no. 2, Helsinki, pp. 222–226. Also in: Paavo Pylkkänen, Pauli Pylkkö & Anti Hautamäki (eds.), *Brain, Mind & Physics*, Amsterdam: IOS Press, pp. 205–210.

Allwood, Jens. 1995. Reasons for Management in Spoken Dialogue. In: Robbert-Jan Beun, M. Baker & M. Reiner (eds.), *Dialogue and Instruction*, Springer-Verlag, pp. 241–250.

Allwood, Jens. 1994a. Om Dialogreglering. In: Nils Jörgenson, Christer Platzack & Jan Svensson (eds.), *Språkbruk, grammatik och språkförändring*. Department of Nordic Languages, University of Lund, [no pages numbers obtained].

Allwood, Jens. 1994b. Obligations and Options in Dialogue. In: *Think*, vol. 3, ITK Tilbury University, pp. 9–18.

Allwood, Jens. 1988a. The Structure of Dialog. In: Martin M. Taylor, Françoise Neél & Don G. Bouwhuis (eds.), *Structure of Multimodal Dialog II*, Amsterdam: John Benjamins, pp. 3–24.

Allwood, Jens. 1988b. Om det svenska systemet för språklig återkoppling. In: Per Linell, Viveca Adelswärd, Torbjörn Nilsson & Per A. Pettersson (eds.), *Svenskans Beskrivning* 16, vol. 1, Tema Kommunikation, University of Linköping, Sweden**,** pp. 89–106.

Allwood, Jens. 1977. A Critical Look at Speech Act Theory. In: Östen Dahl (ed.), *Logic, Pragmatics and Grammar*. Lund: Studentlitteratur, pp. 53–69.

Allwood, Jens. 1976. *Linguistic Communication as Action and Cooperation. A study in pragmatics*. PhD thesis, Göteborg University, Sweden.

Allwood, Jens, Leif Grönqvist, Elisabeth Ahlsén & Magnus Gunnarsson. 2002. In: J. van Kuppevelt (ed.), *Current and New Directions in Discourse and Dialogue*. Kluwer Academic Publishers.

Allwood, Jens, Leif Grönqvist, Elisabeth Ahlsén & Magnus Gunnarsson. 2001a. Annotations and Tools for an Activity Based Spoken Language Corpus. *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, 1–2 September 2001, Aalborg, Denmark, pp. 1–10.

Allwood, Jens, Elisabeth Ahlsén, Joakim Nivre & Staffan Larsson. 2001b. Own Communication Management. Coding Manual v1.0. In: Jens Allwood (ed.), Dialog Coding – Function and Grammar. Göteborg Coding Schemas, *Gothenburg Papers in Theoretical Linguistics 85*, Department of Linguistics, Göteborg University, pp. 45–52.

Allwood, Jens, Åsa Abelin & Leif Grönqvist. 1999. Kort beskrivning och jämförelse av transkriptions-system från Lund, Telia, och Göteborg. In: *Swedish Dialogue Systems (SDS). HSRF/NUTEK. A Platform for Multimodal Spoken Language Corpora*, Dec[ember] 31, 1999, Department of Linguistics, Göteborg University, Sweden. [No page numbers.]

Allwood, Jens & Maria Björnberg. 1999. Coding schemas within the SDS project – A comparison. In: *Swedish Dialogue Systems (SDS). HSRF/NUTEK. A Platform for Multimodal Spoken Language Corpora*, Dec[ember] 31, 1999, Department of Linguistics, Göteborg University, Sweden. [No page numbers.]

Allwood, Jens & Johan Hagman. 1994/1999. Some Simple Automatic Measures of Spoken Interaction. *SSKKII Reports* #6, Department of Linguistics, Göteborg University, Sweden. Also in: *Swedish Dialogue Systems (SDS). HSRF/NUTEK. A Platform for Multimodal Spoken Language Corpora*, Dec[ember] 31, 1999, Department of Linguistics, Göteborg University, Sweden. [No page numbers.]

Allwood, Jens & Björn Haglund. 1992. *Communicative Activity Analysis of a Wizard of Oz Experiment*. Internal Report, PLUS ESPRIT project P5254.

Allwood, Jens, Joakim Nivre & Elisabeth Ahlsén. 1992. On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics*, vol. 9, pt. 1, pp. 1–26.

Allwood, Jens, Joakim Nivre & Elisabeth Ahlsén. 1990. Speech Management—on the Non-written Life of Speech. *Nordic Journal of Linguistics*, vol. 13, no. 1, pp. 3–48.

Alm, Per. 1995. *Stamning*. Borås, Sweden: Natur och kultur.

Althoff, Frederek. 1997. *Ein Modul für den Einsatz morphologischen Wissens bei der Erkennung gesprochener Sprache*. Diplomarbeit, Universität Bielefeld, 29. Juli 1997.

Althoff, Frederek, Guido Drexel, Harald Lüngen, Martina Pampel & Christoph Schillo. 1996. The Treatment of Compounds in a Morphological Component for Speech Recognition. In: Dafydd Gibbon (ed.), *Natural Language and Speech Technology. Results of the 3rd KONVENS Conference*. Bielefeld, October 1996. Berlin: Mouton de Gruyter, pp. 71–76. Also as *Verbmobil Report*, 170.

Amadeus. **http://www.amadeus.net/**

Amalberti, René, Noëlle Carbonell & Pierre Falzon. 1993. User representations of computer systems in human–computer speech interaction. *International Journal of Man–Machine Studies*, vol. 38, pp. 547–566.

Andrews, Gavin, Ashley Craig, Anne-Marie Feyer, Susan Hoddinott, Pauline Howie & Megan Neilson. 1983. Stuttering: A review of research findings and theories circa 1982. *Journal of Speech and Hearing Research*, vol. 48, pp. 226–246.

Andrews, Gavin, Pauline M. Howie, Melinda Dozsa & Barry E. Guitar. 1982. Stuttering: Speech pattern characteristics under fluency-inducing conditions. *Journal of Speech and Hearing Research*, vol. 25, pp. 208–216.

Annett, Marian. 2003. Myths of first cause and asymmetries in human evolution. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 208–209.

Arbib, Michael A. 2003. Protosign and protospeech: An expanding spiral. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 209–210.

Arcadi, Adam Clark. 2003. Is gestural communication more sophisticated than vocal communication in wild chimpanzees? *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 210–211.

Armstrong, David F. 2003. Creative solution to an old problem. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 211–212.

Aschersleben, Gina & Wolfgang Prinz. 1995. Synchronizing actions with events: The role of sensory information. *Perception & Psychophysics*, vol. 57, no. 3, pp. 305–317.

Asp, Annika & Anna Decker. 2001a. Designing with speech acts to elude disfluency in human–computer dialogue systems. *Papers from Fonetik 2001, Örenäs, May 30 – June 1, 2001. Working Papers, no. 49, 2001*, Lund University, Sweden, pp. 2–5.

Asp, Annika & Anna Decker. 2001b. *Reducing disfluency through speech act design*. Bachelor's Degree Thesis, Department of Linguistics, Computational Linguistics, Stockholm University and Telia Research AB, Broadband Services, Farsta, Sweden.

Austin, J[ohn] L. 1962/1975. *How to do things with words*. Oxford: Clarendon Press.

Baars, Bernard J. (ed.). 1992a. *Experimental Slips and Human Error. Exploring the Architecture of Volition*. New York & London: Plenum Press.

Baars, Bernard J. 1992b. The Many Uses of Error. Twelve Steps to a Unified Framework. In: Bernard J. Baars (ed.), *Experimental Slips and Human Error*, New York & London: Plenum Press, ch. 1, pp. 3–34.

Baars, Bernard J. 1992c. A New Ideomotor Theory of Voluntary Control. In: Bernard J. Baars (ed.), *Experimental Slips and Human Error*, New York & London: Plenum Press, ch. 4, pp. 93–120.

Baars, Bernard J. 1992d. A Dozen Competing-Plans Techniques for Inducing Predictable Slips and Speech and Action. In: Bernard J. Baars (ed.), *Experimental Slips and Human Error*, New York & London: Plenum Press, ch. 6, pp. 129–150.

Baars, Bernard J. 1991. A curious coincidence? Consciousness as an object of scientific scrutiny fits our personal experience remarkably well. *Behavioral and Brain Sciences*, vol. 14, no. 4, pp. 669–670.

Baars, Bernard J. 1988. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, Bernard J., Michael T. Motley & Donald G. MacKay. 1975. Output Editing for Lexical Status in Artificially Elicited Slips of the Tongue. *Journal of Verbal Learning and Verbal Behavior*, vol. 14, pp. 382–391.

Baber, C., B. Mellor, R. Graham, J. M. Noyes & C. Tunley. 1996. Workload and the use of automatic speech recognition: The effects of time and resource demands. *Speech Communication*, vol. 20, pp. 37–53.

Bakker, Klaus & Gene J. Brutten. 1990. Speech-related reaction times of stutterers and nonstutterers. Diagnostic implications. *Journal of Speech and Hearing Disorders*, vol. 55, pp. 295–299.

Banks, William P. 2002. On Timing Relations between Brain and World. *Consciousness and Cognition*, vol. 11, pp. 141–143.

Barash, Charles T., Barry Guitar, Rebecca J. McCauley & Richard G. Absher. 2000. Disfluency and Time Perception. *Journal of Speech, Language, and Hearing Research*, vol. 43, pp. 1429–1439.

Barber, Victoria. 1940. Studies in the psychology of stuttering, XVI. Rhythm as a Distraction in Stuttering. *Journal of Speech Disorders*, vol. 5, pp. 29–42.

Bard, Ellen Gurman & Robin Lickley. 1998b. Disfluency Deafness: Graceful Failure in the Recognition of Running Speech. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, University of Wisconsin, Madison, Minnesota, USA, Lawrence Erlbaum, pp. 108–113.

Bard, Ellen Gurman & Robin Lickley. 1997. On Not Remembering Disfluencies. *Proceedings of Eurospeech '97*, 22–25 September 1997, Rhodes, Greece, vol. 5, pp. 2855–2858.

Bargh, John A. & Tanya L. Chartrand. 1999. The Unbearable Automaticity of Being. *American Psychologist*, vol. 54, pp. 462–479.

Baron, Don, Elizabeth Shriberg & Andreas Stolcke. 2002. Automatic Punctuation and Disfluency Detection in Multi-Party Meetings Using Prosodic and Lexical Cues. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2002*, 16–20 September 2002, Denver, Colorado, USA, vol. 2, pp. 949–952.

Baum, L. Frank. 1900. *The Wonderful Wizard of Oz*. New York: George M. Hill Co.

Baumeister, Roy F. 1984. Choking Under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance. *Journal of Personality and Social Psychology*, vol. 46, no. 3, pp. 610–620.

Bavelier, Daphne, David P. Corina & Helen J. Neville. 1998a. Brain and Language: a Perspective from Sign Language. *Neuron*, vol. 21, pp. 275–278.

Bavelier, Daphne, David Corina, Peter Jezzard, Vince Clark, Avi Karni, Anil Lalwani, Josef P. Rauschecker, Allen Braun, Robert Turner & Helen J. Neville. 1998b. Hemispheric specialization for English and ASL: left invariance–right variability. *Neuroreport*, vol. 9, pp. 1537–1542.

Bear, John, John Dowding & Elizabeth E. Shriberg. 1992. Integrating multiple knowledge sources for detection and correction of repairs in human–computer dialog. *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 1992*, 28 June – 2 July 1992, Newark, Delaware, USA, pp. 56–63.

Beaton, Alan A. 2003. Going for Broca? I wouldn't bet on it! *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 212–213.

Beattie, Geoffrey W. & B[rian] L. Butterworth. 1979. Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, vol. 22, part 3, pp. 201–211.

Becker, Wolfgang, Ottomar Hoehne, Katsuhiko Iwase & Hans H. Kornhuber. 1972. *Vision Research*, vol. 12, pp. 421–436.

Becket, Ralph, Pierrette Bouillon, Harry Bratt, Ivan Bretan, David Carter, Vassilis Digalakis, Robert Eklund, Horacio Franco, Jaan Kaja, Martin Keegan, Ian Lewin, Bertil Lyberg, David Milward, Leonardo Neumeyer, Patti Price, Manny Rayner, Per Sautermeister, Fuliang Weng & Mats Wirén. 1997. *Spoken Language Translator: Phase Two Report*. Internal Report, Telia Research and SRI International (SRI Project 6393).

Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2004. The original ToBI system and the evaluation of the evolution of the ToBI framework. In: Sun-Ah Jun (ed.), *Prosodic Typology – The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, ch. 2, pp. 9–54.

Beckman, Mary E. & Julia Hirschberg. 1994. *The ToBi Annotation Conventions*.
http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html

Beckman, Mary E. & Gayle Ayers Elam. 1993/1997. *Guidelines for ToBI Labelling*. Version 3, March 1997.
http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/

Beilock, Sian L. & Thomas H. Carr. 2001. On the Fragility of Skilled Performance: What Governs Choking Under Pressure? *Journal of Experimental Psychology: General*, vol. 130, no. 4, pp. 701–725.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America (JASA)*, vol. 113, no. 2, pp. 1001–1024.

Bell, Linda, Johan Boye & Joakim Gustafson. 2001. Real-Time Handling of Fragmented Utterances. *Proceedings of the NAACL 2001 Workshop on Adaptation in Dialogue Systems*, 4 June 2001, Pittsburgh, Pennsylvania, USA. [No page numbers, CD-ROM proceedings only.]

Bell, Linda, Johan Boye, Joakim Gustafson & Mats Wirén. 2000. Modality Convergence in a Multimodal Dialogue System. *Proceeding Götalog 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue*, 15–17 June 2000, Göteborg University, Sweden, pp. 29–34.

Bell, Linda, Robert Eklund & Joakim Gustafson. 2000. A Comparison of Disfluency Distribution in a Unimodal and a Multimodal Human–Machine Interface *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2000*, 16–20 October 2000, Beijing, China, vol. 3, pp. 626–629.

Bentall, R. P. & P. D. Slade. 1986. Verbal hallucinations, unintendedness, and the validity of the schizophrenia diagnosis. *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 519–520.

Bentall, R. P. & P. D. Slade. 1985. Reality testing and auditory hallucinations: A signal detection analysis. *British Journal of Clinical Psychology*, vol. 24, pp. 159–169.

Bentin, Shlomo, Marta Kutas & Steven A. Hillyard. 1993. Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, vol. 30, pp. 161–169.

Berg, Thomas. 1992. Productive and perceptual constraints on speech-error detection. *Psychological Research*, vol. 54, pp. 114–126.

Berg, Thomas. 1986a. The problems of language control: Editing, monitoring, and feedback. *Psychological Research*, vol. 48, pp. 133–144.

Berg, Thomas. 1986b. The aftermath of error occurrence: Psycholinguistic evidence from cut-offs. *Language & Communication*, vol. 6, no. 3, pp. 195–213.

Berman, Bob. 2004. Space: A Very Noisy Place. *Discover*, February 2004, vol. 25, no. 2, p. 30.

Bernstein, Basil. 1962. Linguistic codes, hesitation phenomena and intelligence. *Language and Speech*, vol. 5, pt. 1, pp. 31–46.

Bernstein, Nan Ratner. See: Ratner, Nan Bernstein.

Berthold, André & Anthony Jameson. 1999. Interpreting Symptoms of Cognitive Load in Speech Input. In: J. Kay (ed.), *User Modeling: Proceedings of the Seventh International Conference on User Modeling, UM99*, 20–24 June 1999, Banff, Canada. Vienna. New York: Springer, pp. 235–244.

Besson, Mireille, Frederique Faita, Claire Czternasty & Marta Kutas. 1997. What's in a pause: event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, vol. 46, pp. 3–23.

Beun, R[obbert].-J[an]. & H[arry]. C. Bunt. 1987. Investigating linguistic behaviour in information dialogues with a computer. *I.P.O. Annual Progress Report* 22, pp. 77–86.

Black, John W. 1951. The Effect of Delayed Side-Tone Upon Vocal Rate and Intensity. *Journal of Speech and Hearing Disorders*, vol 16, pp. 56–60.

Blackmer, Elizabeth R. & Janet L. Mitton. 1991. Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, vol. 39, pp. 173–194.

Blakemore, Sarah-Jayne & Chris Frith. 2003. Self-awareness and action. *Current Opinion in Neurobiology*. vol. 13, pt. 2, pp. 219–224.

Blakemore, S[arah].-J[ayne]., D. A. Oakley & C[hris]. D. Frith. 2003. Delusions of alien control in the normal brain. *Neuropsychologia*, vol. 41, pt. 8, pp. 1058–1067.

Blankenship, Jane. 1964. "Stuttering" In Normal Speech. *Journal of Speech and Hearing Research*, vol. 7, no. 1, pp. 95–96.

Blankenship, Jane & Christian Kay. 1964. Hesitation phenomena in English Speech: A Study in Distribution. *Word*, vol. 20, pp. 360–373.

Blass, Thomas & Aron W. Siegman. 1975. A psycholinguistic comparison of speech, dictation and writing. *Language and Speech*, vol. 18, pp. 20–34.

Bloch, Bernard. 1946. Studies in colloquial Japanese II. Syntax. *Language*, vol. 22, pp. 200–248.

Block, Ned. 1992. Begging the question against phenomenal consciousness. *Behavioral and Brain Sciences*, vol. 15, pp. 205–206.

Blood, Gordon W., Ingrid M. Blood & Stephen B. Hood. 1987. The development of ear preferences in stuttering and nonstuttering children: a longitudinal study. *Journal of Fluency Disorders*, vol. 12, pp. 119–131.

Bloodstein, Oliver. 1969/1987 (4th edition). *A Handbook on Stuttering*. Chicago: National Easter Seal Society.

Bloodstein, Oliver. 1950. A Rating Scale Study Of Conditions Under Which Stuttering Is Reduced Or Absent. *Journal of Speech and Hearing Disorders*, vol. 15, pp. 29–36.

Bloodstein, Oliver, Janice P. Alper & Paulette Kendler Zisk. 1965. Stuttering as an Outgrowth of Normal Disfluency. In: Dominick A. Barbara (ed.), *New Directions in Stuttering*, 31, Springfield, Illinois: Charles E. Thomas, pp. 31–54.

Bock, [J.] Kathryn. 1987. Exploring levels of processing in sentence production. In: Gerard Kempen (ed.), *Natural Language Generation: New Results in Artificial Intelligence, Psychology and Linguistics*. Dordrecht: Kluwer Academic Publishers, ch. 22, pp. 351–363.

Bock, J. Kathryn. 1982. Towards a Cognitive Psychology of Syntax: Information Processing Contributions to Sentence Formulation. *Psychological Review*, vol. 89, no. 1, pp. 1–48.

Boddy, John & Hal Weinberg. 1981. Brain potentials, perceptual mechanisms and semantic categorisation. *Biological Psychology*, vol. 12, pp. 43–61.

Boehmler, Richard M. 1958. Listener Responses to Non-Fluencies. *Journal of Speech and Hearing Research*, vol. 1, pp. 132–141.

Boehmler, R[ichard]. M. & S. I. Boehmler. 1989. The Cause of Stuttering: What's the Question? *Journal of Fluency Disorders*, vol. 14, pp. 447–450.

Bogen, Joseph E. 1969. The other side of the brain II. An appositional mind. *Bulletin of the Los Angeles Neurological Society*, vol. 34, pt. 3, pp. 135–162.

Bogen, Joseph E., E. D. Fischer & P. J. Vogel. 1965. Cerebral Commissurotomy: A Second Case Report. *Journal of the American Medical Association (JAMA)*, vol. 194, no. 12, pp. 1328–1329.

Bogen, Joseph E. & Philip J. Vogel. 1962. Cerebral Commissurotomy in Man: Preliminary Case Report. *Bulletin of the Los Angeles Neurological Society*, vol. 27, p. 169–172.

Bolbecker, Amanda R., Zixi Cheng, Gary Felstein, King-Leung Kong, Corrinne C. M. Lim, Sheryl J. Nisly-Nagele, Lolin T. Wang-Bennett & Gerald S. Wasserman. 2002. Two Asymmetries Governing Mental and Neural Timing. *Consciousness and Cognition*, vol. 11, pp. 265–272.

Bolinger, Dwight. 1983. Where does intonation belong? *Journal of Semantics*, vol.2, no. 2, pp. 101–120.

Bond, Z[inny]. 1973. Perceptual Errors in Ordinary Speech. *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft*, vol. 26, pp. 691–695.

Bond, Z[inny] S. & L[arry]. H. Small. 1984. Detecting and correcting mispronunciations: a note on methodology. *Journal of Phonetics*, vol. 12, pp. 279–283.

Bond, Z[inny] S. & L[arry]. H. Small. 1983. Voicing, vowel, and stress mispronunciations in continuous speech. *Perception and Psychophysics*, vol. 5, pp. 470–474.

Boomer, Donald S. 1970. Review of Goldman-Eisler: Psycholinguistics: experiments in spontaneous speech. *Lingua*, vol. 25, pp. 152–164.

Boomer, Donald S. 1965. Hesitation and grammatical encoding. *Language and Speech*, vol. 8, pp. 148–158.

Boomer, Donald S. 1963. Speech disturbance and body movement in interviews. *Journal of Nervous and Mental Disease*, vol. 136, pp. 263–266.

Boomer, Donald S. & John D. M. Laver. 1968/1973. Slips of the tongue. In: Victoria A. Fromkin (ed.), *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 120–131. Originally published in 1968 in *British Journal of Disorders of Communication*, vol. 3, no. 1, pp. 2–12.

Boomer, Donald S. & Allen T. Dittman. 1964. Speech rate, filled pause, and body movement in interviews. *Journal of Nervous and Mental Disease*, vol. 139, pp. 324–327.

Boomer, Donald S. & Allen T. Dittman. 1963. Hesitation pauses and juncture pauses in speech. *Language and Speech*, vol. 5, p. 215–220.

Boomer, Donald S. & D. Wells Goodrich. 1961. Speech disturbances and judged anxiety. *Journal of Consulting Psychology*, vol. 25,no. 2, p. 160.

Borden, Gloria J. 1979. An Interpretation of Research on Feedback Interruption in Speech. *Brain and Language*, vol. 7, pp. 307–319.

Borden, Gloria J., Thomas Baer & Mary Kay Kenney. 1985. Onset of voicing in stuttered and fluent utterances. *Journal of Speech and Hearing Research*, vol. 28, pp. 363–372.

Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober & Susan E. Brennan. 2001. Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech,* vol. 44, no. 2, pp. 123–147.

Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober & Susan E. Brennan. 1999. Which speakers are most disfluent in conversation, and when? *Proceedings of Disfluency in Spontaneous Speech Workshop*, 1 July 1999, Berkeley, California, USA, pp. 7–10.

Boschert, Jürgen & Lüder Deecke. 1986. Cerebral potentials preceding voluntary toe, knee and hip movements and their vectors in human precentral gyrus. *Brain Research*, vol. 376, pp. 175–179.

Boschert, J[ürgen]., R. F. Hink & L[üder]. Deecke. Finger Movement Versus Toe Movement-Related Potentials: Further Evidence for Supplementary Motor Area (SMA) Participation Prior to Voluntary Action. *Experimental Brain Research*, vol. 52, pp. 73–80.

Bradshaw, John L. 2003. Gesture in language evolution: Could I but raise my hand to it! *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 213–214.

Brady, John Paul. 1969. Studies on the metronome effect on stuttering. *Behavior Research and Therapy*, vol. 7, pp. 197–204.

Brain, Lord. 1963. Some reflections on brain and mind. *Brain*, vol. 86, pt. 3, pp. 381–402.

Brand, Myles. 1986. Indended versus intentional action. *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 520–521.

Branigan, Holly, Robin Lickley & David McKelvie. 1999. Non-linguistic influences on rates of disfluency in spontaneous speech. *Proceedings of the International Congress of Phonetic Sciences (ICPhS) '99*, 1–7 August 1999, San Francisco, California, USA, vol. 1, pp. 387–390.

Branscom, Margaret E., Jeannette Hughes & Eloise Tupper Oxtoby. 1955. Studies of Nonfluency in the Speech of Preschool Children. In: Wendell Johnson (ed.), *Stuttering in Children and Adults. Thirty Years of Research at the University of Iowa*, Minneapolis: University of Minneapolis Press, ch. 5, pp. 157–180.

Brasil-Neto, Joaquim P., Alvaro Pascual-Leone, Josep Valls-Solé, Leonardo G. Cohen & Mark Hallett. 1992. Focal transcranial magnetic stimulation and response bias in a forced-choice task. *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 55, pp. 964–966.

Brayton, Evelyn R. & Edward G. Conture. 1978. Effects of noise and rhythmic stimulation on the speech of stutterers. *Journal of Speech and Hearing Research*, vol. 21, pp. 285–294.

Breitenstein, Caterina, Agnes Floel, Bianca Dräger & Stefan Knecht. 2003. Lateralisation may be a side issue for understanding language development. *Behavioral and Brain Sciences*, vol. 26, no. 2, p. 214.

Breitmeyer, Bruno G. 2002. In Support of Pockett's Critique of Libet's Studies of the Time Course of Consciousness. *Consciousness and Cognition*, vol. 11, pp. 280–283.

Breitmeyer, Bruno G. 1985. Problems with the psychophysics of intention. *Behavioral and Brain Sciences*, vol. 8, pp. 539–540.

Brennan, Susan E. 2000. Processes that Shape Conversation and their Implications for Computational Linguistics. *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, 3–6 October 2000, Hongkong, China. [No page numbers obtained.]

Brennan, Susan E. & Michael F. Schober. 2001. How Listeners Compensate for Disfluencies in Spontaneous Speech. *Journal of Memory and Language*, vol. 44, pp. 274–296.

Brenner, Malcolm, E. Thomas Doherty & Thomas Shipp. 1994. Speech Measures Indicating Workload Demand. *Aviation, Space, and Environmental Medicine*, January 1994, pp. 21–26.

Bretan, Ivan, Robert Eklund, Jaan Kaja, Catriona MacDermid, Manny Rayner & David Carter. 2000. Corpora and Data Collection. In: Manny Rayner, Dave Carter, Pierrette Bouillon, Vassilis Digalakis & Mats Wirén (eds.), *The Spoken Language Translator*, Cambridge, Cambridge University Press, ch. 8, pp. 131–144.

Bretan, Ivan, Robert Eklund & Catriona MacDermid. 1996. Approaches to Gathering Realistic Training Data for Speech Translation Systems. *Proceedings of IVTTA—1996 IEEE Third Workshop, Interactive Voice Technology for Telecommunications Applications*, 30 September–1 October 1996, Basking Ridge, New Jersey, USA, pp. 97–100.

Bridgeman, Bruce. 1985. Free will and the functions of consciousness. *Behavioral and Brain Sciences*, vol. 8, p. 550.

Broen, Patricia A. & Gerald M. Siegel. 1972. Variations in normal speech disfluencies. *Language and Speech*, vol. 15, pt. 3, pp. 219–231.

Browman, Catherine P. 1978. *Tip of the Tongue and Slip of the Ear. Implications for Language Processing*. UCLA Working Papers in Phonetics, July 1978, University of California, Los Angeles, USA.

Brown, C. J., G. N. Zimmerman, R. N. Linville & J. P. Hegmann. 1990. Variations in self-paced behaviors in stutterers and nonstutterers. *Journal of Speech and Hearing Research*, vol. 33, pp. 317–323.

Brown, Roger. 1973. Schizophrenia, language, and reality. *American Psychologist*, vol. 28, pp. 395–403.

Brown, Roger & David MacNeill. 1966. The "Tip of the Tongue" Phenomenon. *Journal of Verbal Learning and Verbal Behavior*, vol. 5, pp. 325–337.

Brown, Spencer F. 1945. The Loci of Stutterings In The Speech Sequence. *Journal of Speech Disorders*, vol. 10, no. 3, pp. 181–192.

Brown, Spencer F. 1937. The influence of grammatical function on the incidence of stuttering. *Journal of Speech Disorders*, vol. 2, no. 3, pp. 207–215.

Brunia, C. H. M. What is Wrong with Legs in Motor Preparation? In: H[ans]. H. Kornhuber & L[üder]. Deecke (eds.), *Motivation, motor and sensory processes: electrical potentials, behaviour and clinical use. Prog. Brain Research*, Amsterdam: Elsevier, vol. 54, pp. 232–236.

Brutten, Eugene J. 1963. Palmar Sweat Investigation of Disfluency and Expectancy Adaptation. *Journal of Speech and Hearing Research*, vol. 6, pp. 40–48.

Brutten, Gene J. & Alice C. Trotter. 1986. A dual-task investigation of young stutterers and nonstutterers. *Journal of Fluency Disorders*, vol. 11, pp. 275–284.

Burian, K., G. F. Gestring & M. Haider. 1969. EEG-Computer-Analyse — Sinnloser und Sinnvoller Akustischer Reize. *Acta oto-laryngologica*, vol. 67, pp. 333–340.

Butterworth, Brian. 1981. Speech errors: old data in search of new theories. *Linguistics*, vol. 19, pp. 627–662.

Butterworth, Brian. 1980. Evidence from pauses in speech. In: Brian Butterworth (ed.), *Language Production: vol. 1. Speech and Talk*. London: Academic Press, ch. 7, pp. 155–176.

Butterworth, Brian. 1975. Hesitation and Semantic Planning in Speech. *Journal of Psycholinguistic Research*, vol. 4, no. 1, pp. 75–87.

Butterworth, Brian & Geoffrey Beattie. 1978. Gesture and silence as indicators of planning in speech. In: Robin N. Campbell & Philip T. Smith (eds.), *Recent advances in the psychology of language. Formal and Experimental Approaches*. New York and London: Plenum Press, pp. 347–360.

Cairns, Douglas A. & John H. L. Hansen. 1994. Nonlinear analysis and classification of speech under stressed conditions. *Journal of the Acoustical Society of America (JASA)*, vol. 96, no. 6, pp. 3392–3400.

Callaway, Enoch & Peter R. Harris. 1974. Coupling between Cortical Potentials from Different Areas. *Science*, vol. 183, pp. 873–875.

Caramazza, Alfonso & Michele Miozzo. 1997. The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*, vol. 64, pp. 309–343.

Carletta, Jean, Amy Isard, Stephen Isard, Jacqueline C. Kowtko, Gwyneth Doherty-Sneddon & Anne H. Anderson. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, vol. 23, pp. 13–31.

Carver, Charles S. & Michael F. Scheier. 1978. Self-Focusing Effects of Dispositional Self-Consciousness, Mirror Presence, and Audience Presence. *Journal of Personality and Social Psychology*, vol. 36, no. 3, pp. 324–332.

Cecconi, Christine P., Stephen B. Hood & Raymond K. Tucker. 1977. Influence of reading level difficulty on the disfluencies of normal children. *Journal of Speech and Hearing Research*, vol. 20, pp. 475–484.

Chafe, Wallace [L]. 1994. *Discourse, Consciousness, And Time*. Chicago & London: University of Chicago Press.

Chafe, Wallace L. 1980. Some reasons for hesitating. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 169–180.

Chaika, Elaine O. 1977. Schizophrenic Speech, Slips of the Tongue, and Jargonaphasia: A Reply to Fromkin and to Lecours and Vanier-Clément. *Brain and Language*, vol. 4, pp. 464–475.

Chaika, Elaine [O.]. 1974. A Linguist Looks at "Schizophrenic" Language. *Brain and Language*, vol. 1, pp. 257–276.

Chaney, Carolyn F. 1969. Loci of disfluencies in the speech of nonstutterers. *Journal of Speech and Hearing Research*, vol. 12, no. 3, pp. 667–668.

Chapanis, Alphonse. 1981. Interactive human communication: some lessons learned from laboratory experiments. In: B. Shackel (ed.), *Man–Computer Interaction: Human Factors Aspects of Computers & People*. Rockville, Maryland: Sijthoff and Noordhoff, pp. 65–114.

Chapanis, Alphonse. 1975. Interactive human communication. *Scientific American*, vol. 232, no. 3, pp. 36–42.

Chapanis, Alphonse. 1973. The communication of factual information through various channels. *Information Storage and Retrieval*, vol. 9, pp. 215–231.

Chapanis, Alphonse. 1971. Prelude to 2001: Explorations in human communication. *American Psychologist*, vol. 26, pp. 949–961.

Chapanis, Alphonse, Robert N. Parrish, Robert B. Ochsman & Gerald D. Weeks. 1977. Studies in Interactive Communication: II. The Effects of Four Communication Modes on the Linguistic Performance of Teams during Cooperative Problem Solving. *Human Factors*, vol. 19, no. 2, pp. 101–126.

Chapanis, Alphonse & Charles M. Overbey. 1974. Studies in interactive communication: III. Effects of similar and dissimilar communication channels and two interchange options on team problem solving. *Perceptual and Motor Skills*, vol. 38 (Monograph Supplement), pp. 343–374.

Chapanis, Alphonse, Robert B. Ochsman, Robert N. Parrish & Gerald D. Weeks. 1972. Studies in Interactive Communication: I. The Effects of Four Communication Modes on the Behavior of Teams During Cooperative Problem-Solving. *Human Factors*, vol. 14, no. 6, pp. 487–509.

Chapman, James. 1966. The Early Symptoms of Schizophrenia. *British Journal of Psychiatry*, vol. 112, pp. 225–251.

Cherry, E. Colin. 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustical Society of America (JASA)*, vol. 25, no. 5, pp. 975–979.

Cherry, E. Colin, B. McA. Sayers & Pauline M. Marland. 1955. Experiments on the complete suppression of stammering. *Nature*, vol. 176, pp. 874–875.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The MIT Press.

Christenfeld, Nicholas. 1996. Effects of a Metronome on the Filled Pauses of Fluent Speakers. *Journal of Speech and Hearing Research*, vol. 39, pp. 1232–1238.

Christenfeld, Nicholas. 1995. Does It Hurt To Say Um? *Journal of Nonverbal Behavior*, vol. 19, no. 3, pp. 171–186.

Christenfeld, Nicholas. 1994. Options and Ums. *Journal of Language and Social Psychology*, vol. 13, no. 2, pp. 192–199.

Christenfeld, Nicholas & Beth Creager. 1996. Anxiety, Alcohol, Aphasia, and *Ums*. *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 451–460.

Churchland, Patricia Smith. 2002. *Brain-Wise. Studies in Neurophilosophy*. Cambridge, Massachusetts: The MIT Press.

Churchland, Patricia Smith. 1981. On the alleged backwards referral of experiences and its relevance to the mind–body problem. *Philosophy of Science*, vol. 48, pp. 165–181.

Chwilla, Dorothee J., Herman H. J. Kolk & Gijsbertus Mulder. 2000. Mediated Priming in the Lexical Decision Task: Evidence from Event-Related Potentials and Reaction Times. *Journal of Memory and Language*, vol. 42, pp. 314–341.

Clark, Herbert H. *Using language*. 1996. Cambridge, Cambridge University Press.

Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speech. *Cognition*, vol. 84, pp. 73–111.

Clark, Herbert H. & Thomas Wasow. 1998. Repeating Words in Spontaneous Speech. *Cognitive Psychology*, vol. 37, pp. 201–242.

Clarke, Arthur C. 1968. *2001: A Space Odyssey*. New York: The New American Library.

Claxton, Guy. 1999. Whodunnit? *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 99–113. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 99–113.

Clemmer, Edward J. 1980. Psycholinguistic Aspects of Pauses and Temporal Patterns in Schizophrenic Speech. *Journal of Psycholinguistic Research*, vol. 9, no. 2, pp. 161–185.

Code, Chris. 2003. Vocalisation and the development of hand preference. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 215–216.

Cohen, Anthony. 1968/1973. Errors of Speech and their Implication for Understanding the Strategy of Language Users. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, vol. 21, pp. 177–181. Republished in: Victoria A. Fromkin (ed.). 1973. *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 88–92.

Cohen, Jacob. 1960. A coeffiecient of agreement for nominal scales. *Educational and Psychological Measurement*, vol. XX, pp. 37–46.

Cohen, Philip R. 1984. The Pragmatics of Referring and the Modality of Communication. *Computational Linguistics*, vol. 10, no. 2, pp. 97–146.

Colburn, Norma. 1985. Clustering of disfluency in nonstuttering children's early utterances. *Journal of Fluency Disorders*, vol. 10, pp. 51–58.

Colcord, Roger D. & Martin R. Adams 1979. Voicing durations and vocal SPL changes associated with stuttering reduction during singing. *Journal of Speech and Hearing Research*, vol 22, pp. 468–479.

Cole, Ronald A. 1973. Listening for mispronunciations: A measure of what we hear during speech. *Perception and Psychophysics*, vol. 1, pp. 153–156.

Coles, Michael G. H. 1988. Modern Mind-Brain Reading: Psychophysiology, Physiology, and Cognition. *Psychophysiology*, vol. 26, no. 3, pp. 251–269.

Coles, L. Stephen. 1969. Talking with a robot in English. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) '69*, 7–9 May 1969, Washington, D.C., USA, pp. 587–596.

Connolly, J. F., S. H. Stewart & N. A. Phillips. 1990. The Effects of Processing Requirements on Neurophysiological Responses to Spoken Sentences. *Brain and Language*, vol. 39, pp. 302–318.

Connolly, J. F., N. A. Phillips, S. H. Stewart & W. G. Brake. 1992. Event-Related Potential Sensitivity to Acoustic and Semantic Properties of Terminal Words in Sentences. *Brain and Language*, vol. 43, pp. 1–18.

Conture, Edward G. 1990. Childhood stuttering: what is it and who does it? *ASHA Reports* (American Speech–Language–Hearing Association), no. 18, ch. 1, pp. 2–14.

Conture, Edward G., Raymond H. Colton & John R. Gleason. 1988. Selected temporal aspects of coordination during fluent speech of young stutterers. *Journal of Speech and Hearing Research*, vol. 31, pp. 640–653.

Cook, Mark. 1971. The incidence of filled pauses in relation to part of speech. *Language and Speech*, vol. 14, part 2, pp. 135–139.

Cook, Mark. 1969a. Transition probabilities and the incidence of filled pauses. *Psychonomic Science*, vol. 16, pp. 191–192.

Cook, Mark. 1969b. Anxiety, Speech Disturbances, and Speech Rate. *British Journal of Social and Clinical Psychology*, vol. 8, pp. 13–21.

Cook, Mark, Jacqueline Smith & Mansur G. Lalljee. 1974. Filled pauses and syntactic complexity. *Language and Speech*, vol. 17, no. 1, pp. 11–16.

Cook, Norman D. 2003. Hemispheric dominance has its origins in the control of the midline organs of speech. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 216–217.

Copeland, Jack. 1993. *Artificial Intelligence. A Philosophical Introduction*. Oxford: Blackwell.

Corballis, Michael C. 2003a/2003b. From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 199–260. Includes: Corballis, Michael C. 2003b. Hand-to-hand combat, or mouth-to-mouth resuscitation, pp. 242–260.

Corbetta, Daniela. 2003. Right-handedness may have come first: Evidence from studies in human infants and nonhuman primates. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 217–218.

Cordes, Anne K. 2000. Individual and Consensus Judgments of Disfluency Types in the Speech of Persons Who Stutter. *Journal of Speech, Language and Hearing Research*, vol. 43, no. 4, pp. 951–964.

Cordes, Anne K. & Roger J. Ingham. 1995. Stuttering Includes Both Within-Word and Between-Word Disfluencies. *Journal of Speech and Hearing Research*, vol. 38, pp. 382–386.

Core, Mark G. 1999. *Dialog Parsing: from Speech Repairs to Speech Acts*. PhD thesis, Department of Computer Science, University of Rochester, New York.

Core, Mark G. 1996. Using Parsed Corpora for Structural Disambiguation in the TRAINS domain. *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL) '97*, 24–17 June 1996, University of California, Santa Cruz, California, USA, pp. 345–347.

Core, Mark G. & Lenhart K. Schubert 1999a. A Syntactic Framework for Speech Repairs and Other Disruptions. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 20–26 June 1999, College Park, Maryland, USA, pp. 412–420.

Core, Mark G. & Lenhart K. Schubert. 1999b. Speech repairs: a parsing perspective. *Proceedings of Disfluency in Spontaneous Speech Workshop*, 1 July 1999, Berkeley, California, USA, pp. 47–50.

Core, Mark G. & Lenhart K. Schubert. 1999c. A Model of Speech Repairs and Other Disruptions. *Proceedings of AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, 5–7 November 1999, Cape Cod, Massachusetts, USA, pp. 48–53.

Core, Mark G. & Lenhart K. Schubert. 1998. Implementing Parser Metarules that Handle Speech Repairs and Other Disruptions. In: Diane Cook (ed.), *Proceedings 11th Internatinal FLAIRS Conference*, 17–20 May 1998, Sanibel Island, Florida, USA, pp. 283–288.

Core, Mark G. & Lenhart K. Schubert. 1997. Handling Speech Repairs and Other Disruptions Through Parser Metarules. *Proceedings AAAI Spring Symposium on Computational Models for Mixed Initiative Interaction*, 24–26 March 1997, Stanford, California, USA, pp. 23–29.

Cosmides, Leda, John Tooby, Helena Cronin & Oliver Curry (eds.). In press. *What Is Evolutionary Psychology: Explaining the New Science of the Mind (Darwinism Today)*.

Cosmides, Leda, John Tooby & Jerome H. Barkow. 1992. Introduction: Evolutionary Psychology and Conceptual Integration. In: Jerome H. Barkow, Leda Cosmides & John Tooby (eds.), *The Adapted Mind*. New York and Oxford: Oxford University Press, pp. 3–15.

Covington, Virginia C. 1964/1973 (revised version). Juncture in American Sign Language. *Sign Language Studies*, vol. 2, pp. 29–38.

Cowan, J. Milton & Bernard Bloch. 1948. An experimental study of pause in English grammar. *American Speech*, vol. 23, pp. 89–99.

Creutzfeldt, O., G. Ojemann & E. Lettich. 1989. Neural activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Experimental Brain Research*, vol. 77, pp. 476–489.

Crick, Francis & Christof Koch. 1990. Towards a neurobiological theory of consciousness. In: A[ntonio]. R. Damasio (ed.), *Seminars in the Neurosciences*, vol. 2, pp. 263–273.

Cross, Douglas E. & Harold L. Luper. 1983. Relation between finger reaction time and voice reaction time in stuttering and nonstuttering children and adults. *Journal of Speech and Hearing Research*, vol. 26, pp. 356–361.

Cross, Douglas E. & Harold L. Luper. 1979. Voice Reaction Time of Stuttering and Nonstuttering Children and Adults. *Journal of Fluency Disorders*, vol. 4, pp. 59–77.

Cross, Douglas E., Barbara B. Shadden & Harold L. Luper. 1979. Effects of Stimulus Ear Presentation on the Voice Reaction Time of Adult Stutterers and Nonstutterers. *Journal of Fluency Disorders*, vol. 4, pp. 45–58.

Crow, T[imothy]. J. 2000. Invited commentary on: Functional anatomy of verbal fluency in people with schizophrenia and those at genetic risk. *British Journal of Psychiatry*, vol. 176, pp. 61–63.

Culatta, Richard & Linda Leeper. 1988. Dysfluency isn't always stuttering. *Journal of Speech and Hearing Disorders*, vol. 53, pp. 486–488.

Cullinan, Walter L. & Mark T. Springer. 1980. Voice initiation and termination times in stuttering and nonstuttering children. *Journal of Speech and Hearing Research*, vol. 23, pp. 344–360.

Curlee, Richard F. 1981. Observer agreement on disfluency and stuttering. *Journal of Speech and Hearing Research*, vol. 24, pp. 595–600.

Cutler, Anne (ed.). 1982. *Slips of the Tongue and Language Production*. Berlin: Mouton Publishers.

Cutler, Anne. 1980. Errors of stress and intonation. In: Victoria A. Fromkin (ed.). *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, ch. 4, pp. 67–80.

Dahlbäck, Nils. 1995. Kinds of agents and types of dialogues. *Proceedings 9th Twente Workshop on Language Technology – Corpus Based Approaches to Dialogue Modelling*, Universiteit Twente, Enschede, The Netherlands. [Page numbers not obtained.]

Dahlbäck, Nils, Arne Jönsson & Lars Ahrenberg. 1993. Wizard of Oz studies – why and how. *Knowledge-Based Systems*, vol. 6, no. 4, pp. 258–266. Also in: Mark Maybury & Wolfgang Wahlster (eds.). 1998. *Readings in Intelligent User Interfaces*, Morgan Kaufmann, pp. 610–619.

Dahlbäck, Nils & Arne Jönsson. 1988. The Wizard of Oz in Computer Science: Simulations of Natural Language Interfaces. *Research Report NLPLAB-Memo 88-01*, June 1988, Department of Computer and Information Science, Linköping University, Sweden.

Dahlbäck, Nils & Arne Jönsson. 1986. A System for Studying Human–Computer Dialogues in Natural Language, *Research Report LiTH-IDA-R-86-42*, Department of Computer and Information Science, Linköping University, Sweden.

Dale, Paulette. 1977. Factors Related to Dysfluent Speech in Bilingual Cuban-American Adolescents. *Journal of Fluency Disorders*, vol. 2, pp. 311–314.

Dale, Rick, Daniel C. Richardson & Michael J. Owren. 2003. Pumping for gestural origins: The well may be rather dry. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 218–219.

Damasio, Antonio R. 1999. *The Scientific American Book of the Brain*. Guilford, Connecticut: The Lyons Press.

Danto, Arthur C. 1985. Consciousness and motor control. *Behavioral and Brain Sciences*, vol. 8, pp. 540–541.

Darley, Frederic L. 1955. The Relationship of Parental Attitudes and Adjustments to the Development of Stuttering. In: Wendell Johnson (ed.), *Stuttering in Children and Adults. Thirty Years of Research at the University of Iowa*, Minneapolis: University of Minneapolis Press, ch. 4, pp. 74–153.

Davidsen-Nielsen, Niels. 1971. A phonological analysis of English *sp*, *st*, *sk* with special reference to speech error evidence. *Journal of the International Phonetic Association*, vol. 5, pp. 3–25.

Davies, Paul. 1995. *Are We Alone?* London: Penguin.

Davies, Paul. 1987. *The Cosmic Blueprint. Order and Complexity at the Edge of Chaos*. London: Penguin.

Deacon, Terrence. 1997. *The Symbolic Species. The co-evolution of language and the human brain*. London: Allen Lane, The Penguin Press.

Debaisieux, Jeanne-Marie & José Deulofeu. 2001. Grammatically unacceptable utterances are communicatively accepted by native speakers, why are they? *Proceedings of DiSS '01 Disfluency in Spontaneous Speech*, 29–31 August 2001, University of Edinburgh, Scotland, pp. 69–72.

Dechert, Hans W. 1980. Pauses and intonation as indicators of verbal planning in second-language speech productions: Two examples from a case study. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 271–285.

Deecke, Lüder. 1987a. The natural explanation for the two components of the readiness potential. *Behavioral and Brain Sciences*, vol. 10, no. 4, pp. 781–782.

Deecke, L[üder]. 1987b. Bereitschaftspotential as an indicator of movement preparation in supplementary motor area and motor cortex. In: *Motor Areas of the Cerebral Cortex*, Ciba Foundation Symposium 132, Chicester: John Wiley & Sons, pp. 231–250.

Deecke, Lüder, Bernd Heise, Hans Helmut Kornhuber, Michael Lang & Wilfried Lang. 1984. Brain Potentials Associated with Voluntary Manual Tracking: Bereitschaftspotential, Conditioned Premotion Positivity, Directed Attention Potential, and Relaxation Potential. *Brain and Information. Event-related Potentials. Annals of the New York Academy of Sciences*, vol. 425, pp. 450–464.

Deecke, L[üder], J. Boschert, H. Weinberg & P. Brickett. 1983. Magnetic Fields of the Human Brain (Bereitschaftsmagnetfeld) Preceding Voluntary Foot and Toe Movements. *Experimental Brain Research*, vol. 52, pp. 81–86.

Deecke, L[üder], H. Weinberg & P. Brickett. 1982. Magnetic Fields of the Human Brain Accompanying Voluntary Movement: Bereitschaftsmagnetfeld. *Experimental Brain Research*, vol. 48, pp. 144–148.

Deecke, Lüder, Berta Grözinger & H[ans]. H. Kornhuber. 1976. Voluntary Finger Movement in Man: Cerebral Potentials and Theory. *Biological Cybernetics*, vol. 23, pp. 99–119.

Deecke, Lüder, Peter Scheid & Hans H. Kornhuber. 1969. Distribution of Readiness Potential, Pre-Motion Positivity, and Motor Potential of the Human Cerebral Cortex Preceding Voluntary Finger Movements. *Experimental Brain Research*, vol. 7, pp. 158–168.

Deese, James. 1986. Reality and control. *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 521–522.

Deese, James. 1978. Thought into speech. *American Scientist*, vol. 66, pp. 314–421.

DeJoy, Daniel A. & Hugo H. Gregory. 1985. The relationship between age and frequency of disfluency in preschool children. *Journal of Fluency Disorders*, vol. 10, pp. 107–122.

Dell, Gary S. 1986. A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, vol. 93, no. 3, pp. 283–321.

Dell, Gary S. 1984. Representation of Serial Order in Speech: Evidence From the Repeated Phoneme Effect in Speech Errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 10, no. 2, pp. 222–233.

Dell, Gary S. & Peter A. Reich. 1980. Toward a unified model of slips of the tongue. In: Victoria A. Fromkin (ed.). *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, ch. 19, pp. 273–286.

Dell, Gary S. & Peter A. Reich. 1975. A model of slips of the tongue. *The Third LACUS Forum*, vol. 3, pp. 448–455.

Delomier, Dominique, André Meunier & Mary-Annick Morel. 1989. Linguistic features of Human–Machine oral Dialogue. *Proceedings of Eurospeech '89*, September 1989, Paris, France, vol. 2, pp. 236–243.

Dembowski, James & Ben C. Watson. 1991. Preparation Time and Response Complexity Effects of Stutterers' and Nonstutterers' Acoustic LRT. *Journal of Speech and Hearing Research*, vol. 34, pp. 49–59.

Den, Yasuharu. 2003. Some strategies in prolonging speech segments in spontaneous Japanese. In: Robert Eklund (ed.), *Proceedings of DiSS '03, Disfluency in Spontaneous Speech Workshop*, 5–8 September 2003, Göteborg University, Sweden. *Gothenburg Papers in Theoretical Linguistics 90*, ISSN 0349–1021, pp. 87–90.

De Nil, Luc F. & Gene J. Brutten. 1991. Speech-Associated Attitudes of Stuttering and Nonstuttering Children. *Journal of Speech and Hearing Research*, vol. 34, pp. 60–66.

Dennett, Daniel C. 1991. *Consciousness Explained*. London: Penguin Books.

Dennett, Daniel C. & Marcel Kinsbourne. 1992a/1992b. Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, vol. 15, pp. 183–247. Includes: Daniel C. Dennett & Marcel Kinsbourne. 1992b. Escape from the Cartesian Theater. *Behavioral and Brain Sciences*, vol. 15, pp. 234–247.

Deschamps, Alain. 1980. The syntactical distribution of pauses in English spoken as a second language by French speakers. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 255–262.

De Smedt, Koenrad & Gerard Kempen. 1987. Incremental sentence production, self-correction and coordination. In: Gerard Kempen (ed.), *Natural language generation: New results in artificial intelligence, psychology and linguistics*. Dordrecht: Kluwer Academic Publishers, ch. 23, pp. 365–376.

Devitt, Michael & Kim Sterelny. 1987. *Language & Reality. An Introduction to the Philosophy of Language*. Oxford: Basil Blackwell.

Dewar, Ann., A. D. Dewar, W. T. S. Austin & H. M. Brash. 1979. The Long Term Use of an Automatically Triggered Auditory Feedback Masking Device in the Treatment of Stammering. *British Journal of Disorders of Communication*, vol. 14, pp. 219–229.

Dibner, Andrew S. 1958. Ambiguity and Anxiety. *Journal of Abnormal and Social Psychology*, vol. 56, pp. 165–174.

Dibner, Andrew S. 1956. Cue-Counting: A Measure of Anxiety in Interviews. *Journal of Consulting Psychology*, vol. 20, no. 6, pp. 475–478.

Dickins, Thomas E. 2003. Possible phylogenies: The role of hypotheses, weak inferences, and falsification. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 219–220.

Dierks, Thomas, David E. J. Linden, Martin Jandl, Elia Formisano, Rainer Goebel, Heinrich Lanfermann & Wolf Singer. 1999. Activation of Heschl's Gyrus during Auditory Hallucinations. *Neuron*, vol. 22, no. 3, pp. 615–621.

Dinnan, James A., Eugene McGuinness & Lawrence Perrinn. 1970. Auditory feedback: Stutterers versus nonstutterers. *Journal of Learning Disabilities*, vol. 3, pp. 209–213.

Dittman, Allen T. & Lynn G. Llewellyn. 1969. Body movement and speech rhythm in social conversation. *Journal of Personality and Social Psychology*, vol. 11, no. 2, pp. 98–106.

Dixon, Norman F. 1989. Unconscious perception and general anaesthesia. *Baillières Clinical Anaesthesiology*, vol. 3, no. 3, pp. 473–485.

Dixon, Norman F. 1981. *Preconscious Processing*. Chichester: John Wiley & Sons.

Donzel, Monique E van. See: Van Donzel, Monique.

Doty, Robert W. 1998. The five mysteries of the mind, and their consequences. *Neuropsychologia*, vol. 36, no. 10, pp. 1069–1076.

Doty, Robert W. 1985. The time course of conscious processing: Vetoes by the uninformed? *Behavioral and Brain Sciences*, vol. 8, pp. 541–542.

Duez, Danielle. 1995. Perception of Hesitations in Spontaneous French Speech. *Proceedings of the International Congress of Phonetic Sciences (ICPhS) '95*, 13–19 August 1985, Stockholm, Sweden, vol. 2, pp. 498–501.

Duez, Danielle. 1985. Perception of silent pauses in continuous speech. *Language and Speech*, vol. 28, pt. 4, pp. 377–389.

Duez, Danielle. 1983/84. Perception des pauses silencieuses dans la parole continue. *Travaux de L'Institut de Phonetique D'Aix*, vol. 9, pp. 31–83.

Duez, Danielle. 1982. Silent and non-silent pauses in three speech styles. *Language and Speech*, vol. 25, part 1, pp. 11–28.

Duez, Danielle. 1981/82. Pauses silencieuses et pauses non silencieuses dans trois types de messages oraux. *Travaux de L'Institut de Phonetique D'Aix*, vol. 8, pp. 85–114.

Duez, D[anielle] & R[ené] Carré. 1983. Perception of silent pauses in continuous speech. *Proceedings of the 10th International Congress of Phonetic Sciences (ICPhS) '83*, 1–6 August 1983, Utrecht, The Netherlands, Dordrecht: Foris Publications, p. 558.

Duffy, Robert J., Martin F. Hunt Jr. & Thomas G. Giolas. 1975. Effects of Four Types of Disfluency on Listener Reactions. *Folia Phoniatrica*, vol. 27, no. 2, pp. 106–115.

Dumas, Roland & Arlene Morgan. 1975. EEG asymmetry as a function of occupation, task, and task difficulty. *Neuropsychologia*, vol. 13, pp. 219–228.

Duval, Shelley & Robert Wicklund. 1972. *A theory of objective self awareness*. New York and London: Academic Press.

Eccles, John C. 1985. Mental summation: The timing of voluntary intentions by cortical activity. *Behavioral and Brain Sciences*, vol. 8, pp. 542–543.

Edelman, Gerald M. & Guilio Tononi. 2000. *A Universe of Consciousness*. New York: Basic Books.

Edelman, Gerald M. 1992. *Bright Air, Brilliant Fire*. New York: BasicBooks.

Edelsky, Carole. 1981. Who's got the floor? *Language and Society*, vol. 10, pp. 383–421.

Egland, George O. 1955. Repetitions and Prolongations in the Speech of Stuttering and Nonstuttering Children. In: Wendell Johnson (ed.), *Stuttering in Children and Adults. Thirty Years of Research at the University of Iowa*, Minneapolis: University of Minneapolis Press, ch. 6, pp. 181–188.

Eimer, Martin. 1998. The lateralized readiness potentials as an on-line measure of response activation processes. *Behavior Research Methods, Instruments, & Computers*, vol. 30, no. 1, pp. 146–156.

Eklund, Robert (ed.). 2003. *Proceedings of DiSS '03, Proceedings of Disfluency in Spontaneous Speech*, 5–8 September 2003, Göteborg University, Sweden. *Gothenburg Papers in Theoretical Linguistics 90*, ISSN 0349–1021.

Eklund, Robert. 2002. Ingressive Speech As An Indication That Humans Are Talking To Humans (And Not To Machines). *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2002*, 16–20 September 2002, Denver, Colorado, USA, vol. 2, pp. 837–840.

Eklund, Robert. 2001. Prolongations: A dark horse in the disfluency stable. *Proceedings of DiSS '01 Disfluency in Spontaneous Speech*, 29–31 August 2001, University of Edinburgh, Scotland, pp. 5–8.

Eklund, Robert. 2000a. Crosslinguistic Disfluency Modeling: A Comparative Analysis of Swedish and Tok Pisin Human–Human ATIS Dialogues. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2000*, 16–20 October 2000, Beijing, China, vol. 2, pp. 991–994.

Eklund, Robert. 2000b. Wanpela deitabeis long Tok Pisin bilong baim tiket bilong balus. (An ATIS database in Tok Pisin.) Methodological observations with regard to the collection of human–human data. *Proceedings of Fonetik 2000*, The Swedish Phonetics Conference, May 24–26, 2000, University of Skövde, Sweden, pp. 49–52.

Eklund, Robert. 1999. A Comparative Study of Disfluencies in Four Swedish Travel Dialogue Corpora. *Proceedings of Disfluency in Spontaneous Speech Workshop*, 1 July 1999, Berkeley, California, USA, pp. 3–6.

Eklund, Robert. 1997. Interaction between prosody and discourse structure in a simulated man–machine dialogue. *The Journal of the Acoustical Society of America*, vol. 102, no. 5, part. 2, December 1997, p. 3202.

Eklund, Robert, Jaan Kaja, Leonardo Neumeyer, Fuliang Weng & Vassilis Digalakis. 2000. Porting a Recogniser to a New Language. In: Manny Rayner, Dave Carter, Pierrette Bouillon, Vassilis Digalakis & Mats Wirén (eds.), *The Spoken Language Translator*, Cambridge, Cambridge University Press, ch. 17, pp. 265–273.

Eklund, Robert & Elizabeth Shriberg. 1998. Crosslinguistic Disfluency Modeling: A Comparative Analysis of Swedish and American English Human–Human and Human–Machine Dialogs. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 1998*, 30 November–5 December 1998, Sydney, Australia, vol. 6, pp. 2631–2634.

Eling, Paul. 1986. Speech and the Left Hemisphere: What Broca Actually Said. *Folia Phoniatrica*, vol. 38, pp. 13–15.

Elliot, Rogers. 1968. Simple visual and simple auditory reaction time: A comparison. *Psychonomic Science*, vol. 10, no. 10, pp. 335–336.

Ertl, J. & E[dward]. W. P. Schafer. 1969. Erratum. *Life Sciences*, vol. 8, no. 9, p. 559.

Ertl, J. & E[dward]. W. P. Schafer. 1967. Cortical activity preceding speech. *Life Sciences*, vol. 6, no. 5, pp. 473–479.

Esposito, Anna, Susan Duncan & Francis Quek. 2002. Holds as gestural correlates to empty and filled speech pauses. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2002*, 16–20 September 2002, Denver, Colorado, USA, vol. 1, pp. 541–544.

ESPS Programs. 1996. Version 5.1, Entropic Research Laboratory Inc. *ESPS/waves+* downloadable from **http://www.speech.kth.se/software/#esps**

Ezrati-Vinacour, Ruth, Rozanne Platzky & Ehud Yairi. 2001. The Young Child's Awareness of Stuttering-Like Disfluency. *Journal of Speech, Language, and Hearing Research*, vol. 44, pp. 368–380.

Faaborg-Anderson, K. & Å. W. Edfeldt. 1958. Electromyography of intrinsic and extrinsic laryngeal muscles during silent speech: correlation with reading activity. *Acta oto-laryngologica*, vol. 49, pp. 478–482.

Faure, Marc. 1980. Results of a contrastive study of hesitation phenomena in French and German. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 287–290.

Faurie, Charlotte & Michel Raymond. 2003. Handedness: Neutral or adaptive? *Behavioral and Brain Sciences*, vol. 26, no. 2, p. 220.

Fay, David & Anne Cutler. 1977. Malapropisms and the Structure of the Mental Lexicon. *Linguistic Inquiry*, vol. 8, no. 3, pp. 505–520.

Federmeier, Kara D., Jessica B. Segal, Tania Lombrozo & Marta Kutas. 2000. Brain responses to nouns, verbs and class-ambiguous words in context. *Brain*, vol. 123, pt. 12, pp. 2552–2566.

Feldstein, Stanley. 1962. The relationship of interpersonal involvement and affectiveness of content to the verbal communication of schizophrenic patients. *Journal of Abnormal and Social Psychology*, vol. 64, pp. 39–45.

Feldstein, Stanley, Marcia S. Brenner & Joseph Jaffe. 1963. The effect of subject sex, verbal interaction and topical focus on speech disruption. *Language and Speech*, vol. 6, p. 229–239.

Fenigstein, Allan, Michael F. Scheier & Arnold H. Buss. 1975. Public and Private Self-Consciousness: Assessment and Theory. *Journal of Consulting and Clinical Psychology*, vol. 43, no. 4, pp. 522–527.

Ferber, Rosa. 1995. The reliability and validity of slip-of-the-tongue corpora. A methodological note. *Linguistics*, vol. 33, pp. 1169–1190.

Feyereisen, Paul. 2003. Are human gestures in the present time a mere vestige of a former sign language? Probably not. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 220–221.

Feyereisen, Paul & Jacques-Dominique de Lannoy. 1991. *Gestures and speech: Psychological investigations*. Cambridge: Cambridge University Press.

Fillmore, Charles J. 1979. On Fluency. In: Charles J. Fillmore, Daniel Kempler & William S-.Y. Wang (eds.), *Individual Differences in Language Ability and Language Behavior*, New York: Academic Press, pp. 85–101.

Finlayson, Sheena, Victoria Forrest, Robin Lickley & Janet Mackenzie Beck. 2003. Effect of the restriction of hand gestures on disfluency. In: Robert Eklund (ed.), *Proceedings of DiSS '03, Disfluency in Spontaneous Speech Workshop*, 5–8 September 2003, Göteborg University, Sweden. *Gothenburg Papers in Theoretical Linguistics 90*, ISSN 0349–1021, pp. 21–24.

Fitts, Paul M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391.

Flanagan, Bruce, Israel Golddiamond & Nathan H. Azrin. 1959. Instatement of Stuttering in Normally Fluent Individuals through Operant Procedures. *Science*, vol. 130, pp. 979–981.

Flanagan, Bruce, Israel Golddiamond & Nathan [H.] Azrin. 1958. Operant Stuttering: The Control of Stuttering Behavior Through Response-Contingent Consequences. *Journal of the Experimental Analysis of Behavior* vol. 1, pp. 173–177.

Flanagan, Owen. 1992. *Consciousness Reconsidered*. Cambridge, Massachusetts: The MIT Press.

Flor-Henry, Pierre. 1986. Auditory hallucinations, inner speech, and the dominant hemisphere. *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 523–524.

Floyd, Susan & William H. Perkins. 1974. Early syllable dysfluency in stutterers and nonstutterers: A preliminary report. *Journal of Communication Disorders*, vol. 7, pp. 279–282.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, Massachusetts: The MIT Press.

Fodor, J[erry] A., T. G. Bever & M. F. Garrett. 1974. *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill.

Foit, A[lexander]., B[erta]. Grözinger & H[ans]. H. Kornhuber. 1982. Brain potential differences related to programming, monitoring and outcome of aimed and non-aimed, fast and slow Movements to a visual target: the Movement monitoring potential (MMP) and the task outcome evaluation potential (TEP). *Neuroscience*, vol. 7, p. 571.

Ford, Marilyn & Virginia M. Holmes. 1978. Planning units and syntax in sentence production. *Cognition*, vol. 6, pp. 35–53.

Forrest, A. D., A. J. Hay & A. W. Kushner. 1969. Studies in Speech Disorder in Schizophrenia. *British Journal of Psychiatry*, vol. 115, pp. 833–841.

Foulkes, David. 1991. Dream processing. *Behavioral and Brain Sciences*, vol. 14, no. 4, p. 678.

Foulkes, David. 1990. Dreaming and Consciousness. *European Journal of Cognitive Psychology*, vol. 2, no. 1, pp. 39–55.

Fourneret, Pierre & Marc Jeannerod. 1998. Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, vol. 36, no. 11, pp. 1133–1140.

Fouts, Roger S. & Gabriel Waters. 2003. Unbalanced human apes and syntax. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 221–222.

Fox, Barbara A., Makoto Hayashi & Robert Jasperson. 1996. Resources and repair: a cross-linguistic study of syntax and repair. In: Elinor Ochs, Emanuel E. Schegloff & Sandra A. Thompson (eds.), *Interaction and grammar*. Cambridge: Cambridge University Press, ch. 4, pp. 185–237.

Fox, Peter T., Roger J. Ingham, Janis C. Ingham, Frank Zamarripa, Jin-Hu Xiong & Jack L. Lancaster. 2000. Brain correlates of stuttering and syllable production. *Brain*, vol. 123, pp. 1985–2004.

Fox Tree, Jean E. 2001. Listeners' uses of *um* and *uh* in speech comprehension. *Memory and Cognition*, vol. 29, no. 2, pp. 320–236.

Fox Tree, Jean E. 1995. The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language*, vol. 34, pp. 709–728.

Fransella, Fay & H. R. Beech. 1965. An experimental analysis of the effect of rhythm on the speech of stutterers. *Behavior Research and Therapy*, vol. 3, pp. 195–201.

Fraser, Norman M. & G. Nigel Gilbert. 1991. Simulating speech systems. *Computer Speech and Language*, vol. 5, pp. 81–99.

Freud, Sigmund. 1901/1973. Slips of the tongue. In: Victoria A. Fromkin (ed.), *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 46–81.

Friederici, Angela D. 1995. The Time Course of Syntactic Activation during Language Processing: A Model Based on Neuropsychological and Neurophysiological Data. *Brain and Language*, vol. 50, pp. 259–281.

Friederici, Angela D., Karsten Steinhauer, Axel Mecklinger & Martin Meyer. 1998. Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology*, vol. 47, pp. 193–221.

Friederici, Angela D., Erdmut Pfeier & Anja Hahne. 1993. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, vol. 1, pp. 183–192.

Friedman, Ernest H. 1991a. Speech Pauses and Diagnosis. *Journal of Clinical Psychiatry*, vol. 52, no. 4, pp. 181–182.

Friedman, Ernest H. 1991b. Speech Hesitation Pauses as Markers for Mood Disorder in Stroke Patients? *Journal of Clinical Psychiatry*, vol. 52, no. 3, p. 140.

Frisch, Stefan A. & Richard Wright. 2002. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, vol. 30, pp. 139–162.

Frith, Chris. 2003. Sex, brains, robots and Buddhism: looking for free will. *New Scientist*, 10 May 2003, pp. 46–49. [Report from the Royal Society of Arts; Frith's comments on p. 46.]

Frith, Chris. 2002. Attention to action and awareness of other minds. *Consciousness and Cognition*, vol. 11, pp. 481–487.

Frith, Chris. 1999. How Hallucinations Make Themselves Heard. *Neuron*, vol. 22, no. 3, pp. 414–415.

Frith, C[hris]. D. 1987. The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*, vol. 17, pp. 631–648.

Frith, Chris. 1979. Consciousness, Information Processing and Schizophrenia. *British Journal of Psychology*, vol. 134, pp. 225–235.

Frith, Chris, Sarah-Jayne Blakemore & Daniel M. Wolpert. 2000. Abnormalities in the awareness and control of action. *Philosophical Tansactions. Royal Society of London. Series B. Biological Science*, vol. 355 (1404), pp. 1771–1188.

Frith, Christopher D. & D. John Done. 1989. Experiences of alien control in schizophrenia reflect a dsorder in the central monitoring of action. *Psychological Medicine*, vol. 19, pp. 359–363.

Fromkin, Victoria A. (ed.). 1980. *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press.

Fromkin, Victoria A. 1975. A Linguist Looks at "A Linguist Looks at 'Schizophrenic Language'". *Brain and Language*, vol. 2, pp. 498–503.

Fromkin, Victoria A. 1971/1973. The non-anomalous nature of anomalous utterances. In: Victoria A. Fromkin (ed.). 1973. *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 215–242. First published in *Language*, 1971, vol. 33, no. 1, pp. 27–52.

Frost, Julie A., Jeffrey R. Binder, Jane A. Springer, Thomas A. Hammeke, Patrick S. F. Bellgowan, Stephen M. Rao & Robert W. Cox. 1999. Language processing is strongly left lateralized in both sexes. *Brain*, vol. 122, pt. 2, pp. 199–208.

Funnell, Margaret G., Paul M. Corballis & Michael S. Gazzaniga. 2000. Insights into the functional specificity of the human corpus callosum. *Brain*, vol. 123, pt. 5, pp. 920–926.

Furnas, G[eorge] W., T[homas] K. Landauer, L[ouis] M. Gomez & S[usan] T. Dumais. 1987. The Vocabulary Problem in Human–System Communication. *Communications of the ACM*, vol. 30, no. 11, pp. 964–971.

Gabrea, M[arcel] & D[ouglas] O'Shaughnessy. 2000. Detection of filled pauses in spontaneous conversational speech, *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2000*, 16–20 October 2000, Beijing, China, vol. 2, pp. 678–681.

Gaines, Natalie D., Charles M. Runyan & Susan C. Meyers. 1991. A Comparison of Young Stutterers' Fluent Versus Stutterered Utterances on Measures of Length and Complexity. *Journal of Speech and Hearing Research*, vol. 34, pp. 37–42.

Galin, David & Ron R. Ellis. 1975. Asymmetry in evoked potentials as an index of lateralized cognitive processes: relation to EEG alpha asymmetry. *Neuropsychologia*, vol. 13, pp. 45–50.

Galin, David & Robert Ornstein. 1972. Lateral Specialization of Cognitive Mode: An EEG Study. *Psychophysiology*, vol. 9, no. 4, pp. 412–418.

Garber, Sharon F. & Richard R. Martin. 1978. Effects of noise and increased vocal intensity on stuttering. *Journal of Speech and Hearing Research*, vol. 20, pp. 233–240.

Garnham, Alan, Richard C. Shillcock, Gordon D. A. Brown, Andrew I. D. Mill & Anne Cutler. 1982. Slips of the tongue in the London-Lund corpus of spontaneous conversation. In: Anne Cutler (ed.), *Slips of the Tongue and Language Production*. Berlin: Mouton Publishers, pp. 251–263 (805–817).

Garnsey, Susan M. 1993. Event-related Brain Potentials in the Study of Language: An Introduction. *Language and Cognitive Processes*, vol. 8, pt. 4, pp. 337–356.

Garnsey, Susan M. & Gary S. Dell. 1984. Some Neurolinguistic Implications of Prearticulatory Editing in Production. *Brain and Language*, vol. 23, pp. 64–73.

Garrett, M[errill]. F. 1980a. Levels of Processing in Sentence Production. In: Brian Butterworth (ed.), *Language Production*, *vol. 1: Speech and Talk*. New York: Academic Press, pp. 177–220.

Garrett, Merrill F. 1980b. The limits of accommodation: arguments for independent processing levels in sentence production. In: Victoria A. Fromkin (ed.). *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, ch. 18, pp. 263–271.

Garrett, M[errill]. F. 1975. The analysis of sentence production. In: G. Bower (ed.): *The psychology of learning and motivation, vol. 9*. New York: Academic Press. pp. 133–177.

Gazzaniga, Michael S. 2002. The Split Brain Revisited. *Scientific American, Special Edition: The Hidden Mind*, August 31, 2002, pp. 27–31. [Updated version of Gazzaniga, 1998.]

Gazzaniga, Michael S. 2000. Cerebral specialization and interhemispheric communication. *Brain*, vol. 123, pp. 1293–1326.

Gazzaniga, Michael S. 1999. The Split Brain Revisited. In: Antonio R. Damasio (ed.), *The Scientific American Book of the Brain*, Guilford, Connecticut: The Lyons Press, pp. 129–138. [Same as Gazzaniga, 1998.]

Gazzaniga, Michael S. 1998. The Split Brain Revisited. *Scientific American*, July 1998, pp. 34–39.

Gazzaniga, Michael S. 1992. *Nature's Mind. The Biological Roots of Thinking, Emotions, Sexuality, Language and Intelligence*. London: Penguin Books.

Gazzaniga, Michael S. 1983. Right Hemisphere Language Following Brain Bisection. *American Psychologist*, vol. 38, pp. 525–537.

Gazzaniga, Michael S. 1970. *The Bisected Brain*. New York: Appleton-Century-Crofts.

Gazzaniga, Michael S. 1967. The Split Brain in Man. *Scientific American*, vol. 217, pp. 24–29.

Gazzaniga, M[ichael] S. & S[teven] A. Hillyard. 1973. Attention Mechanisms following Brain Bisection. In: Sylvan Kornblum (ed.), *Attention and Performance IV*, New York: Academic Press, pp. 221–238.

Gazzaniga, M[ichael]. S. & R[oger]. W. Sperry. 1967. Language after section of the cerebral commissures. *Brain*, vol. 1967, vol. 90, pp. 131–148.

Gazzaniga, M[ichael]. S., J[oseph]. E. Bogen & R[oger]. W. Sperry. 1965. Observations on visual perception after disconnexion of the cerebral hemispheres in man. *Brain*, vol. 88, pt. 2, pp. 221–236.

Geldard, Frank A. & Carl E. Sherrick. 1972. The Cutaneous "Rabbit": A Perceptual Illusion. *Science*, vol. 178, pp. 178–179.

Gilden, L., H. G. Vaughan Jr. & L[ouis]. D. Costa. 1966. Summated human EEG potentials with voluntary movement. *Electroencephalography and Clinical Neurophysiology*, vol. 20, pp. 433–438.

Giles, Howard & Jennifer A. Giles. 1976. Comments on "Speech Fluency Fluctuations During the Menstrual Cycle". *Journal of Speech and Hearing Research*, vol. 19, no. 1, pp. 187–188.

Gillett, Grant R. 2003. Word and talk – handedness and the stuff of life. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 222–223.

Glausiusz, Josie. 2002. The Art of, Um, Speaking Clearly. *Discover*, vol. 23, no. 11, November 2002, R&D, p. 13.

Glynn, Ian M. 1990. Consciousness and time. *Nature*, vol. 348, pp. 477–479.

Goffman, Erving. 1978. Response Cries. *Language*, vol. 54, pp. 787–815.

Goldman-Eisler, Frieda. 1972. Pauses, clauses, sentences. *Language and Speech*, vol. 15, pp. 103–113.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in spontaneous speech*. London & New York: Academic Press.

Goldman-Eisler, Frieda. 1961. The distribution of pause durations in speech. *Language and Speech*, vol. 4, part 4, pp. 232–237.

Goldman-Eisler, Frieda. 1958a. Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology*, vol. 10, pp. 96–106.

Goldman-Eisler, Frieda. 1958b. Speech analysis and mental processes. *Language and Speech*, vol. 1, pp. 59–75.

Goldman-Eisler, Frieda, 1958c. The predictability of words in context and the length of pauses in speech. *Language and Speech*, vol. 1, pp. 226–231.

Goldman-Eisler, F[rieda]. 1957. Speech Production and Language Statistics. *Nature*, vol. 28, December 1957, p. 1497.

Goldman-Eisler, Frieda. 1955. Speech-breathing activity—A measure of tension and affect during interviews. *British Journal of Psychology*, vol. 46, pp. 53–63.

Goldman-Eisler, F[rieda]. 1954a. A study of individual differences and of interaction in the behaviour of some aspects of language in interviews. *Journal of Mental Science*, vol. 100, pp. 177–197.

Goldman-Eisler, Frieda. 1954b. On the variability of the speech of talking and on its relation to the length of utterances in conversations. *British Journal of Psychology*, vol. 45, pp. 94–107.

Gomes, Gilberto. 2002. Problems in the Timing of Conscious Experience. *Consciousness and Cognition*, vol. 11, pp. 191–197.

Gomes, Gilberto. 1999. Volition and the Readiness Potential. *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 59–76. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 59–76.

Gomes, Gilberto. 1998. The Timing of Conscious Experience: A Critical Review and Reinterpretation of Libet's Research. *Consciousness and Cognition*, vol. 7, pp. 559–595.

Gordon, Pearl A. & Harold L. Luper. 1989. Speech Disfluencies in Nonstutterers: Syntactic Complexity and Production Task Effects. *Journal of Fluency Disorders*, vol. 14, pp. 429–445.

Grabow, Jack D. & Frederick W. Elliot. 1974. The electrophysiologic assessment of hemispheric asymmetries during speech. *Journal of Speech and Hearing Research*, vol. 17, pp. 64–72.

Gray, Jeffrey A. 1991. What is the relation between language and consciousness? *Behavioral and Brain Sciences*, vol. 14, no. 4, p. 679.

Greenfield, Patricia M. 1991. Language, tools and the brain: The ontogeny and phylogeny of hierarchically organized sequential behaviors. *Behavioral and Brain Sciences*, vol. 14, no. 4, pp. 531–595.

Greiner, Jay R., Hiram Fitzgerald & Paul A. Cooke. 1986. Speech fluency and hand performance on a sequential tapping task in left- and right-handed stutterers and nonstutterers. *Journal of Fluency Disorders*, vol. 11, pp. 55–69.

Grice, Martine, Matthias Reyelt, Ralf Benzmüller, Jörg Mayer & Anton Batliner. 1996. Consistency in Transcription and Labelling of German Intonation with GToBI. *Proceedings of the International Conference on Spoken Language Processing (ICLSP) '96*, 3–6 October 1996, Philadelphia, Pennsylvania, USA, vol. 3, pp. 1716–1719.

Grice, [H.] Paul. 1989/1997. Utterer's Meaning and Intentions. Studies in the Way of Words, Cambridge: Harvard University Press. Republished in: Peter Ludlow (ed.), *Readings in the Philosophy of Language*, Cambridge, Massachusetts: The MIT Press, ch. 4, pp. 59–88.

Grice, H. P[aul]. 1975. Logic and conversation. In: P. Cole & J. L. Morgan (eds.), *Syntax and Semantics III: Speech Acts*. New York: Academic Press, pp. 41–58.

Grosjean, François. 1980a. Linguistic structures and performance structures: Studies in pause distribution. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 91–106.

Grosjean, François. 1980b. Comparative studies of temporal variables in spoken and sign languages: A short review. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 307–312.

Grosjean, François. 1979. A Study of Timing in a Manual and a Spoken Language: American Sign Language and English. *Journal of Psycholinguistic Research*, vol. 8, no. 4, pp. 379–405.

Grosjean, François & Maryann Collins. 1979. Breathing, Pausing and Reading. *Phonetica*, vol. 36, pt. 2, pp. 98–114.

Grosjean, François, Lysiane Grosjean & Harlan Lane. 1979. The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology*, vol. 11, pp. 58–81.

Grosjean, François & Harlan Lane. 1977. Pauses and syntax in American Sign Language. *Cognition*, vol. 5, pp. 101–117.

Grosjean, François & Alain Deschamps. 1975. Analyses contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, vol. 31, pp. 144–184.

Grosjean, F[rançois]. & A[lain]. Deschamps. 1972. Analyse des variables temporelles du français spontané. *Phonetica*, vol. 26, no. 3, pp. 129–157.

Grözinger, B[erta]., H[ans]. H. Kornhuber, J. Kriebel, J[an]. Szirtes & K. T. Westphal. 1980. The Bereitschaftspotential Preceding the Act of Speaking. Also an Analysis of Artifacts. *Progress in Brain Research*, vol. 54, pp. 798–804.

Grözinger, B[erta], H[ans]. H. Kornhuber & J. Kriebel. 1975. Methodological problems in the investigation of cerebral potentials preceding speech: determining the onset and suppressing artefacts caused by speech. *Neurophysiologia*, vol. 13, pp. 263–270.

Grözinger, B[erta]., J. Kriebel, H[ans]. H. Kornhuber & K. Murata. 1974. Cerebral potentials during respiration and preceding vocalization. *Electroencephalography and Clinical Neurophysiology*, vol. 36, pp. 435.

Grözinger, B[erta]., J. Kriebel & H[ans]. H. Kornhuber. 1974. Respiration Correlated Brain Potentials. *Journal of Interdiscliplinary Cycle Research*, vol. 5, nos. 3–4, pp. 287–294.

Grözinger, B[erta]., H[ans]. H. Kornhuber & J. Kriebel. 1973. Inter- and intra-hemispheric asymmetries of brain potentials preceding speech and phonation. *Electroencephalography and Clinical Neurophysiology*, vol. 34, no. 7, pp. 737–738.

Guindon, Raymonde. 1988. A Multidisciplinary Perspective on Dialogue Structure in User–Advisor Dialogues. In: Raymonde Guindon (ed.), *Cognitive Science and its Applications for Human–Computer Interaction*. Hillsdale, New Jersey: Erlbaum, pp. 163–197.

Gulick, Robert Van. See: Van Gulick, Robert.

Gurman Bard, Ellen. See: Bard, Ellen Gurman.

Hadar, U., T. J. Steiner & F. Rose. 1984. The relationship between head movements and speech dysfluencies. *Language and Speech*, vol. 27, pt. 4, pp. 333–342.

Haggard, Patrick. 2001. The psychology of action. *British Journal of Psychology*, vol. 92, pp. 113–128.

Haggard, Patrick, Sam Clark & Jeri Kalogeras. 2002. Voluntary action and conscious awareness. *Nature Neuroscience*, vol. 5, pt. 4, pp. 382–385.

Haggard, Patrick & Benjamin Libet. 2001. Conscious Intention and Brain Activity. *Journal of Consciousness Studies*, vol. 8, no. 11, pp. 47–63.

Haggard, Patrick & Martin Eimer. 1999. On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, vol. 126, pt. 1, pp. 128–133.

Haggard, Patrick, Chris Newman & Elena Magno. 1999. On the perceived time of voluntary actions. *British Journal of Psychology*, vol. 90, pp. 291–303.

Haggard, Patrick & Elena Magno. 1999. Localising awareness of action with transcranial magnetic stimulation. *Experimental Brain Research*, vol. 127, pp. 102–107.

Hagoort, Peter & Colin M. Brown. 2000. ERP effects of listening to speech: semantic ERP effects. *Neurophysiologia*, vol. 38, pp. 1518–1530.

Hahn, Walther von. 1986. Pragmatic considerations in man–machine discourse. *Proceedings of COLING '86*, 25–29 August 1986, Bonn, Germany, pp. 520–526.

Ham, Richard E. 1990. What is stuttering: variations and stereotypes. *Journal of Fluency Disorders*, vol. 15, pp. 259–273.

Hameroff, Stuart. 1998a. Quantum computation in brain microtubules? The Penrose-Hameroff "Orch OR" model of consciousness. *Philosophical Transactions Royal Society London (A)*, vol. 356, pp. 1869–1896.

Hameroff, Stuart. 1998b. Anesthesia, Consciousness and Hydrophobic Pockets – A Unitary Quantum Hypothesis of Anesthetic Action. *Toxicology Letters*, vol. 100/101, pp. 31–39.

Hamilton, John. 1985. Auditory Hallucinations in Nonverbal Quadriplegics. *Psychiatry*, vol. 48, pp. 382–392.

Hand, C. Rebekah & William O. Haynes. 1983. Linguistic processing and reaction time differences in stutterers and nonstutterers. *Journal of Speech and Hearing Research*, vol. 26, pp. 181–185.

Hanna, Richmond & Stephen Morris. 1977. Stuttering, speech rate, and the metronome effect. *Perceptual and Motor Skills*, vol. 44, pp. 452–454.

Hannah, Elaine P. & Joanne G. Gardner. 1968. A note on syntactic relationships in nonfluency. *Journal of Speech and Hearing Research*, vol. 11, pp. 835–860.

Hansson, Petra. 1998. Pausing in Spontaneous Speech. *Proceedings of FONETIK 98*, 27–29 May 1998, Stockholm University, Sweden, pp. 158–161.

Harley, Trevor A. 1986. Speech errors and hallucinations in schizophrenia – no difference? *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 525–526.

Harley, Trevor A. 1984. A Critique of Top-Down Independent Levels Models of Speech Production: Evidence from Non-plan-Internal Speech Errors. *Cognitive Science*, vol. 8, pp. 191–219.

Hartsuiker, Robert J., Martin Corley, Robin Lickley & Melanie Russell. 2003. Perception of disfluency in people who stutter and people who do not stutter: Results from magnitude estimation. In: Robert Eklund (ed.), *Proceedings of DiSS '03, Disfluency in Spontaneous Speech Workshop*, 5–8 September 2003, Göteborg University, Sweden. *Gothenburg Papers in Theoretical Linguistics 90*, ISSN 0349–1021, pp. 35–38.

Hauptmann, Alexander G. 1989. Speech and Gestures for Graphic Image Manipulation. *Proceedings of CHI '89*, 30 April–4 May 1989, Austin, Texas, USA, pp. 241–245.

Hauptmann, Alexander G. & Alexander I. Rudnicky. 1990. A Comparison of Speech and Typed Input. *Proceedings of the DARPA Speech and Natural Language Workshop*, June 1990, Hidden Valley, Pennsylvania, USA, pp. 219–223.

Hauptmann, Alexander G. & Alexander I. Rudnicky. 1988. Talking to computers: an empirical investigation. *International Journal of Man–Machine Studies* , vol. 28, pp. 583–604.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, vol. 298, pp. 1569–1579.

Hawkins**,** P. R. 1971. The syntactic location of hesitation pauses. *Language and Speech*, vol. 14, part 3, pp. 277–288.

Hayden, Paul A., Martin R. Adams & Nanette Jordahl. 1982. The Effects of Pacing and Masking on Stutterers' and Nonstutterers' Speech Initiation Times. *Journal of Fluency Disorders*, vol. 7, pp. 9–19.

Hayes, Cathy. 1951. *The APE in Our House*. New York: Harper & Brothers.

Hayes, Keith J. & Catherine Hayes. 1952. Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology*, vol. 45, no. 5, pp. 450–459.

Hayes, Keith J. & Cathy Hayes. 1951. The intellectual development of a home-raised chimpanzee. *Proceedings of the American Philosophical Society*, vol. 95, no. 2, pp. 105–109.

Healey, E. Charles & Bonnie Bernstein. 1991. Acoustic analyses of young stutterers' and nonstutters' disfluencies. In: Herman F. M. Peters, Wouter Hulstijn & C. Woodruff Starkweather (eds.), *Speech motor control and stuttering*. New York: Elsevier, ch. 37, pp. 401–407.

Healey, E. Charles & Barbara Gutkin. 1984. Analysis of stutterers' voice onset times and fundamental frequency contours during fluency. *Journal of Speech and Hearing Research*, vol. 27, pp. 219–225.

Healey, Charles, A. R. Mallard III & Martin R. Adams. 1976. Factors contributing to the reduction of stuttering during singing. *Journal of Speech and Hearing Research*, vol. 19, pp. 475–480.

Heeman, Peter Anthony. 1997. *Speech Repairs, Intonational Boundaries and Discourse Markers: Modeling Speakers' Utterances in Spoken Dialog*. PhD thesis, Department of Computer Science, University of Rochester, New York.

Heeman, Peter A. & James F. Allen. 1999. Speech Repairs, Intonational Phrases and Discourse Markers: Modeling Speakers' Utterances in Spoken Dialogues. *Computational Linguistics*, vol. 25, no. 4, pp. 527–571.

Heeman, Peter A. & Kyung-ho Loken-Kim. 1999. Detecting And Correcting Speech Repairs In Japanese. *Proceedings of Disfluency in Spontaneous Speech Workshop*, 1 July 1999, Berkeley, California, USA, pp. 43–46.

Heeman, Peter A. & James F. Allen. 1997. Intonational Boundaries, Speech Repairs and Discourse Markers: Modeling Spoken Dialog. *Proceedings of ACL/EACL '97*, 7–11 July 1997, Madrid, Spain, pp. 254–261.

Heeman, Peter A., Kyung-ho Loken-Kim & James F. Allen. 1996. Combining the Detection and Correction of Speech Repairs. *Proceedings of ISSD 96*, 2–3 October 1996, Philadelphia, Pennsylvania, USA, pp. 133–136.

Heeman, Peter A. & Kyung-ho Loken-Kim. 1995. Using Structural Information to Detect Speech Repairs. *Technical Report IEICE SP95-91, December 1995*.

Heeman, Peter A. & James [F.] Allen. 1994a. Detecting and Correcting Speech Repairs. *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, 27 June–1 July 1994, Las Cruces, Mexico pp. 295–302.

Heeman, Peter A. & James [F.] Allen. 1994b. Tagging Speech Repairs. *Proceedings of the Human Technology Workshop*, 8–11 March 1994, Princeton, New Jersey, USA, pp. 187–192.

Hegedüs, Lajos. 1953. On the problem of the pauses of speech. *Acta Linguistica Academiae Scientarium Hungaricae*, vol. 3, pp. 1–34.

Heldner, Mattias & Eva Strangert. 2001. Temporal effects of focus in Swedish. *Journal of Phonetics*, vol. 29, no. 3, pp. 329–361.

Helenius, Päivi, Riita Salmelin, Elisabet Service & John F. Connolly. 1998. Distinct time courses of word and context comprehension in the left temporal cortex. *Brain*, vol. 121, pt. 6, pp. 1133–1142.

Helfrich, Hede. 1980. A digital method of pause extraction. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 247–252.

Hemphill, Charles T., John J. Godfrey & George R. Doddington. 1990. The ATIS Spoken Language Systems Pilot Corpus. *Proceedings of DARPA Speech and Natural Language Workshop*, pp. 96–101. **http://www.ldc.upenn.edu/readme_files/atis/sspcrd/corpus.html**

Hensel, Herbert & Kurt A. Boman. 1960. Afferent impulses in cutaneous sensory nerves in human subjects. *Journal of Neurophysiology*, vol. 23, pp. 564–578.

Herning, Ronald I., Reese T. Jones & Johanna S. Hunt. 1987. Speech Event Related Potentials Reflect Linguistic Content Processing. *Brain and Language*, vol. 30, pp. 116–129.

Herning, Ronald I. & Reese T. Jones. 1984. Slow Potentials during Speech Processing. *Annals of the New York Academy of Sciences*, vol. 425, pp. 212–215.

Hieke, Adolph E. 1981. A content-processing view of hesitation phenomena. *Language and Speech*, vol. 24, no. 2, pp. 147–160.

Hieke, Adolph E., Sabine Kowal & Daniel C. O'Connell. 1983. The trouble with "articulatory" pauses. *Language and Speech*, vol. 26, part 3, pp. 203–214.

Hill, Archibald A. 1973. A theory of speech errors. In: Victoria A. Fromkin (ed.), *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 205–214.

Hill, Harris E. 1954. An Experimental Study of Disorganization Of Speech And Manual Responses In Normal Subjects. *Journal of Speech and Hearing Disorders*, vol. 19, pp. 295–305.

Hindle, Donald. 1983. Deterministic Parsing of Syntactic Non-fluencies. *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, 15–17 June 1983, Cambridge, Massachussetts, USA, pp. 123–128.

Hink, R. F., H. Kohler, L[üder]. Deecke & H[ans]. H. Kornhuber. 1982. Risk-taking and the human Bereitschaftspotential. *Electroencephalography and Clinical Neurophysiology*, vol. 53, pp. 361–373.

Hirst, Graeme, Susan McRoy, Peter Heeman, Philip Edmonds & Diane Horton. 1994. Repairing conversational misunderstandings and non-understandings. *Speech Communication*, vol. 15, pp. 213–229.

Hockett, Charles F. 1967/1973. Where the Tongue Slips, There Slip I. In: Victoria A. Fromkin (ed.), *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 93–119. Originally published in: *To Honor Roman Jakobson, vol. II (Janua Linguarum, Series Maior XXXII*, The Hague: Mouton, pp. 910–936.

Hodgson, David. 1999. Hume's Mistake. 1999. *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 201–224. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 201–224.

Hoffman, Ralph E. 1986. Verbal hallucinations and language production processes in schizophrenia. *Behavioural and Brain Sciences*, vol. 9, pp. 503–548.

Hoffman, Ralph E. & Richard E. Kravitz. 1987. Feedforward action regulation and the experience of will. *Behavioral and Brain Sciences*, vol. 10, no. 4, p. 782–783.

Hoffman, Ralph E., George L. Hogben, Harry Smith & WM. Ford Calhoun. 1985. Message disruption during syntactic processing in zchizophrenia. *Journal of Communication Disorders*, vol. 18, pp. 183–202.

Hofstadter, Douglas R. 1986. The architecture of Jumbo. In: R. Michalski, J. Carbonell, & T. Mitchell (eds.), *Proceedings of the International Machine Learning Workshop*, University of Illinois, Urbana, Illinois, USA, pp. 161–170.

Hokkanen, Tapio. 2001. *Slips of the tongue. Errors, repairs, and a model*. Studia Fennica Linguistics 10. Helsinki: Finnish Literature Society.

Holcomb, Phillip J. 1993. Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, vol. 30, pp. 47–61.

Holcomb, Phillip J. 1988. Automatic and Attentional Processing: An Event-Related Brain Potential Analysis of Semantic Priming. *Brain and Language*, vol. 35, pp. 66–85.

Holcomb, Phillip J. & Helen J. Neville. 1991. Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, vol. 19, no. 4, pp. 286–300.

Holloway, Ralph. 2003. Was a manual gesturing stage really necessary? *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 223–224.

Holloway, Ralph. 1976. Paleoneurological evidence for language origins. In: Stevan R. Harnad, Horst D. Steklis & Jane Lancaster (eds.), *Origins and Evolution of Language and Speech*, *Annals of the New York Academy of Sciences*, vol. 280, Part VII. The Fossil Record and Neural Organization, pp. 330–348.

Holmes, V[irginia]. M. 1988. Hesitations and Sentence Planning. *Language and Cognitive Processes*, vol. 3, pt. 4, pp. 323–361.

Honderich, Ted. 1984. The Time of a Conscious Sensory Experience and Mind–Brain Theories. *Journal of Theoretical Biology*, vol. 110, pp. 115–129.

Hopkins, Williams D. & Claudio Cantalupo. 2003. Brodmann's area 44, gestural communication, and the emergence of right handedness in chimpanzees. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 224–225.

Horii, Yoshiyuki & Peter R. Ramig. 1987. Pause and utterance durations and fundamental frequency characteristics of repeated oral readings by stutterers and nonstutterers. *Journal of Fluency Disorders*, vol. 12, pp. 257–270.

Hotopf, W. H. N. 1983. Lexical Slips of the Pen and Tongue: What they tell us about Language Production. In: Brian Butterworth (ed.). *Language Production, Volume 2: Development, Writing and Other Language Processes*. London: Academic Press, pp. 147–199.

Houde, John F. & Michael I. Jordan. 1998. Sensorimotor Adaptation in Speech Production. *Science*, vol. 279, pp. 1213–1216.

Howell, Peter. 1990. Changes in voice level caused by several forms of altered feedback in fluent speakers and stutterers. *Language and Speech*, vol. 33, no. 4, pp. 325–338.

Howell, Peter, Karima Kadi-Hanifi & Keith Young. 1991. Phrase revisions in fluent and stuttering children. In: Herman F. M. Peters, Wouter Hulstijn & C. Woodruff Starkweather (eds.), *Speech Motor Control and Stuttering*, ch. 39, pp. 415–422.

Hubbard, Carol P. & Ehud Yairi. 1988. Clustering of disfluencies in the speech of stuttering and nonstuttering preschool children. *Journal of Speech and Hearing Research*, vol. 31, pp. 228–233.

Hulit, Lloyd M. 1976. Effects of nonfluencies of comprehension. *Perceptual and Motor Skills*, vol. 42, pp. 1119–1122.

Hulit, Lloyd M. & Sharon K. Haasler. 1989. Influence of suggestion on the nonfluencies of normal speakers. *Journal of Fluency Disorders*, vol. 14, pp. 359–369.

Iacobini, Marco, Alain Ptito, Nicole Y. Weekes & Eran Zaidel. 2000. Parallel visuomotor processing in the split brain: cortico-subcortical interactions. *Brain*, vol. 123, pt. 4, pp. 759–769.

Indefrey, Peter, Colin M. Brown, Frauke Hellwig, Katrin Amunts, Hans Herzog, Rüdiger J. Seitz & Peter Hagoort. 2001. *Proceedings of the National Academy of Sciences*, vol. 98, no. 10, pp. 5933–5936.

Ingvar, David H. On Volition. 1999. *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 1–10. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 1–10.

Irwin, Donald A., John R. Knott, Dale W. McAdam & Charles S. Rebert. 1966. Motivational Determinants of the "Contingent Negative Variation". *Electroencephalography and Clinical Neurophysiology*, vol. 21, pp. 538–543.

Iverson, Jana M. & Esther Thelen. 2003. The hand leads to the mouth in ontogenesis too. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 225–226.

Iverson, Jana M. & Susan Goldin-Meadow. 1998. Why people gesture when they speak. *Nature*, vol. 396, p. 228.

Ivry, Richard B. & Lynn C. Robertson. 1998. *The Two Sides of Perception*. Cambridge, Massachusetts: MIT Press.

Jackendoff, Ray. 1995. *Languages of the Mind. Essays on Mental Representation*. Cambridge, Massachusetts: The MIT Press.

James, William. 1890 (facsimile edition 1950). *The Principles of Psychology*. New York: Dover Publications.

Jasper, Herbert H. 1985. Brain mechanisms of conscious experience and voluntary action. *Behavioral and Brain Sciences*, vol. 8, p. 543.

Jayaram, M. 1984. Distribution of stuttering in sentences: relationship to sentence length and clause position. *Journal of Speech and Hearing Research*, vol. 27, pp. 338–341.

Jaynes, Julian. 1990. Verbal Hallucinations and Preconscious Mentality. In: Manfred Spitzer & Brendan H. Maher (eds.), *Philosophy and Psychopathology*, New York: Springer Verlag, pp. 159–170.

Jaynes, Julian. 1986. Hearing voices and the bicameral mind. *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 525–527.

Jaynes, Julian. 1980. Consciousness and the voices of the mind. *Canadian Psychology*, vol. 27, pp. 128–195.

Jaynes, Julian. 1976/2000. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. New York: First Mariner Books.

Johns, L. C., S. Rossell, C[hris]. Frith, F. Ahmad, D. Hemsley, E. Kuipers & P[hilip]. K. McGuire. 2001. Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine*, vol. 31, pt. 4, pp. 704–715.

Johnson, Helen & Patrick Haggard. 2003. The effect of attentional cueing on conscious awareness of stimulus and response. *Experimental Brain Research*, vol. 150, pp. 490–496.

Johnson, Raymond L. & Miles D. Miller. 1965. Auditory hallucinations and intellectual deficit. *Journal of Psychiatric Research*, vol. 3, pp. 37–41.

Johnson, Wendell. 1961. Measurement of Oral Reading and Speaking Rate and Disfluency of Adult Male and Female Stutterers and Nonstutterers. *Journal of Speech and Hearing Disorders Monograph Supplement Number 7*, pp. 1–20.

Johnson, Wendell and Associates. 1959. *The Onset of Stuttering: Research Findings and Implications*. Minneapolis: University of Minnesota Press.

Johnson, Wendell (ed.). 1955. *Stuttering in Children and Adults. Thirty Years of Research at the University of Iowa*. Minneapolis: University of Iowa Press.

Johnson, Wendell, Spencer F. Brown, James F. Curtis, Clarence W. Edney & Jacqueline Keaster. 1948. Stuttering. In: Wendell Johnson, Spencer F. Brown, James F. Curtis, Clarence W. Edney & Jacqueline Keaster (eds.), *Speech Handicapped School Children*. New York: Harper & Brothers Publishers, ch 5, pp. 179–257.

Johnson-Frey, Scott. 2003. Mirror neurons, Broca's area and language: Reflecting on the evidence. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 226–227.

Johnston, Sharon J., Kenneth L. Watkin & Peter T. Macklem. 1993. Lung volume changes during relatively fluent speech in stutterers. *Journal of Applied Physiology*, vol. 75, no. 2, pp. 696–703.

Jones, Gregory V. & Maryanne Martin. 2003. Dual asymmetries in handedness. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 227–228.

Jönsson, Arne & Nils Dahlbäck. 1988. Talking to a computer is not like talking to your best friend. *Research Report LiTH-IDA-R-88-34*, September 1988, Department of Computer and Information Science, Linköping University, Sweden. Also published in *Proceedings of The first Scandinavian Conference on Artificial Intelligence*, 9–11 March 1988, Tromsø, Norway, pp. 53–68.

Josse, Goulven & Nathalie Tzourio-Mazoyer. 2003. What functional imaging of the human brain can tell us about handedness and language. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 228–229.

Jung, Richard. 1985. Voluntary intention and conscious selection in complex learned action. *Behavioral and Brain Sciences*, vol. 8, pp. 544–545.

Junginger, John. 1986. Distinctiveness, unintendedness, location, and nonself attribution of verbal hallucinations. *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 527–528.

Jürgens, Uwe. 2003. From mouth to mouth and hand to hand. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 229–230.

Jurich, Anthony P. & Charyl J. Polson. 1985. Nonverbal assessment of anxiety as a function of intimacy of sexual attitude questions. *Psychological Reports*, vol. 57, pp. 1247–1253.

Kaplan, Bernard. 1957. Critique and Notes on the Phenomena of "Opposite Speech". *Journal of Abnormal and Social Psychology*, vol. 55, pp. 389–393.

Karniol, Rachel. 1995. Stuttering, Language, and Cognition: A Review and a Model of Stuttering as Suprasegmental Sentence Plan Alignment (SPA). *Psychological Bulletin*, vol. 117, no. 1, pp. 104–124.

Kasl, Stanislav V. & George F. Mahl. 1987. Speech Disturbances and Experimentally Induced Anxiety. In: George F. Mahl (ed.), *Explorations in Nonverbal and Vocal Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum, ch. 12, pp. 203–213.

Kasl, Stanislav V. & George F. Mahl. 1965. The relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, vol. 1, pp. 425–433.

Kasl, Stanislav V. & George F. Mahl. 1958. Experimentally induced anxiety and speech disturbances. *American Psychologist*, vol. 13, p. 349.

Keller, I. & H. Heckenhausen. 1990. Readiness potentials preceeding spontaneous motor acts: voluntary vs. involuntary control. *Electroencephalography and clinical Neurophysiology*, vol. 76, pp. 351–361.

Kelley, J. F. 1984. An Iteractive Design Methodology for User-Friendly Natural Language Office Information Applications. *Association for Computing Machinery Transactions on Office-Information Systems*, vol. 2, pp. 26–41.

Kelley, J. F. 1983. An empirical methodology for writing User-Friendly Natural Language computer applications. *Proceedings of CHI '83,* 12–15 December 1983, Boston, Massachussetts, USA, pp. 193–196.

Kellog, Winthrop N. 1968. Communication and Language in the Home-Raised Chimpanzee. *Science*, vol. 162, pp. 423–427.

Kellog, W[inthrop] N. & L. A. Kellog. 1933. *The Ape and the Child*. NewYork and London: Whittlesey House.

Kelly, Ellen M. & Edward G. Conture. 1992. Speaking Rates, Response Time Latencies, and Interrupting Behaviors of Young Stutterers, Nonstutterers, and Their Mothers. *Journal of Speech and Hearing Research*, vol. 35, pp. 1256–1267.

Kelly, Spencer D. 2003. From past to present: Speech, gesture, and brain in present-day human communication. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 230–231.

Kempen, Gerard & Edward Hoenkamp. 1987. An Incremental Procedural Grammar for Sentence Formulation. *Cognitive Science*, vol. 11, pp. 201–258.

Kemper, Susan. 1992. Adults Sentence Fragments: Who, What, When, Where and Why? *Communications Research*, vol. 19, pp. 444–458.

Kendon, Adam. 1972. Some Relationships Between Body Motion and Speech. In: Aron Wolfe Siegman & Benjamin Pope (eds.), *Studies in Dyadic Communication*, Elmsford, New York: Pergamon, pp. 177–210.

Kennedy, Alan, Alan Wilkes, Leona Elder & Wayne S. Murray. 1988. Dialogues with machines. *Cognition*, vol. 30, pp. 37–72.

Kent, Ray D. 1983. Facts about stuttering: neuropsychologic perspectives. *Journal of Speech and Hearing Research*, vol. 48, pp. 249–255.

Khedr, Eman, Waguih Abd El-Nasser, Emad K. Abdel Haleem, M. Salama Bakr & Mohamed N. Trakhan. 2000. Evoked Potentials and Electroencephalography in Stuttering. *Folia Phoniatrica et Logopaedica*, vol. 52, pp. 178–186.

Kikui, Gen-ichiro & Tsuyoshi Morimoto. 1994. Similarity-based identification of repairs in Japanese spoken language. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '94*, 18–22 September 1994, Yokohama, Japan, vol. 2, pp. 915–918.

Kimble, Gregory A. & Lawrence C. Perlmuter. 1970. The Problem of Volition. *Psychological Review*, vol. 77, no. 5, pp. 361–384.

Klein, Daniel. 1981. "Pay no attention to the man behind the curtain". Open Channel, *Computer*, vol. 14, no. 11, p. 112.

Klein, Stanley. 2002a. Libet's Research on the Timing of Conscious Intention to Act: A Commentary. *Consciousness and Cognition*, vol. 11, pp. 273–279.

Klein, Stanley. 2002b. Libet's Timing of Mental Events: Commentary on the Commentaries. *Consciousness and Cognition*, vol. 11, pp. 326–333.

Knight, Chris. 2003. The secret of lateralisation is trust. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 231–232.

Knott, John R. & Donald A. Irwin. 1973. Anxiety, Stress, and the Contingent Negative Variation. *Archives of General Psychiatry*, vol. 29, pp. 538–541.

Koch, Cristof & Francis Crick. 1991. Understanding awarenss at the neuronal level. *Behavioral and Brain Sciences*, vol. 14, no. 4, pp. 683–685.

Kohn, Susan E., Arthur Wingfield, Lise Menn, Harold Goodglass, Jean Berko Gleason & Mary Hyde. 1987. Lexical Retrieval: The tip-of-the-tongue phenomenon. *Applied Psycholinguistics*, vol. 8, pp. 245–266.

Kolers, Paul A. & Michael von Grünau. 1976. Shape and color in apparent motion. *Vision Research*, vol. 16, pp. 329–335.

Kolk, Herman. 1991. Is stuttering a sympton of adaptation or of impairment? In: Herman F. M. Peters, Wouter Hulstijn & C. Woodruff Starkweather (eds.), *Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering*, Nijmegen, The Netherlands, ch. 9, pp. 131–140.

Kools, Joseph A. & Joan D. Berryman. 1971. Differences in disfluency behavior between male and female nonstuttering children. *Journal of Speech and Hearing Research*, vol. 14, no. 1, pp. 125–130.

Koomen, Willem & Wil Dijkstra. 1975. Effects of question length on verbal behavior in a bias-reduced interview situation. *European Journal of Social Psychology*, vol. 5, pp. 399–403.

Koopmans, Marina, Iman Slis & Toni Rietveld. 1991. The influence of word position and word type on the incidence of stuttering. In: Herman F. M. Peters, Wouter Hultijn & C. Woodruff Starkweather (eds.), *Speech motor control and stuttering*. New York: Elsevier, ch. 30, pp. 333–340.

Kornhuber, Hans H. 1987. Voluntary Activity, Readiness Potential, and Motor Program. In: George Adelman (ed.), *Encyclopedia of Neuroscience*. Boston: Birkhäuser, pp. 1302–1303.

Kornhuber, Hans H. & Lüder Deecke. 1965. Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv*, vol. 284, pp. 1–17.

Kremen, William S., Larry J. Seidman, Stephen V. Faraone & Ming T. Tsuang. 2003. Is there disproportionate impairment or phonemic fluency in schizophrenia? *Journal of the International Neuropsychological Society*, vol. 9, no. 1, pp. 79–88.

Kurzweil, Ray. 1999. *The Age of Spiritual Machines*. New York: Penguin Books.

Kutas, Marta & Steven A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature*, vol. 307, pp. 161–163.

Kutas, Marta & Steven A. Hillyard. 1980. Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, vol. 207, pp. 203–205.

Kutas, Marta & Emanuel Donchin. 1980. Preparation to respond as manifested by movement-related brain potentials. *Brain Research*, vol. 202, pp. 95–115.

Lackner, James R. & Betty H. Tuller. 1979. Role of Efference Monitoring in the Detection of Self-Produced Speech Errors. In: William E. Cooper (ed.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, Hillsdale, New Jersey: Erlbaum, pp. 281–294.

Laffal, Julius, L. Duoglas Lenkoski & Lane Ameen. 1956. "Opposite Speech" in a schizophrenic patient. *Journal of Abnormal and Social Psychology*, vol. 52, pp. 409–413.

Lalljee, Mansur [G.] & Mark Cook. 1973. Uncertainty in first encounters. *Journal of Personality and Social Psychology*, vol. 26, no. 1, pp. 137–141.

Lallgee, Mansur G. & Mark Cook. 1969. An experimental investigation of the function of filled pauses in speech. *Language and Speech*, vol. 12, pt. 1, pp. 24–28.

Landis, J. Richard & Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, vol. 33, pp. 159–174.

Lane, Harlan & Bernhard Tranel. 1971. The Lombard Sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, vol. 14, pp. 677–709.

Langer, Ellen J. & Lois G. Imber. 1979. When Practice Makes Imperfect: Debilitating Effects of Overlearning. *Journal of Personality and Social Psychology*, vol. 37, no. 11, pp. 2014–2024.

Lashley, K. S. 1951. The Problem of Serial Order in Behavior. In: L. A. Jeffress (ed.), *Hixon Symposium on Cerebral Mechanisms in Behavior*, New York, pp. 112–136.

Lassen, Carol L. 1987. Effect of Proximity on Anxiety and Communication in the Initial Psychiatric Interview. In: George F. Mahl (ed.), *Explorations in Nonverbal and Vocal Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum, ch. 5, pp. 108–119.

Latto, Richard. 1985. Consciousness as an experimental variable: Problems of definition, practice, and interpretation. *Behavioral and Brain Sciences*, vol. 8, pp. 545–546.

Laver, John [D. M.]. 1980a. Slips of the tongue as neuromuscular evidence for a model of speech production. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 21–26.

Laver, John [D. M.]. 1980b. Monitoring systems in the neurolinguistic control of speech perception. In: Victoria A. Fromkin (ed.). *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, ch. 20, pp. 287–305.

Laver, John [D. M.]. 1970. The production of speech. In: John Lyons (ed.), *New Horizons in Linguistics*, Harmondsworth: Penguin Books, pp. 53–75.

Laver, John D. M. 1969/1973. The detection and correction of slips of the tongue. *Work in Progress*, vol. 3, D epartment of Phonetics and Linguistics, University of Edinburgh, Scotland. Republished in: Victoria A. Fromkin (ed.). 1973. *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 132–143.

Lay, Clarry H. & Allan Paivio. 1969. The effects of task difficulty and anxiety on hesitations in speech. *Canadian Journal of Behavioural Sciences*, vol. 1, pp. 25–37.

Leavens, David A. 2003. Integration of visual and vocal communication: Evidence for Miocene origins. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 232–233.

Lebrun, Yvan & John Van Borsel. 1990. Final sound repetitions. *Journal of Fluency Disorders*, vol. 15, pp. 107–113.

Lecours, André Roch & Marie Vanier-Clément. 1976. Schizophasia and Jargonaphasia. *Brain and Language*, vol. 3, pp. 516–565.

Lee, Bernard S. 1951. Artificial stutter. *Journal of Speech and Hearing Disorders*, vol. 16, pp. 53–55.

Lee, Bernard S. 1950. Effects of Delayed Speech Feedback. *Journal of the Acoustical Society of America, (JASA)*, vol. 22, no. 6, pp. 824–826.

Lee, Chia-Ying, Wen-Jui Kuo, Jeng-Ren Duann, Yu-Chen Liang, Daisy L. Hung, Ovid J. L. Tzeng & Jen-Chuen Hsieh. 1999. A fMRI Study on Chinese Phonemic and Semantic Verbal Fluency Task. Poster presented at the Fifth International Conference on Human Brain Mapping (HBM'99), 22–26 June 1999, Düsseldorf, Germany. Abstract published in *Neuroimage*, vol. 9, no. 6, p. 1062.

Lee, Yue-Shi & Hsin-His Chen. 1997. Using acoustic and prosodic cues to correct Chinese speech repairs. *Proceedings of Eurospeech '97*, 22–25 September 1997, Rhodes, Greece, vol. 4, pp. 2211–2214.

Lee, Tzu-Lun, Ya-Fang He, Yun-Ju Huang, Shu-Chuan Tseng & Robert Eklund. [Submitted for publication]. Prolongation in spontaneous Mandarin.

Leeper, Linda H. & Richard Culatta. 1995. Speech Fluency: Effect of Age, Gender and Context. *Folia Phoniatrica et Logopaedica*, vol. 47, pp. 1–14.

Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: John Wiley & Sons.

Lerea, Louis. 1956. A preliminary study of the verbal behaviour of speech fright. *Speech Monographs*, vol. 23, pp. 229–233.

Levelt, Willem J. M. 1989. *Speaking. From Intention to Articulation*. Cambridge, Massachusetts: MIT Press.

Levelt, Willem J. M. 1983a. Spontaneous Self-Repairs in Speech: Processes and Representations. *Proceedings of the Tenth International Congress of Phonetic Sciences (ICPhS) '83*, 1–6 August 1983, Utrecht, The Netherlands, Dordrecht: Foris Publications, pp. 105–117.

Levelt, Willem J. M. 1983b. Monitoring and self-repair in speech. *Cognition*, vol. 14, pp. 41–104.

Levelt, Willem J. M. & Anne Cutler. 1983. Prosodic Marking in Speech Repair. *Journal of Semantics*, vol. 2, no. 2, pp. 205–217.

Levin, Harry & Irene Silverman. 1965. Hesitation phenomena in children's speech. *Language and Speech*, vol. 8, pt. 2, pp. 67–85.

Levitt, Andrea G. & Alice F. Healy. 1985. The Roles of Phoneme Frequency, Similarity, and Availability in the Experimental Elicitation of Speech Errors. *Journal of Memory and Language*, vol. 24, pp. 717–733.

Levow, Gina-Anne. 2002. Adaptations in spoken corrections: Implications for models of conversational speech. *Speech Communication*, vol. 36, pp. 147–163.

Levow, Gina-Anne. 1998. Characterizing and Recognizing Spoken Corrections in Human–Computer Dialogue. *Proceedings of COLING-ACL '98*, 10–14 August 1998, Montreal, Quebec, Canada, vol. 1, pp. 736–742.

Libet, Benjamin. 2002. The Timing of Mental Events: Libet's Experimental Findings and Their Implications. *Consciousness and Cognition*, vol. 11, pp. 291–299.

Libet, Benjamin. 1999. Do We Have Free Will? *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 47–57. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 47–57.

Libet, Benjamin. 1993. The neural time factor in conscious and unconscious events. *Experimental and Theoretical Studies of Consciousness*, Ciba Foundation Symposium 174, Chicester: Wiley, pp. 123–146.

Libet, Benjamin. 1992a. Voluntary acts and readiness potentials. *Electroencephalography and Clinical Neurophysiology*, vol. 82, pp. 85–86.

Libet, Benjamin. 1992b. Models of conscious timing and the experimental evidence. *Behavioral and Brain Sciences*, vol. 15, pp. 213–215.

Libet, Benjamin. 1991a. Conscious vs neural time. *Nature*, vol. 352, p. 27.

Libet, Benjamin. 1991b. Conscious functions and brain processes. *Behavioral and Brain Sciences*, vol. 14, no. 4, pp. 685–686.

Libet, Benjamin. 1990. Time-delays in conscious processes. *Behavioral and Brain Sciences*, vol. 13, no. 4, p. 672.

Libet, Benjamin. 1987. Are the mental experiences of will and self-control significant for the performance of a voluntary act? *Behavioral and Brain Sciences*, vol. 10, no. 4, pp. 783–786.

Libet, Benjamin. 1985a/1985b. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, vol. 8, pp. 529–539. Includes: Libet, Benjamin. 1985b. Theory and evidence relating to cerebral processes to conscious will. *Behavioral and Brain Sciences*, vol. 8, pp. 558–566.

Libet, Benjamin. 1985c Subjective Antedating of a Sensory Experience and Mind–Brain Theories: Reply to Honderich (1984). *Journal of Theoretical Biology*, vol. 114, pp. 563–570.

Libet, Benjamin. 1981. The experimental evidence for subjective referral of a sensory experience backwards in time: Reply to P. S. Churchland. *Philosophy of Science*, vol. 48, pp. 182–197.

Libet, Benjamin. 1966. Brain Stimulation and the Threshold of Conscious Experience. In: John C. Eccles (ed.), *Brain and conscius experience. Study Week September 28 to October 4, 1964, of the Pontifica Academia Scientiarum*, Città del Vaticano. New York: Springer-Verlag, ch. 7, pp. 165–181.

Libet, Benjamin. 1965. Cortical activation in conscious and unconscious experience. *Perspectives in Biology and Medicine*, vol. 9, pp. 77–86.

Libet, Benjamin, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic.

Libet, B[enjamin], E[lwood] W. Wright, B[ertram] Feinstein & D[ennis] K. Pearl. 1992. Retroactive Enhancement of a Skin Sensation by a Delayed Cortical Stimulus in Man: Evidence for Delay of a Conscious Sensory Experience. *Consciousness and Cognition*, vol. 1, pp. 367–375.

Libet, Benjamin, Dennis K. Pearl, David. E. Morledge, Curtis A. Gleason, Yoshio Hosobuchi & Nicholas M. Barbaro. 1991. Control of the transition from sensory detection to sensory awareness in man by the duration of a thalamic stimulus. *Brain*, vol. 114, pp. 1731–1757.

Libet, Benjamin, Curtis A. Gleason, Elwood W. Wright & Dennis K. Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness potential). The unconscious initiation of a freely voluntary act. *Brain*, vol. 106, pp. 623–642.

Libet, Benjamin, Elwood W. Wright, Jr., Bertram Feinstein & Dennis K. Pearl. 1979. Subjective referral of the timing for a conscious sensory experience. A functional role for the somatosensory specific projection system in man. *Brain*, vol. 102, pp. 193–224.

Libet, B[enjamin], W. W. Alberts, E. W. Wright, Jr. & B. Feinstein. 1967. Responses of Human Somatosensory Cortex below Threshold for Conscious Sensation. *Science*, vol. 158, pp. 1597–1600.

Libet, B[enjamin], W. W. Alberts, E. W. Wright, Jr. L. D. Delattre, G. Levin & B. Feinstein. 1964. Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex. *Journal of Neurophysiology*, vol. 27, pp. 546–578.

Lickley, Robin. 1996. Juncture cues to disfluency. *Proceedings of the International Conference on Spoken Language Processing (ICPSLP) '96*, 3–6 October 1996, Philadelphia, Pennsylvania, USA, vol. 4, pp. 2478–2481.

Lickley, Robin. 1995. Missing disfluencies. *Proceedings of the International Congress of Phonetic Sciences (ICPhS) '95*, 13–19 August 1995, Stockholm, Sweden, vol. 4, pp. 192–195.

Lickley, Robin J. 1994. *Detecting Disfluency in Spontaneous Speech*. PhD Thesis, University of Edinburgh, Scotland.

Lickley, Robin, David McKelvie & Ellen Gurman Bard. 1999. Comparing human and automatic speech recognition using word gating. *Proceedings of Disfluency in Spontaneous Speech Workshop*, 1 July 1999, Berkeley, California, USA, pp. 23–26.

Lickley, R[obin]. J. & E[llen]. G[urman]. Bard. 1998a. When Can Listeners Detect Disfluency in Spontaneous Speech? *Language and Speech*, vol. 41. no. 2, pp. 203–226.

Lickley, Robin & Ellen Gurman Bard. 1998b. Disfluent Speech. The Transcriber Problem. *Proceedings of Sounds Patterns of Spontaneous Speech: Production and Perception*. ESCA workshop, 24–26 September 1998, La Baume-les-Aix, France, pp. 105–108.

Lickley, R[obin]. J. & E[llen]. G[urman]. Bard. 1996. On not Recognizing Disfluencies in Dialogue. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '96*, 3–6 October 1996, Philadelphia, Pennsylvania, USA, vol. 3, pp. 1876–1879.

Lickley, Robin & Ellen Gurman Bard. 1992. Processing disfluent speech: recognising disfluency before lexical access. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '92*, 12–16 October 1992, Banff, Canada, vol. 2, pp. 935–938.

Lickley, R[obin]. J., R[ichard]. C. Shillcock & E[llen]. G[urman]. Bard. 1991. Processing disfluent speech: how and when are disfluencies found? *Proceedings of Eurospeech '91*, 24–26 September 1991, Genova, Italy, vol. 3, pp. 1499–1502.

Liu, Yang, Elizabeth Shriberg & Andreas Stolcke. 2003. Automatic Disfluency Identification in Conversational Speech Using Multiple Knowledge Sources. *Proceedings Eurospeech 2003*, 1–4 September 2003, Geneva, Switzerland, vol. 2, pp. 957–960.

Livant, William Paul. 1963. Antagonistic functions of verbal pauses: filled and unfilled pauses in the solution of additions. *Language and Speech*, vol. 6, part 1, pp. 1–4.

Loftus, Josephine, Lynn E. Delisi & Timothy J. Crow. 2000. Factor structure and familiarity of first-rank symptoms in sibling pairs with schizophrenia and schizoaffective disorder. *British Journal of Psychiatry*, vol. 177, pp. 15–19.

Long, Karen M. & Rebekah H. Pindzola. 1985. Manual reaction time to linguistic stimuli in child stutterers and nonstutterers. *Journal of Fluency Disorders*, vol. 10, pp. 143–149.

Lombard, E. 1911. Le signe du l'elevation de la voix. *Annales des maladies de l'oreille dy larynx, du nez et du pharynx*, vol. 37, pp. 101–119.

Lounsbury, Floyd G. 1954. Transitional Probability, Linguistic Structure, and Systems of Habit-family Hierarchies. In: Charles E. Osgood & Thomas A. Sebeok (eds.), *Psycholinguistics. A Survey of Theory and Research Problems*. Baltimore: Waverly Press, pp. 93–101.

Love, Laura Russ & Lloyd A. Jeffress. 1971. Identification of brief pauses in the fluent speech of stutterers and nonstutterers. *Journal of Speech and Hearing Research*, vol. 14, pp. 229–240.

Love, William Robert. 1955. The Effect of Pentobarbital Sodium (Nembutal) and Amphetamine Sulphate (Benzedrine) on the Severity of Stuttering. In: Wendell Johnson (ed.), *Stuttering in Children and Adults. Thirty Years of Research at the University of Iowa*, Minneapolis: University of Minneapolis Press, ch. 23, pp. 298–310.

Lüngen, Harald, Martina Pampel, Guido Drexel, Dafydd Gibbon, Frederek Althoff & Christoph Schillo. 1996. Morphology and speech technology. *Proceedings of the ACL–SIGPHON Conference*, 28 June 1996, Santa Cruz, California, USA, pp. 25–30.

Luper, Harold L. 1956. Consistency Of Stuttering In Relation To The Goal Gradient Hypothesis. *Journal of Speech and Hearing Disorders*, vol. 21, no. 3, pp. 336–342.

Luria, A[lexander]. R[omanivich]. 1961. *The role of speech in the regulation of normal and abnormal behavior*. Oxford: Pergamon Press.

Luria, Alexander Romanivich. 1960. Verbal regulation of behavior. In: Mary A. R. Brazier (ed.), *The central nervous system and behavior*. Madison Printing Company, pp. 359–423.

MacDermid, Catriona & Camilla Eklund. 1997. *Report on the first WOZ Simulation for the SLT–DB Project*. Technical Report, 11 November 1997, Telia Research AB.

MacDermid, Catriona & Camilla Eklund. 1996. *Simulering av en automatiserad översättningstjänst för rese-bokningar*. Technical Report, 17 May 1996, Telia Research AB.

MacDermid, Catriona & Mikael Goldstein. 1996. The 'Storyboard' Method: Establishing an Unbiased Vocabulary For Keyboard and Voice Applications. *Adjunct Proceedings of Human–Computer Interaction '96*, London, England, pp. 104–109.

MacDonald, James D. & Richard R. Martin. 1973. Stuttering and disfluency as two reliable and unambiguous response classes. *Journal of Speech and Hearing Research*, vol. 16, pt. 4. pp. 691–699.

MacKay, Donald G. 1987. *The Organization of Perception and Action: A Theory for Language and Other Cognitive Skills*. New York: Springer.

MacKay, Donald G. 1971. Stress Pre-Entry in Motor Systems. *American Journal of Psychology*, vol. 84, pp. 35–51.

MacKay, Donald G. 1970. Spoonerisms: the structure of errors in the serial order of speech. *Neuropsychologia*, vol. 8, pp. 323–350. Also published in: Victoria A. Fromkin (ed.). 1973. *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 164–194.

MacKay, Donald M. 1969. Evoked Brain Potentials as Indicators of Sensory Information Processing. *Neurosciences Research Program Bulletin*, vol. 7, no. 3, June 1969, Brookline, Massachusetts: MIT Press.

Maclay, Howard & Charles E. Osgood. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word*, vol. 5, pp. 19–44.

MacNeilage, Peter F. 2003. Mouth to hand and back again? Could language have made those journeys? *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 233–234.

MacNeilage, Peter F. & Barbara L. Davis. 2000. On the Origin of Internal Structure of Word Forms. *Science*, vol. 288, pp. 527–531.

Maddox, John. 1997. The price of language? *Nature*, vol. 388, pp. 424–425.

Maher, Brendan. 1972. The Language of Schizophrenia: A Review and Interpretation. *British Journal of Psychiatry*, vol. 120, pp. 3–17.

Mahl, George F. (ed.). 1987a. *Explorations in nonverbal and vocal Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum.

Mahl, George F. 1987b. Everyday Disturbances in Speech. In: Robert L. Russell (ed.), *Language in Psychotherapy: Strategies of Discovery*. New York and London: Plenum Press, ch. 6, pp. 213–269.

Mahl, George F. 1958. On the use of "ah" in spontaneous speech: Quantitative, developmental, characterological, situational, and linguistic aspects. *American Psychologist*, vol. 13, p. 349.

Mahl, George F. 1956. Disturbances and silences in the patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology*, vol. 53, pp. 1–15.

Mahl, George F. & Arthur S. Bender. 1987. Dialect, Stress, and Identity Feelings. In: George F. Mahl (ed.), Explorations in Nonverbal and Vocal Behavior. Hillsdale, New Jersey: Lawrence Erlbaum, ch. 20, pp. 310–327.

Mahl, George F, Gene Schulze & Edward J. Murray. 1987. Speech Disturbances and Manifest Verbal Content in Psychotherapeutic Interviews. In: George F. Mahl (ed.), *Explorations in Nonverbal and Vocal Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum, ch. 17, pp. 267–276.

Malhotra, Ashok. 1975. Knowledge-based English language systems for management support: an analysis of requirements. *Proceedings of the fourth International Joint Conference on Artificial Intelligence (IJCAI-75)*, Tblisi, Georgia, USSR, 3–8 September 1975, pp. 842–847.

Marks, Lawrence E. 1985. Toward a psychophysics of intention. *Behavioral and Brain Sciences*, vol. 8, pp. 547.

Martin, James G. 1970. On Judging Pauses in Spontaneous Speech. *Journal of Verbal Learning and Verbal Behavior*, vol. 9, pp. 75–78.

Martin, James & Winifred Strange. 1968. The perception of hesitation in spontaneous speech. *Perception and Psychophysics*, vol. 3, no. 6, pp. 427–438.

Martin, Richard R., Samuel K. Haroldson & Patricia Kuhl. 1972a. Disfluencies in child–child and child–mother speaking situations. *Journal of Speech and Hearing Research*, vol. 15, no. 4, pp. 753–756.

Martin, Richard R., Samuel K. Haroldson & Patricia Kuhl. 1972b. Disfluencies of young children in two speaking situations. *Journal of Speech and Hearing Research*, vol. 15, no. 4, pp. 831–836.

Mayo, Catherine, Matthew Aylett & D. Robert Ladd. 1997. Prosodic transcription of Glasgow English: an evaluation study of GlaToBI. *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications*, 18–20 September 1997, Athens, Greece, pp. 231–234.

McAdam, Dale W. & Harry A. Whitaker. 1971. Language Production: Electroenchephalographic Localization in the Normal Human Brain. *Science*, vol. 172, pp. 499–502.

McAdam, Dale W. & David M. Seales. 1969. Bereitschaftspotential enhancement with increased level of motivation. *Electroencephalography and Clinical Neurophysiology*, vol. 27, pp. 73–75.

McCloskey, D. I., J. G. Colebatch, Erica K. Potter & D. Burke. 1983. Judgements About Onset of Rapid Voluntary Movements in Man. *Journal of Neurophysiology*, vol. 49, no. 4, pp. 851–863.

McCroskey, James C. & R. Samuel Mehrley. 1969. The effects of disorganization and nonfluency on attitude and source credibility. *Speech Monographs*, vol. 36, pp. 13–21.

McDonough, Alanna (Neilsen) & Robert W. Quesal. 1988. Locus of control orientation of stutterers and nonstutterers. *Journal of Fluency Disorders*, vol. 13, pp. 97–106.

McFarlane, Stephen C. & Kenneth G. Shipley. 1981. Latency of vocalization onset for stutterers and nonstutterers under conditions of auditory and visual cueing. *Journal of Speech and Hearing Disorders*, vol. 46, pp. 307–312.

McGuire, P[hilip]. K., D. Robertson, A. Thacker, A. S. David, N. Kitson, R. S. J. Frackowiak & C. D. Frith. 1997. Neural correlates of thinking in sign language. *NeuroReport*, vol. 8, pp. 695–698.

McGuire, P[hilip]. K., D. A. Silberzweig, I. Wright, R. M. Murray, R. S. J. Frackowiak & C. D. Frith. 1996. The Neural Correlates of Inner Speech and Auditory Verbal Imagery in Schizophrenia: Relationship to Auditory Verbal Hallucinations. *British Journal of Psychiatry*, vol. 169, pp. 148–159.

McGuire, P[hilip]. K., G. M. S. Shah & R. M. Murray. 1993. Increased blod flow in Broca's area during auditory hallucinations in schizophrenia. *Lancet*, vol. 342, pp. 703–706.

McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature*, vol. 264, pp. 746–748.

McKee, George, Brian Humphrey & Dale W. McAdam. 1973. Scaled Lateralization of Alpha Activity During Linguistic and Musical Tasks. *Psychophysiology*, vol. 10, no. 4, pp. 441–443.

McKinnon, Richard, Mark Allen & Lee Osterhout. 2003. Morphological decomposition involving non-productive morphemes: ERP evidence. *NeuroReport*, vol. 14, no. 6, pp. 883–886.

Meisels, Murray. 1967. Test anxiety, stress, and verbal behavior. *Journal of Consulting Psychology*, vol. 31, pp. 577–582.

Mellor, C. S. 1970. First Rank Symptoms of Schizophrenia. *British Journal of Psychiatry*, vol. 117, pp. 15–23.

Merikle, Philip M. & Jim Cheesman. 1985. Conscious and unconscious processes: Same or different? *Behavioral and Brain Sciences*, vol. 88, pp. 547–548.

Meteer, Marie et al. 1995. *Dysfluency Annotation Stylebook for the Switchboard Corpus*. Unpublished manuscript, February 1995, revised version by Ann Taylor, June 1995.

Meyers, Susan C. 1986. Qualitative and quantitative differences and patterns of variability in disfluencies emitted by preschool stutterers and nonstutterers during dyadic conversations. *Journal of Fluency Disorders*, vol. 11, pp. 293–306.

Meyers, Susan C. & Frances J. Freeman. 1985. Interruptions as a variable in stuttering and disfluency. *Journal of Speech and Hearing Research*, vol. 28, pp. 428–435.

Miall, R. C. & D. M. Wolpert. 1996. Forward Models for Physiological Motor Control. *Neural Networks*, vol. 9, no. 8, pp. 1265–1279.

Michel, George F. 2003. Ontogenetic constraints on the evolution of right-handedness. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 234–235.

Miller, George A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, vol. 63, pp. 81–97.

Miller, Gerald R. & Murray A. Hewgill. 1964. The effects of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech*, vol. 50, pp. 36–44.

Miller, Miles, Raymond L. Johnson & Lewis H. Richmond. 1965. Auditory hallucinations and descriptive langage skills. *Journal of Psychiatry Research*, vol. 3, pp. 43–56.

Minsky, Marvin. 1985. *The Society of Mind*. New York: Simon & Schuster.

Mohrhoff, Ulrike. 1999. The Physics of Interactionism. *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 165–184. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 165–184.

Moore, Jr., Walter H. 1976. Bilateral Tachistoscopic Word perception of Stutterers and Normal Subjects. *Brain and Language*, vol. 3. pp. 434–442.

Moore, Jr., Walter H. & Einer Boberg. 1987. Hemispheric Processing and Stuttering. In: Lena Rustin, Harry Purser & David Rowley (eds.), *Progress in the treatment of fluency disorders*. London: Taylor & Francis, ch. 2, pp. 19–42.

Moore, Jr., Walter H. & William O. Haynes. 1980. Alpha hemispheric asymmetry and stuttering: Some support for a segmentation dysfunction hypothesis. *Journal of Speech and Hearing Research*, vol. 23, pp. 229–247.

Moore, Jr., Walter H. & Mary K. Lang. 1977. Alpha asymmetry over the right and left hemispheres of stutterers and control subjects preceding massed oral readings: a preliminary investigation. *Perceptual and Motor Skills*, vol. 44, pp. 223–230.

Moray, Neville. 1959. Attention in dichotic listening: affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, vol. 9, pp. 56–60.

Morel, M[ary].-A[nnick]. 1989. Computer–Human Communication. In: M. M. Taylor, F. Néel & D. G. Bouwhuis (eds*.), The Structure of Multimodal Dialogue*, Amsterdam: North-Holland, ch. 24, pp. 323–330.

Morrell, Lenore K. & Dorothy A. Huntington. 1972. Cortical potentials time-locked to speech production: evidence for probable cerebral origin. *Life Sciences*, vol. 11, pp. 921–929.

Morrell, Lenore K. & Dorothy A. Huntington. 1971. Electrocortical Localization of Language Production. *Science*, vol. 174, pp. 1359–1360.

Morrell, Lenore K. & Joseph G. Salamy. 1971. Hemispheric Asymmetry of Electrocortical Responses to Speech Stimuli. *Science*, vol. 174, pp. 164–166.

Mortensen, Chris. 1985. Conscious decisions. *Behavioral and Brain Sciences*, vol. 8, pp. 548–549.

Morton, John. 1964. A model for continuous language behaviour. *Language and Speech*, vol. 7, pp. 40–70.

Morton, John, Steve Marcus & Clive Frankish. 1976. Perceptual Centers (P-centers). *Psychological Review*, vol. 83, no. 5, pp. 405–408.

Motley, Michael T., Carl T. Camden & Bernard J. Baars. 1982. Covert Formulation and Editing of Anomalies in Speech Production: Evidence from Experimentally Elicited Slips of the Tongue. *Journal of Verbal Learning and Verbal Behavior*, vol. 21, pp. 578–594.

Motley, Michael T. & Bernard J. Baars. 1975. Encoding sensitivities to phonological markedness and transitional probability: evidence from spoonerisms. *Human Communication Research*, vol. 2, pp. 353–361.

Murphy, Marianne & John M. Baumgartner. 1981. Voice Initiation and Termination Time in Stuttering and Nonstuttering Children. *Journal of Fluency Disorders*, vol. 6, pp. 257–264.

Murray, Iain R., Chris Baber & Allan South. 1996. Towards a definition and working model of stress and its effects on speech. *Speech Communication*, vol. 20, pp. 3–12.

Näätänen, R. 1985. Brain physiology and the unconscious initiation of movements. *Behavioral and Brain Sciences*, vol. 8, pp. 549.

Nakatani, Christine & Elizabeth Shriberg. 1993. *Guidelines for labeling disfluencies in ToBI*. SRI Technical Report, November 1993, Menlo Park, California, USA.

Nakatani, Christine & Julia Hirschberg. 1994. A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustic Society of America (JASA)*, vol. 95, no. 3, pp. 1603–1616.

Nakatani, Christine & Julia Hirschberg. 1993. A speech-first model for repair detection and correction. *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pp. 46–53.

Narayan, Srikanth & Alexandros Potamianos. 2002. Creating Conversational Interfaces for Children. *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 2, pp. 65–78.

Nass, Clifford & Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, vol. 56, no. 1, pp. 81–103.

Neelley, James N. 1961. A Study of the Speech Behavior of Stutterers and Nonstutterers under Normal and Delayed Auditory Feedback. *Journal of Speech and Hearing Disorders. Monograph Supplement*, vol. 7, pp. 63–82.

Neelley, James N. & Roy J. Timmons. 1967. Adaptation and constistency in the disfluent speech behavior of young stutterers and nonstutterers. *Journal of Speech and Hearing Research*, vol. 10, pp. 250–256.

Neely, James H. 1977. Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention. *Journal of Experimental Psychology*, vol. 106, no. 3, pp. 226–254.

Neilson, Megan D. & Peter D. Neilson. 1987. Speech motor control and stuttering: a computational model of adaptive sensory-motor processing. *Speech Communication*, vol. 6, pp. 325–333.

Nelson, R. J. 1985. Libet's dualism. *Behavioral and Brain Sciences*, vol. 8, p. 550.

Neville, Helen J. 1985. Brain potentials reflect meaning in language. *Trends in Neuroscience*, vol. 8, pp. 91–92.

Neville, Helen J. 1980. Event-related Potentials in Neuropsychological Studies of Language. *Brain and Language*, vol. 11, pp. 300–318.

Newell, A[lan]. F. 1989. Speech Simulation Studies – Performance and Dialogue. In: Jeremy Peckham (ed.), *Recent Developments and Applications of Natural Language Understanding*. Unicom Seminar, December 1987, London, England, London: Kogan Page, ch. 10, pp. 141–157.

Newell, A[lan]. F. 1984. Speech – the natural modality for man–machine interaction? In: B. Shackel (ed.), *Proceedings of INTERACT '84, the First Conference on Human–Computer Interaction*, 4–7 September 1984, London, England. North-Holland: Amsterdam, pp. 231–235.

Newkirk, Don, Edward S. Klima, Carlene Canday Pedersen & Ursula Bellugi. 1980. Linguistic evidence from slips of the hand. In: Victoria A. Fromkin (ed.), *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, ch. 13, pp. 165–197.

Nicolosi, Lucille, Elizabeth Harryman & Janet Kresnick. 1978/1983/1989. *Terminology of communication disorders: Speech–Language–Hearing*. Baltimore: Williams & Wilkins.

Nickerson, R. S. 1969. Man–Computer Interaction: A Challenge for Human Factors Research. *Ergonomics*, vol. 12, no. 4, pp. 501–517.

Nisbett, Richard E. & Timothy DeCamp Wilson. 1977. Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, vol. 84, no. 3, pp. 231–259.

Nivre, Joakim. 1999. Transcription Standard. Version 6.2. In: *Swedish Dialogue Systems (SDS). HSRF/NUTEK. A Platform for Multimodal Spoken Language Corpora*, Dec[ember] 31, 1999, Department of Linguistics, Göteborg University, Sweden. [No page numbers.]

Nivre, Joakim, Jens Allwood, Jenny Holm, Dario Lopez-Kasten, Kristina Tullgren, Elisabeth Ahlsén, Leif Grönqvist & Sylvia Sofkova. 1999. Towards Multimodal Spoken Language Corpora: TransTool and SyncTool. In: *Swedish Dialogue Systems (SDS). HSRF/NUTEK. A Platform for Multimodal Spoken Language Corpora*, Dec[ember] 31, 1999, Department of Linguistics, Göteborg University, Sweden. [No page numbers.]

Nivre, Joakim, Jens Allwood & Elisabeth Ahlsén. 1999. Interactive Communication Management. Coding Manual V1.0. In: *Swedish Dialogue Systems (SDS). HSRF/NUTEK. A Platform for Multimodal Spoken Language Corpora*, Dec[ember] 31, 1999, Department of Linguistics, Göteborg University, Sweden. [No page numbers.]

Nolan, Francis & Esther Grabe. 1997. Can "ToBI" Transcribe Intonational Variation in British English? *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications*, 18–20 September 1997, Athens, Greece, pp. 259–262.

Nooteboom, Sieb G. 1980. Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech. In: Victoria A Fromkin (ed.). *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, ch. 6, pp. 87–95.

Nooteboom, Sieb G. 1973. The tongue slips into patterns. In: Victoria A. Fromkin (ed.), *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 144–156. Same as Nooteboom (1969).

Nooteboom. Sieb G. 1969. The Tongue Slips into Patterns. In: A. G. Sciarone (ed.), *Nomen: Leyden Studies in Linguistics and Phonetics*, The Hague: Mouton, pp. 114–132.

Norman, Donald A. 1981a. Categorization of Action Slips. *Psychological Review*, vol. 88, pp. 1–15.

Norman, Donald A. 1981b. A psychologist views human processing: human errors and other phenomena suggest processing mechanisms. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence, (IJCAI) '81*, 24–28 August 1981, Vancouver, British Columbia, Canada, vol. II, pp. 1097–1101.

Nørretranders, Tor. 1991/1993 (Swedish edition). *Märk världen*. Stockholm: Månpocket.

Novick, Barbara, Deborah Lovrich & Herbert G. Vaughan, Jr. 1985. Event-related potentials associated with the discrimination of acoustic and semantic aspects of speech. *Neurophysiologia*, vol. 23, no. 1, pp. 87–101.

Obeso, J. A., J. C. Rothwell & C. D. Marsden. 1981. Simple tics in Gilles de la Tourette's syndrome are not prefaced by a normal premovement EEG potential. *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 44, pp. 735–738.

Ochsman, Robert B. & Alphonse Chapanis. 1974. The Effects of 10 Communication Modes on the Behavior of Teams During Co-operative Problem-solving. *International Journal of Man–Machine Studies*, vol. 6, pp. 579–619.

Onslow, Mark. 1995. A Picture Is Worth More Than Any Words. *Journal of Speech and Hearing Research*, vol. 38, no. 3, pp. 586–588.

Onslow, Mark, Kate Gardner, Kathryn M. Bryant, Cathi L. Stuckins & Tamsin Knight. Stuttered and Normal Speech Events in Early Childhood: The Validity of a Behavioral Data Language. 1992. *Journal of Speech and Hearing Research*, vol. 35, pp. 79–87.

Opperman, Daniela, Florian Schiel, Silke Steininger & Nicole Beringer. 2001. Off-Talk – a Problem for Human–Machine-Interaction? *Proceedings of Eurospeech 2001*, 3–7 September 2001, Aalborg, Denmark, vol. 3, pp. 2197–2200.

Orton, Samuel T. 1927. Studies in stuttering. *Archives of Neurology and Psychiatry*, vol. 18, pp. 671–672.

O'Shaughnessy, Douglas. 1999. Better detection of hesitations in spontaneous speech. *Proceedings of Disfluency in Spontaneous Speech Workshop*, 1 July 1999, Berkeley, California, USA, pp. 39–42.

O'Shaughnessy, Douglas. 1994. Correcting complex false starts in spontaneous speech. *The International Conference on Acoustics, Speech & Signal Processing (ICAASP)'94*, 19–20 April 1994, Adelaide, Australia, vol. 1, pp. 349–352.

O'Shaughnessy, Douglas. 1993. Analysis and automatic recognition of false starts in spontaneous speech. *The International Conference on Acoustics, Speech & Signal Processing (ICAASP)'93*, 27–30 April 1993, Minneapolis, Minnesota, USA, vol. 2, pp. 724–727.

O'Shaughnessy, Douglas. 1992a. Acoustical analysis of false starts in spontaneous speech. *Proceedings of 124th ASA Meeting*, 31 October–4 November, New Orleans, Louisiana, USA, *Journal of the Acoustical Sociey of America (JASA)*, vol. 92, no. 4, pt. 2, p. 2341.

O'Shaughnessy, Douglas. 1992b. Analysis of false starts in spontaneous speech. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '92*, 12–16 October 1992, Banff, Canada, vol. 2, pp. 931–934.

O'Shaughnessy, Douglas. 1992c. Recognition of hesitations in spontaneous speech. *The International Conference on Acoustics, Speech & Signal Processing (ICAASP)'92*, 23–26 March 1992, San Francisco, California, USA, vol. 1, pp. 521–524.

Ostendorf, Mari, Patti Price, Stefanie Shattuck-Hufnagel. 1997. *Evaluating the Use of Prosodic Information in Speech Recognition and Understanding*. Final Report submitted to National Science Foundation and Advanced Research Projects Administration, April 1997. Boston University.

Oviatt, Sharon. 2000. Talking to thimble jellies: Children's conversational speech with animated characters. *Proceedings of Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2000*, 16–20 October 2000, Beijing, China, vol. 3, pp. 877–880.

Oviatt, Sharon. 1995. Predicting spoken disfluencies during human–computer interaction. *Computer Speech and Language*, vol. 9, pp. 19–35.

Oviatt, Sharon, Margaret MacEachern & Gina-Anne Levow. 1998. Predicting hyperarticulate speech during human–computer error resolution. *Speech Communication*, vol. 24, no. 2, pp. 1–23.

Paivio, Allan. 1965. Personality and audience influence. In: B[rendan] A. Maher (ed.), *Progress in experimental personality research*, New York: Academic Press, pp. 127–173.

Panek, David M. & Barclay Martin. 1959. The relationship between GSR and speech disturbances in psychotherapy. *Journal of Abnormal and Social Psychology*, vol. 58, pp. 402–405.

Paulesu, E., C[hris]. D. Frith & R. S. J. Frackowiak. 1993. The neural correlates of the verbal component of working memory. *Nature*, vol. 362, pp. 342–345.

Pearce, Toby M. 2003. Did they talk their way out of Africa? *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 235–236.

Peckham, Jeremy. 1990. Speech Understanding and Dialogue over the Telephone: an overview of the ESPRIT SUNDIAL project. *Acoustics Bulletin*, vol. 15, pp. 12–19.

Pedersen, Arve Vorland & Beatrix Vereijken. 2003. Laterality probabilities fluctuate during ontogenetic development. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 236–237.

Penfield, Wilder & Phanor Perot. 1963. The brain's record of auditory and visual experience. *Brain*, vol. 86, pt. 4, pp. 595–696.

Penrose, Roger. 1990. Précis of The Emperor's New Mind: Concerning computers, minds and the laws of physics. *Behavioral and Brain Sciences*, vol. 13, pp. 643–705.

Penrose, Roger. 1989. *The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics*. New York: Oxford University Press.

Perkins, William H. 1990. What is stuttering? *Journal of Speech and Hearing Disorders*, vol. 55, pp. 370–382.

Perkins, William H. 1983. The problem of definition: Commentary on "Stuttering". *Journal of Speech and Hearing Disorders*, vol. 48, pp. 246–249.

Perkins, William H., Joanna Rudas, Linda Johnson & Jody Bell. 1976. Stuttering: discoordination of phonation with articulation and respiration. *Journal of Speech and Hearing Research*, vol. 19, pp. 509–522.

Peters, Herman F. M., Wouter Hulstijn & C. Woodruff Starkweather. 1989. Acoustic and physiological reaction times of stutterers and nonstutterers. *Journal of Speech and Hearing Research*, vol. 32, pp. 668–680.

Picton, Terence & Jerome Cohen. 1984. Event-Related Potentials: Whence? Where? Whither? *Annals of the New York Academy*, vol. 425, pp. 753–765.

Pinker, Steven. 1995. *The Language Instinct. How the Mind Creates Language*. New York: HarperPerennial. [Originally published: New York: W. Morrow & Co., 1994.]

Pitrelli, John F., Mary E. Beckman & Julia Hirschberg. 1994. Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '94*, 18–22 September 1994, Yokohama, Japan, vol. 1, pp. 123–126.

Plauché, Madelaine & Elizabeth Shriberg. 1999. Data-driven subclassification of disfluent repetitions based on prosodic features. *Proceedings of the International Congress of Phonetic Sciences (ICPhS) 1999*, 1–7 August 1999, San Francisco, California, USA, vol. 2, pp. 1513–1516.

Pockett, Susan. 2002a. On Subjective Back-Referral and How Long It Takes to Become Conscious of a Stimulus: A reinterpretation of Libet's Data. *Consciousness and Cognition*, vol. 11, pp. 144–161.

Pockett, Susan. 2002b. Backward Referral, Flash-lags, and Quantum Free Will: A Response to Commentaries on Articles by Pockett, Klein, Gomes, and Trevena and Miller. *Consciousness and Cognition*, vol. 11, pp. 314–325.

Pope, Benjamin, Thomas Blass, Aron W[olfe]. Siegman & Jack Raher. 1970. Anxiety and depression in speech. *Journal of Consulting and Clinical Psychology*, vol. 35, no. 1, pp. 128–133.

Pope, Benjamin & Aron W[olfe]. Siegman. 1962. The effect of therapist verbal activity level and specificity on patient productivity and speech disturbance in the initial interview. *Journal of Consulting Psychology*, vol. 26, no. 5, p. 489.

Posey, Thomas B. 1986. Verbal hallucinations also occur in normals. *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 530.

Posey, Thomas B. & Mary E. Losch. 1983. Auditory hallucinations of hearing voices in 375 normal subjects. *Imagination, Cognition and Personality*, vol. 3, no. 2, pp. 99–113.

Posner, Michael I. & Charles R. R. Snyder. 1975. Facilitation and Inhibition in the Processing of Signals. In: P. M. A. Rabbitt & S. Dornick (eds.), *Attention and Performance V*, pp. 669–682.

Postma, Albert. 2000. Detection of errors during speech production: a review of speech monitoring models. *Cognition*, vol. 77, pp. 97–131.

Postma, Albert & Herman Kolk. 1993. The Covert Repair Hypothesis: Prearticulatory Repair Processes in Normal and Stuttered Disfluencies. *Journal of Speech and Hearing Research*, vol. 36, pp. 472–487.

Postma, Albert & Herman Kolk. 1992. The Effects of Noise Masking and Required Accuracy on Speech Errors, Disfluencies, and Self-Repairs. *Journal of Speech and Hearing Research*, vol. 35, pp. 537–544.

Postma, Albert, Herman Kolk & Dirk-Jan Povel. 1991. Disfluencies as resulting from covert self-repairs applied to internal speech errors. In: Herman F. M. Peters, Wouter Hulstijn & C. Woodruff Starkweather (eds.), *Speech Motor Control and Stuttering*, ch. 10, pp. 141–147.

Postma, Albert & Herman Kolk. 1990. Speech Errors, Disfluencies, and Self-Repairs of Stutterers in Two Accuracy Conditions. *Journal of Fluency Disorders*, vol. 15, pp. 291–303.

Postma, Albert, Herman Kolk & Dirk-Jan Povel. 1990. On the Relation Among Speech Errors, Disfluencies, and Self-Repairs. *Language and Speech*, vol. 33, no 1, pp. 19–29.

Prescott, John. 1988. Event-related potential indices of speech motor programming in stutterers and non-stutterers. *Biological Psychology*, vol. 27, pp. 259–273.

Prins, David. 1991. Theories of stuttering as event and disorder: implications for speech production theories. In: Herman F. M. Peters, Wouter Hultijn & C. Woodruff Starkweather (eds.), *Speech motor control and stuttering*. New York: Elsevier, ch. 54, pp. 571–580.

Prosek, Robert A., Allen A. Montgomery, Brian E. Walden & Daniel M. Schwartz. 1979. Reaction-Time Measures of Stutterers and Nonstutterers. *Journal of Fluency Disorders*, vol. 4, pp. 269–278.

Quesal, Robert W. 1988. Inexact Use of "Disfluency" and "Dysfluency" in Stuttering Research. *Journal of Speech and Hearing Disorders*, vol. 53, no. 3. pp. 349–351.

Ragsdale, J. Donald & Catherine Fry Silvia. 1982. Distribution of kinesic hesitation phenomena in spontaneous speech. *Language and Speech*, vol. 25, pt. 2, pp. 185–190.

Ramig, Peter & Martin R. Adams. 1980. Rate Reduction Strategies Used by Stutterers and Nonstutterers During High- and Low-Pitched Speech. *Journal of Fluency Disorders*, vol. 5, pp. 27–41.

Random House dictionary of the English Language (unabridged). 1987. New York: Random House.

Rapp, Brenda & Matthew Goldrick. 2000. Discreteness and Interactivity in Spoken Word Production. *Psychological Review*, vol. 107, no. 3, pp. 460–499.

Ratner, Nan Bernstein. 1988. Response to Quesal: Terminology in Stuttering Research. *Journal of Speech and Hearing Disorders*, vol. 53, pp. 350–351.

Ratner, Nan Bernstein & Catherine Costa Sih. 1987. Effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders*, vol. 52, pp. 278–287.

Ratner, Nan Bernstein & Mercedes Benitez. 1985. Linguistic analysis of a bilingual stutterer. *Journal of Fluency Disorders*, vol. 10, pp. 211–219.

Raupach, Manfred. 1980. Temporal variables in first and second language speech production. In: Hans W. Dechert & Manfred Raupach (eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, pp. 263–270.

Rayner, Manny, Beth Ann Hockey, Jim Hieronymus, John Dowding, Greg Aist & Susana Early. 2003. An Intelligent Procedure Assistant Built Using `regulus 2` and `alterf`. *Proceedings of the 41st Annual Meeting for Computational Linguistics*, 7–12 July 2003, Sapporo, Japan, pp. 193–196.

Rayner, Manny, Dave Carter, Pierrette Bouillon, Vassilis Digalakis & Mats Wirén (eds.), 2000. *The Spoken Language Translator*, Cambridge: Cambridge University Press.

Raz, Amir & Opher Donchin. 2003. A zetetic's perspective on gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 237–238.

Razzak, Lailaa & Nan Bernstein Ratner. 1999. Auditory Feedback Responses of Young Fluent and Stuttering Children. *ASHA Reports* (American Speech–Language–Hearing Association), vol. 41, p. 89.

Reeves, Byron & Clifford Nass. 1996. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Stanford: CSLI Publications.

Reich, Alan, James Till & Howard Goldsmith. 1981. Laryngeal and manual reaction times of stuttering and nonstuttering adults. *Journal of Speech and Hearing Research*, vol. 24, pp. 192–196.

Reilly, Ronan. 1987. Ill-formedness and miscommunication in person–machine dialogue. *Information and Software Technology*, vol. 29, pp. 69–74.

Rialland, Annie & Stéphane Robert. 2001. The intonational system of Wolof. *Linguistics*, vol. 39, pp. 893–939.

Richards, M. A. & K. M. Underwood. 1985. How should people and computers speak to each other? In: B. Shackel (ed.), *Proceedings of INTERACT '84, The First Conference on Human–Computer Interaction*, 4–7 September 1984, London, England. Elsevier Science Publishers B.V. (North-Holland): Amsterdam, pp. 215–218.

Richards, M. A. & K. M. Underwood. 1984. Talking to machines. How are people naturally inclined to speak? In: E. D. Megaw (ed.), *Contemporary Ergonomics*, London: Taylor and Francis, pp. 62–67.

Rieger, Caroline L. 2003. Disfluencies and hesitation strategies in oral L2 test. *Proceedings of DiSS '03, Disfluency in Spontaneous Speech Workshop*, Robert Eklund (ed.), *Gothenburg Papers in Theoretical Linguistics 90*, ISSN 0349–1021, pp. 41–44.

Rimé, Bernard & Loris Schiaratura. 1991. Gesture and speech. In: Robert S. Feldman & Bernard Rimé (eds.), *Fundamentals of nonverbal behavior*. Cambridge: Cambridge University Press, ch. 7, pp. 239–281.

Ringel, Robert L. & Fred D. Minifie. 1966. Protensity estimates of stutterers and nonstutterers. *Journal of Speech and Hearing Research*, vol. 9, pp. 289–296.

Ringo, James L. 1985. Timing volition: Questions of what and when about W. *Behavioral and Brain Sciences*, vol. 8, pp. 550–551.

Riper, Charles van. See: Van Riper, Charles.

Roach, Peter. 1994. Conversion between prosodic transcription systems: "Standard British" and ToBI. *Speech Communication*, vol. 15, pp. 91–99.

Rochester, Sherry R. 1975/1976. Defining the silent pause in speech. *Journal of the Ontario Speech and Hearing Association*, vol. VIII, pp. 1–4.

Rochester, S[herry]. R. 1973. The Significance of Pauses in Spontaneous Speech. *Journal of Psycholinguistic Research*, vol. 2, no. 1, pp. 51–81.

Röder, Brigitte, Frank Rösler & Helen J. Neville. Event-related potentials during auditory language processing in congenitally blind and sighted people. *Neuropsychologia*, vol. 38, pp. 1482–1502.

Rohrbaugh, John W., Karl Syndulko & Donald B. Lindsley. 1976. Brain Wave Components of the Contingent Negative Variation in Humans. *Science*, vol. 191, pp. 1055–1057.

Rollman, Gary B. 1985. Sensory events with variable central latencies provide inaccurate clocks. *Behavioral and Brain Sciences*, vol. 8, pp. 551–552.

Rönnqvist, Louise. 2003. Developmentally, the arm preference precedes handedness. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 238–239.

Rosenberg, Seymour & Bertram D. Cohen. 1966. Referential processes of speakers and listeners. *Psychological Review*, vol. 73, vol. 3, pp. 208–231.

Rosenfield, David B. & Harvey B. Nudelman. 1987. Neuropsychological Models of Speech Dysfluency. In: Lena Rustin, Harry Purser & David Rowley (eds.), *Progress in the Treatment of Fluency Disorders*. London: Taylor & Francis, ch. 1, pp. 3–18.

Rosenthal, David M. 2002. The Timing of Conscious States. *Consciousness and Cognition*, vol. 11, pp. 215–220.

Rosenthal, David M. 1992. Time and consciousness. *Behavioral and Brain Sciences*, vol. 15, pp. 220–221.

Rothenberger, Aribert, Berta Grözinger, Alexander Foit & Wolfgang Woerner. 1987. Do event-related potentials (ERPs) reflect right hemisphere's processing of semantics? *International Journal of Neuroscience*, vol. 33, pp. 93–101.

Rothwell, John. 1998. Transcranial magnetic stimulation. *Brain*, vol. 121, pt. 3, pp. 397–398.

Rubino, Carl. 1998. The morphological realization and production of a nonprototypical morpheme: the Tagalog derivational clitic. *Linguistics*, vol. 36, pp. 1147–1166.

Rubino, Carl. 1996. Morphological integrity in Ilocano: a corpus-based study of the production of polymorphemic words in a polymorphic language. *Studies in Language*, vol. 20, no. 3, pp. 633–666.

Rugg, Michael D. 1985. Are the origins of any mental process available to introspection? *Behavioral and Brain Sciences*, vol. 8, pp. 552.

Rugg, Michael D. 1984. Event-related potentials and the phonological processing of words and non-words. *Neurophysiologia*, vol. 22, no. 4, pp. 435–443.

Rumelhart, David E. & James L. McClelland (eds.). 1986. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. Cambridge, Massachussetts: The MIT Press.

Rumsey, J. M., B. Horwitz, B. C. Donohue, K. Nace, J. M. Maisog & P. Andreason. 1997. Phonological and orthographic components of word recognition. *Brain*, vol. 120, pt. 5, pp. 739–759.

Ryan, Bruce P. 1992. Articulation, Language, Rate, and Fluency Characteristics of Stuttering and Nonstuttering Preschool Children. *Journal of Speech and Hearing Research*, vol. 35, pp. 333–342.

Sacks, Harvey, Emanuel A. Schegloff & Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, vol. 50, pp. 696–735.

Salmelin, R., A. Schnitzler, F. Schmitz & H.-J. Freund. 2000. Single word reading in developmental stutterers and fluent speakers. *Brain*, vol. 123, pt. 6, pp. 1184–1202.

Sarason, Irwin G. 1981. Test anxiety, stress, and social support. *Journal of Personality*, vol. 49, pp. 101–114.

Schachter, Stanley, Nicholas Christenfeld, Bernard Ravina & Frances Bilous. 1991. Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*, vol. 60, no. 3, pp. 362–367.

Schafer, Edward W. P. 1967. Cortical Activity preceding Speech: Semantic Specificity. *Nature*, vol. 216, pp. 1338–1339.

Schäfersküpper, Paul & Thomas Simon. 1983. The Mean Fundamental Frequency in Stutterers and Nonstutterers During Reading and Spontaneous Speech. *Journal of Fluency Disorders*, vol. 8, pp. 125–132.

Scheerer, Eckhart. 1985. Conscious intention to act is a mental fiat. *Behavioral and Brain Sciences*, vol. 8, no. 4, pp. 552–553.

Schegloff, Emanuel A. 1979. The relevance of repair to syntax-for-conversation. In: T. Givón (ed.), *Syntax and Semantics, vol. 12, Discourse and Syntax*. New York: Academic Press, pp. 261–286.

Schegloff, Emanuel A., Gail Jefferson & Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, vol. 53, no. 2, pp. 361–382.

Schegloff, Emanuel A. & Harvey Sacks. 1973. Opening up Closings. *Semiotica*, vol. 8, pp. 289–327.

Schlenk, Klaus-Jürgen, Walter Huber & Klaus Willmes. 1987. "Prepairs" and Repairs: Different Monitoring Functions in Aphasic Language Production. *Brain and Language*, vol. 30, pp. 226–244.

Schmeer, Bill. 2003. Speech! Speech! *Discover*, vol. 24, no. 3, March 2003, letters, p. 8.

Schneiderman, Ben. 2000. The limits of speech recognition. *Communications of the ACM*, vol. 43, no. 9, pp. 63–35.

Schönle, P. W. & B. Conrad. 1985. Hesitation vowels: a motor speech respiration hypothesis. *Neuroscience Letters*, vol. 55, pp. 293–296.

Schuckers, Gordon H., & Carol S. Lefkov. 1979. Children's perception of misarticulations in contextual speech. *Journal of Phonetics*, vol. 7, pp. 177–186.

Schulze, Gene. 1987. Speech Disturbances During Induced Psychological Stress. In: George F. Mahl (ed.), *Explorations in Nonverbal and Vocal Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum, ch. 14, pp. 223–244.

Schultz, Wolfram. 1999. The Primate Basal Ganglia and the Voluntary Control of Behaviour. *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 31–45. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 31–45.

Schumacher, E. F. 1977. *A Guide for the Perplexed*. London: Abacus.

Schwartz, Steven. 1986. Hallucination, rationalization, and response set. *Behavioral and Brain Sciences*, vol. 9, no. 3, pp. 532–533.

Searle, John. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Sechenov, I. M. 1863/1935. *Collected works*. Moscow: State Publishing House. Includes the 1863 refererence.

Sellen, Abigail J. & Donald A. Norman. 1992. The Psychology of Slips. In: Bernard J. Baars (ed.), *Experimental Slips and Human Error*, New York & London: Plenum Press, ch. 13, pp. 317–339.

Sereno, Kenneth K. & Gary J. Hawkins. 1967. The effects of variations in speaker's nonfluency upon audience ratings of attitude toward the speech topic and speaker's credibility. *Speech Monographs*, vol. 34, pp. 58–64.

Serzisko, Fritz. 1992. *Sprechhandlungen und Pausen*. Tübingen: Max Niemeyer Verlag.

Seyfeddinipur, Mandana & Sotaro Kita. 2001. Gesture as an Indicator of Early Error Detection in Self-Monitoring of Speech. *Proceedings of DiSS '01 Disfluency in Spontaneous Speech*, 29–31 August 2001, University of Edinburgh, Scotland, pp. 29–32.

Shames, George H. & Carl E. Sherrick, Jr. 1963. A Discussion of Nonfluency and Stuttering as Operant Behavior. *Journal of Speech and Hearing Disorders*, vol. 28, pp. 3–18.

Shane, Mary Lou Sternberg. 1955. Effect on stuttering of alteration in auditory feedback. In: Wendell Johnson (ed.), *Stuttering in Children and Adults*. Minneapolis: University of Minnesota Press, pp. 286–297.

Shannon, C[laude]. E. 1951. Prediction and Entropy of Printed English. *Bell System Technical Journal*, vol. 30, pp. 50–64.

Shapiro, Arnold I. & Barbara A. DeCicco. 1982. The relationship between normal dysfluency and stuttering: An old question revisited. *Journal of Fluency Disorders*, vol. 7, pp. 109–121.

Shattuck-Hufnagel, Stefanie. 1979. Speech Errors as Evidence for a Serial-Ordering Mechanism in Sentence Production. In: W. E. Cooper & E. C. T. Walker (eds.), *Psycholinguistic Studies Presented to Merrill Garrett*, Hillsdale, New Jersey: Erlbaum, pp. 295–342.

Shattuck-Hufnagel, Stefanie & Dennis H. Klatt. 1980. How single phoneme error data rule out two models of error generation. In: Victoria A. Fromkin (ed.). *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, ch. 2, pp. 35–46.

Sheehan, Joseph [G]. 1958. Conflict theory of stuttering. In: J. Eisenson (ed.), *Stuttering: A Symposium*, NewYork: Harper, pp. 123–166.

Shergill, Sukhwinder, S., Michael J. Brammer, Rimmei Fukuda, Steven C. R. Williams, Robin M. Murray & Philip K. McGuire. 2003. Engagement of brain areas implicated in processing inner speech in people with auditory hallucinations. *British Journal of Psychiatry*, vol. 182, pp. 525–531.

Sherrick, Carl E. & Ronald Rogers. 1966. Apparent Haptic Movement. *Perception and Psychophysics*, vol. 1, pp. 175–180.

Shillcock, Richard, Simon Kirby, Scott McDonald & Chris Brew. 2001. Filled pauses and their status in the mental lexicon. *Proceedings of DiSS '01 Disfluency in Spontaneous Speech*, 29–31 August 2001, University of Edinburgh, Scotland, pp. 53–56.

Shriberg, Elizabeth. 2001. To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, vol. 31, no. 1, pp. 153–169.

Shriberg, E[lizabeth] E. 1995. Acoustic properties of disfluent repetitions. *Proceedings of the International Congress of Phonetic Sciences (ICPhS) 1995*, 13–19 August 1995, Stockholm, Sweden, vol. 4, pp. 384–387.

Shriberg, Elizabeth Ellen. 1994. *Preliminaries to a Theory of Speech Disfluencies*. PhD thesis, University of California, Berkeley.

Shriberg, Elizabeth & Andreas Stolcke. 2004. Prosody modeling for automatic speech recognition and understanding. In: M. Johnson, S. Khudanpur, M. Ostendorf & R. Rosenfeld (eds.), *Mathematical Foundations of Speech and Language Processing. IMA Volumes in Mathematics and its Applications*, New York: Springer-Verlag, vol. 138, pp. 105–113.

Shriberg, Elizabeth, Andreas Stolcke & Don Baron. 2001. Can Prosody Aid the Automatic Processing of Multi-Party Meetings? Evidence from Multi-Party Punctuation, Disfluencies, and Overlapping Speech. *Proceedings of the ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*, Red Bank, New Jersey, USA, pp. 139–146.

Shriberg, Elizabeth, Andreas Stolcke, Dilek Hakkani-Tür & Gökhan Tür. 2000. Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, vol. 32, pp. 127–154.

Shriberg, Elizabeth, Rebecca A. Bates & Andreas Stolcke. 1997. A prosody-only decision-tree model for disfluency detection. *Proceedings of Eurospeech '97*, 22–25 September 1997, Rhodes, Greece, vol. 5, pp. 2383–2386.

Shriberg, Elizabeth, D. Robert Ladd, Jacques Terken & Andreas Stolcke. 1996. Modeling Pitch Range Variation Within and Across Speakers: Predicting F0 Targets When "Speaking Up". *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '96*, 3–6 October 1996, Philadelphia, Pennsylvania, USA, Addendum, pp. 1–4.

Shriberg, Elizabeth, Rebecca A. Bates & Andreas Stolcke. 1996. Integrated acoustic and language modeling of speech disfluencies. *Journal of the Acoustical Society of America*, vol. 100, no. 4, pt. 2, p. 2848 [abstract].

Shriberg, Elizabeth & Andreas Stolcke. 1996. Word predictability after hesitations: A corpus-based study. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '96*, 3–6 October 1996, Philadelphia, Pennsylvania, USA, vol. 3, pp. 1868–1871.

Shriberg, Elizabeth E. & Robin J. Lickley. 1993. Intonation of Clause-Internal Filled Pauses. *Phonetica*, vol. 50, pp. 172–179.

Shriberg, Elizabeth E. & Robin J. Lickley. 1992a. The Relationship of Filled-Pause F0 to Prosodic Context. *Proceedings of the IRCS Workshop on Prosody in Natural Speech*, Philadelphia, Pennsylvania, USA, pp. 201–209.

Shriberg, Elizabeth & Robin [J.] Lickley. 1992b. Intonation of clause-internal filled pauses. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '92*, 12–16 October 1992, Banff, Alberta, Canada, vol. 2, pp. 991–994.

Shriberg, Elizabeth, John Bear & John Dowding. 1992. Automatic Detection and Correction of Repairs in Human–Computer Dialog. In: Mitchell Marcus (ed.): *Proceedings of DARPA Speech and Natural Language Workshop*, Morgan Kaufmann, pp. 419–424.

Siegel, Gerald M., Joanne Lenske & Patricia Broen. 1969. Suppression of normal speech disfluencies through response costs. *Journal of Applied Behavior Analysis*, vol. 2, p. 265–276.

Siegel, Gerald M. & Richard R. Martin. 1967. Verbal punishment of disfluencies during spontaneous speech. *Language and Speech*, vol. 10, part 4, pp. 244–251.

Siegel, Gerald M. & Richard R. Martin. 1966. Punishment of disfluencies in normal speakers. *Journal of Speech and Hearing Research*, vol. 9, pp. 208–218.

Siegel, Gerald M. & Richard R. Martin. 1965a. Experimental modification of disfluency in normal speakers. *Journal of Speech and Hearing Research*, vol. 8, pp. 235–244.

Siegel, Gerald M. & Richard R. Martin. 1965b. Verbal punishment of disfluencies in normal speakers. *Journal of Speech and Hearing Research*, vol. 8, pp. 245–251.

Siegman, Aron Wolfe & Benjamin Pope. 1966. Ambiguity and verbal fluency in the TAT. *Journal of Consulting Psychology*, vol. 30, no. 3, pp. 239–245.

Siegman, Aron Wolfe & Benjamin Pope. 1965a. Effects of question specificity and anxiety-producing messages on verbal fluency in the initial interview. *Journal of Personality and Social Psychology*, vol. 2, no. 4, pp. 522–530.

Siegman, Aron Wolfe & Benjamin Pope. 1965b. Personality variables associated with the productivity and verbal fluency in the initial interview. *Proceedings of the 73rd Annual Convention of the American Psychological Association*, Washington, D.C., pp. 273–274.

Silverman, Ellen-Marie. 1974, Word position and grammatical function in relation to preschoolers' speech disfluency. *Perceptual and Motor Skills*, vol. 39, pp. 267–272.

Silverman, Ellen-Marie. 1973a. Clustering: a characteristic of preschoolers' speech disfluency. *Journal of Speech and Hearing Research*, vol. 16, no. 4, pp. 578–583.

Silverman, Ellen-Marie. 1973b. The influence of preschoolers' speech usage on their disfluency frequency. *Journal of Speech and Hearing Research*, vol. 16, no. 3, pp. 474–481.

Silverman, Ellen-Marie. 1972. Generality of disfluency data collected from preschoolers. *Journal of Speech and Hearing Research*, vol. 15, no. 1, pp. 84–92.

Silverman, Ellen-Marie. 1971. Situational variability of preschooler's disfluency: Preliminary Study. Perceptual and Motor Skills, vol. 33, pp. 1021–1022.

Silverman, Ellen-Marie & Catherine H. Zimmer. 1975. Speech fluency fluctuations during the menstrual cycle. *Journal of Speech and Hearing Research*, vol. 18, no. 1, pp. 202–206.

Silverman, Ellen-Marie, Catherine H. Zimmer & Franklin H. Zimmerman. 1974. Varibility in stutterers' speech disfluency: the mentrual cycle. *Percepual and Motor Skills*, vol. 38, pp. 1037–1038.

Silverman, Franklin H. 1995. Can Disfluencies Be Categorized Reliably Using Wendell Johnson's Scheme. *Journal of Speech and Hearing Research*, vol. 38, no. 3, pp. 586.

Silverman, Franklin H. 1992. *Stuttering and Other Fluency Disorders*. Englewood Cliffs, New Jersey: Prentice Hall.

Silverman, Franklin H. 1988. The "Monster" Study. *Journal of Fluency Disorders*, vol. 13, pp. 225–231.

Silverman, Franklin H. 1974. Disfluency Behavior of Elementary-School Stutterers and Nonstutterers. *Language, Speech, and Hearing Services in Schools*, vol. 5, pp. 32–37.

Silverman, Franklin H. 1972. Disfluency and word length. *Journal of Speech and Hearing Research*, vol. 15, no. 4, pp. 788–791.

Silverman, Franklin H. & Marjorie Tylke Goodban. 1972. The effect of auditory masking on the fluency of normal speakers. *Journal of Speech and Hearing Research*, vol. 15, pp. 543–546.

Silverman, Franklin H. & Dean E. Williams. 1967a. Loci of disfluencies in the speech of stutterers. *Perceptual and Motor Skills*, vol. 24, pp. 1085–1086.

Silverman, Franklin H. & Dean E. Williams. 1967b. Loci of disfluencies in the speech of nonstutterers during oral reading. *Journal of Speech and Hearing Research*, vol. 10, no. 4, pp. 790–794.

Silverman, Gerald. 1973. Redundancy, repetition and pausing in schizophrenic speech. *British Journal of Psychiatry*, vol. 122, pp. 407–413.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Osterdorf, Colin Wightman, Patti Price, Janet Pierrehumbert & Julia Hirschberg. 1992. TOBI: A Standard for Labeling English Prosody. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '92*, 12–16 October 1992, Banff, Alberta, Canada, vol. 1, pp. 867–870.

Siu, Man-Hung & Mari Ostendorf. 1996. Modeling Disfluencies in Conversational Speech. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '96*, 3–6 October 1996, Philadelphia, Pennsylvania, USA, vol. 1, pp. 386–389.

Smith, Edward E. 1970. Associative and editing processes in schizophrenic communication. *Journal of Abnormal Psychology*, vol. 75, no. 2, pp. 182–186.

Smith, Hugh. 1980. Human–Computer Communication. In: H[ugh]. T. Smith & T. R. G. Green (eds.), *Human Interaction with Computers*. London: Academic Press, ch. 1, pp. 5–38.

Soderberg, George A. 1967. Linguistic factors in stuttering. *Journal of Speech and Hearing Research*, vol. 10, pp. 801–810.

Sokolov, A[leksandr] N. 1972. *Inner speech and thought*. New York and London: Plenum Press.

Sommer, Iris E. C. & René Kahn. 2003. The left hemisphere as the redundant hemisphere. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 239–240.

Sommer, Iris, André Aleman, Nick Ramsey, Anke Bouma & René Kahn. 2001. Handedness, language lateralisation and anatomical asymmetry in schizophrenia. *British Journal of Psychiatry*, vol. 178, pp. 344–351.

SPSS. **http://www.spss.com/**

Spence, Sean A. 1996. Free Will in the Light of Neuropsychiatry. *Philosophy, Psychiatry and Psychology*, vol. 3, pp. 75–90.

Spence, Sean A., Peter F. Liddle, Martin D. Stefan, Jonathan S. E. Hellewell, Tonmoy Sharma, Karl J. Friston, Steven R. Hirsh, Christopher D. Frith, Robin M. Murray, J. F. William Deakin & Paul M. Grasby. 2000. Functional anatomy of verbal fluency in people with schizophrenia and those at genetic risk. *British Journal of Psychiatry*, vol. 176, pp. 52–60.

Spence, Sean A. & Chris D. Frith. 1999. Towards a Functional Anatomy of Volition. *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 1–10. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 11–29.

Sperry, Roger W. 1980. Mind–Brain Interaction: Mentalism, Yes; Dualism, No. *Neuroscience*, vol. 5, pp. 195–206.

Sperry, Roger W. 1976. Changing concepts of consciousness and free will. *Perspectives in Biology and Medicine*, vol. 20, pp. 9–19.

Sperry, Roger W. 1968. Hemisphere deconnection and unity in conscious awareness. *American Psychologist*, vol. 23, pp. 723–733.

Sperry, Roger W. 1967. Mental unity following surgical disconnection of the hemispheres. *The Harvey Lectures. Series 62*. New York: Academic Press, pp. 293–323.

Sperry, Roger W. 1966. Brain Bisection and Mechanisms of Consciousness. In: J[ohn]. C. Eccles (ed.), *Brain and conscius experience. Study Week September 28 to October 4, 1964, of the Pontifica Academia Scientiarum*, Città del Vaticano. New York: Springer-Verlag, ch. 13, pp. 298–313.

Sperry, Roger W. & Michael S. Gazzaniga. 1967. Language following Surgical Disconnection of the Hemispheres. In: Clark H. Milikan (ed.), *Brain mechanisms underlying speech and language*. New York: Grune & Strutton, pp. 108–121.

Spilker, Jörg, Martin Klarner & Günther Görz. 2000. In: Werner Zühlke, Ernst Günter Schukat-Talamazzini (eds.), *KONVENS 2000 / Sprachkommunikation, Vorträge der gemeinsamen Veranstaltung 5. Konferenz zur Verarbeitung natürlicher Sprache (KONVENS), 6. ITG-Fachtagung "Sprachkommunikation"*, 9. bis 12. Oktober 2000, Technische Universität Ilmenau. VDE Verlag 2000, ISBN 3-8007-2564-9, pp. 27–31.

St. Louis, Kenneth O., Audrey R. Hinzman & Forrest M. Hull. 1985. Studies of cluttering: Disfluency and language measurers in young possible clutterers and stutterers. *Journal of Fluency Disorders*, vol. 10, pp. 151–172.

Staats, Lorin C. Jr. 1955. Sense of Humor in Stutterers and Nonstutterers. In: Wendell Johnson (ed.), *Stuttering in Children and Adults. Thirty Years of Research at the University of Iowa*, Minneapolis: University of Minneapolis Press, ch. 24, pp. 313–316.

Stager, Sheila V. & Christy L. Ludlow. 1993. Speech production changes under fluency-evoking conditions in nonstuttering speakers. *Journal of Speech and Hearing Research*, vol. 36, pp. 245–253.

Stamm, John S. 1985. The uncertainty principle in physiology. *Behavioral and Brain Sciences*, vol. 8, pp. 553–554.

Stapp, Henry P. 2001. Quantum Theory and the Role of Mind in Nature. *Foundations of Physics*, vol. 31, no. 10, pp. 1465–1499.

Stapp, Henry P. 1999. Attention, Intention, and Will in Quantum Physics. *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 143–164. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 143–164.

Starbuck, H. B. & M. D. Steer. 1953. The Adaptation Effect In Stuttering Speech Behavior And Normal Speech Behavior. *Journal of Speech Hearing Disorders*, vol. 18, pp. 252–255.

Starkweather, C. Woodruff. 1987. *Fluency and Stuttering*. Englewood Cliffs, New Jersey: Prentice Hall.

Starkweather, C. Woodruff, Sharon Franklin & Therese M. Smigo. 1984. Vocal and finger reaction times in stutterers and nonstutterers. differences and correlations. *Journal of Speech and Hearing Research*, vol. 27, pp. 193–196.

Starkweather, C. Woodruff, Paula Hirschman & Robert S. Tannenbaum. 1976. Latency of vocalization onset: Stutterers versus nonstutterers. *Journal of Speech and Hearing Research*, vol. 19, pp. 481–492.

Stassi, Eugene J. 1961. Disfluency of Normal Speakers and Reinforcement. *Journal of Speech and Hearing Research*, vol. 4, pp. 358–361.

Steeneken, Hermann J. M. & John H. L. Hansen. 1999. Speech under stress conditions: overview of the effect on speech production and on system performance. *The International Conference on Acoustics, Speech & Signal Processing (ICAASP)'99*, 15–19 March 1999, Phoenix, Arizona, USA, vol. 4, pp. 2079–2082.

Steinhauer, Karsten, Kai Alter & Angela D. Friederici. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, vol. 2, no. 2, pp. 191–196.

Stemberger, Joseph Paul. 1992. The Reliability and Replicability of Naturalistic Speech Error Data. A Comparison with Expreimentally Induced Errors. In: Bernard J. Baars (ed.), *Experimental Slips and Human Error*, New York & London: Plenum Press, ch. 8, pp. 195–215.

Sternberg, Mary Lou Shane. See: Shane, Mary Lou Sternberg

Sternberg, Saul, Stephen Monsell, Ronald L. Knoll & Charles E. Wright. 1978. The Latency and Duration of Rapid Movement Sequences: Comparisons of Speech and Typewriting. In: G. E. Stelmach (ed.), *Information processing in motor control and learning*. Amsterdam: North-Holland, pp. 117–152.

Sternberg, Saul & Ronald L. Knoll. 1973. The Perception of Temporal Order: Fundamental Issues and a General Model. In: Sylvan Kornblum (ed.), *Attention and Performance IV*, New York: Academic Press, pp. 629–685.

Stirling, Lesley, Janet Fletcher, Ilana Mushin & Roger Wales. 2001. Representational issues in annotation: Using the Australian map task corpus to relate prosody and discourse structure. *Speech Communication*, vol. 33, pp. 113–134.

Stolcke, Andreas, Elizabeth Shriberg, Rebecca Bates, Mari Ostendorf, Dilek Hakkani, Madeleine Plauché, Gökhan Tur, & Yu Lu. 1998. Automatic detection of sentence boundaries and disfluencies based on recognized words. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '98*, 30 November–5 December 1998, Sydney, Australia, vol. 5, pp. 2247–2250.

Stolcke, Andreas & Elizabeth Shriberg. 1996. Statistical language modeling for speech disfluencies. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICAASP) '96*, 7–10 May 1996, Atlanta, Georgia, USA, vol. 1, pp. 405–408.

Stoll, François C., Douglas G. Hoecker, Gerald P. Krueger & Alphonse Chapanis. 1976. The effects of four communication modes on the structure of language used during cooperative problem solving. *Journal of Psychology*, vol. 94, pp. 13–26.

Strandburg, Robert J., James T. Marsh, Warren S. Brown, Robert F. Asarnow, Donald Guthrie, Rebecca Harper, Cindy M. Yee & Keith H. Nuechterlein. Event-Related Potential Correlates of Linguistic Information Processing in Schizophrenics. 1997. *Biological Psychiatry*, vol. 42, issue 7, pp. 596–608.

Strauss, Evelyn. 1998. Writing, Speech Separated in Split Brain. *Science*, vol. 280, p. 827.

Streeck, Jürgen. 1996. A little Ilokano grammar as it appears in interaction. *Journal of Pragmatics*, vol. 26, pp. 189–213.

Stromsta, Courtney. 1964. EEG power spectra of stutterers and nonstutterers. *ASHA Reports* (American Speech–Language–Hearing Association), vol. 6, pp. 418–419.

Susca, Michael & E. Charles Healey. 2002. Listener perceptions along a fluency–disfluency continuum: A phenomenological analysis. *Journal of Fluency Disorders*, vol. 27, no. 2. pp. 135–160.

Swerts, Marc, Anne Wichmann & Robbert-Jan Beun. 1996. Filled Pauses as Markers of Discourse Markers. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) '96,* 3–6 October 1996, Philadelphia, Pennsylvania, USA, vol. 2, pp. 1033–1036.

Syrdal, Ann K. & Julia McGory. 2000. Inter-Transcriber Reliability of ToBI Prosodic Labeling. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2000*, 16–20 October 2000, Beijing, China, vol. 3, pp. 235–238.

Szabadi, E., C. M. Bradshaw & J. A. Besson. 1976. Elongation of Pause-Time: A Simple, Objective Measure of Motor Retardation in Depression. *British Journal of Psychiatry*, vol. 129, pp. 592–597.

Szirtes, J[an]. & H[erbert] G. Vaughan Jr. 1977. Characteristics of cranial and facial potentials associated with speech production. *Electroencephalography and Clinical Neurophysiology*, vol. 43, pp. 386–396.

Szirtes, J[an]. & H[erbert] G. Vaughan Jr. 1973. Topographic analysis of speech-related cerebral responses. *Electroencephalography and Clinical Neurophysiology*, vol. 34, p. 754.

Taylor, Wilson L. 1953. "Cloze Procedure": A New Tool For Measuring Readability. *Journalism Quarterly*, Fall 1953, pp. 415–433.

Tennant, Harry. 1979. Experience with the Evaluation of Natural Language Question Answerers. *Proceedings of the Sixth International Joint Conference on Artifical Intelligence (IJCAI) 1979*, 20–23 August 1979, Tokyo, Japan, pp. 874–876.

Tent, J. & J. E. Clark. 1980. An experimental investigation into the perception of slips of the tongue. *Journal of Phonetics*, vol. 8, no. 3, pp. 317–325.

Tessari, Alessia, Raffaella I. Rumiati & Patrick Haggard. 2002. Imitation without awareness. *Cognitive Science and Neuropsychology*, vol. 13, no. 18, pp. 2531–2535.

Thierry, Guillaume, Dominique Cardebat & Jean-François Démonet. Electrophysiological comparison of grammatical processing and semantic processing of single spoken nouns. *Cognitive Brain Research*, vol. 17, pp. 535–547.

Timsit, M. 1970. The Kornuhuber and Deecke phenomenon in schizophrenics and borderline cases. *Electroencephalography and Clinical Neurophysiology*, vol. 29, p. 535.

ToBI. **http://www.ling.ohio-state.edu/~tobi/**

Trager, George L. 1964. Paralanguage: A First Approximation. In: D. Hymes (ed.), *Language in culture and society*. New York: Harper and Row, pp. 274–288.

Trager, George L. 1958. Paralanguage: a first approximation. *Studies in Linguistics*, vol. 13,nos. 1–2, pp. 1–13.

Trask, R. L. 1996. *A Dictionary of Phonetics and Phonology*. London and New York: Routledge.

TravelLink™. **http://www.travellink.net/**

Travis, Lee Edward. 1978. The cerebral dominance theory of stuttering. *Journal of Speech and Hearing Disorders*, vol. 43, pp. 278–281.

Travis, Lee Edward. 1931. *Speech Pathology*. New York and London: D. Appleton and Company.

Tremblay, Stéphanie, Douglas M. Shiller & David J. Ostry. 2003. Somatosensory basis of speech production. *Nature*, vol. 423, pp. 866–869.

Trevena, Judy Arnel & Jeff Miller. 2002. Cortical Movement Preparation before and after a Conscious Decision to Move. *Consciousness and Cognition*, vol. 11, pp. 162–190.

Tsakiris, Manos & Patrick Haggard. 2003. Awareness of somatic events associated with a voluntary action. *Experimental Brain Research*, vol. 149, pt. 4, pp. 439–446.

Tseng, Shu-Chuan. 2000. Modelling Speech Repairs in German and Mandarin Chinese Spoken Dialogues. *Proceedings of COLING 2000*, 31 July–4 August 2000, Saarbrücken, Germany, vol. 2, pp. 864–870.

Tseng, Shu-Chuan. 1999. *Grammar, Prosody and Speech Disfluencies in Spoken Dialogues*. PhD thesis, Department of Linguistics and Literature, University of Bielefeld, Germany.

Turennout, Miranda van. See: Van Turennout, Miranda.

Tuthill, Curtis E. 1946. A quantitative study of extensional meaning with special reference to stuttering. *Speech Monographs*, vol. 13, pp. 81–98.

Tuthill, Curtis E. 1940. A quantitative study of extensional meaning with special reference to stuttering. *Journal of Speech Disorders*, vol. 5, pp. 189–191.

Tweney, Ryan D., Sharon Tkacz & Sally Zaruba. 1975. Slips of the tongue and lexical storage. *Language and Speech*, vol. 18, part 4, pp. 388–396.

Tyler, Lorraine K., Richard Russell, Jalal Fadili & Helen E. Moss. 2001. The neural representation of nouns and verbs: PET studies. *Brain*, vol. 124, pt. 8, pp. 1619–1634.

Underwood, Geoffrey. 1991. Attention is necessary for word integration. *Behavioral and Brain Sciences*, vol. 14, no. 4, p. 698.

Underwood, Geoffrey & Pekka Niemi. 1985. Mind before matter? *Behavioral and Brain Sciences*, vol. 8, pp. 554–555.

Vanderwolf, C. H. 1985. Nineteenth-century psychology and twentieth-century electrophysiology do not mix. *Behavioral and Brain Sciences*, vol. 8, p. 555.

Van Donzel, Monique E. & Florien J. Koopmans-van Beinum. 1998. Pausing strategies in discourse in Dutch. *Proceedings of the International Conference on Spoken Language Processing (ICLSP) '96*, 3–6 October 1996, Philadelphia, Pennsylvania, USA, vol 2, pp. 1029–1032.

Van Gulick, Robert. 1991. Consciousness may still have a processing role to play. *Behavioral and Brain Sciences*, vol. 14, no. 4, pp. 699–700.

Van Gulick, Robert. 1985. Conscious wants and self-awareness. *Behavioral and Brain Sciences*, vol. 8, pp. 555–556.

Van Petten, Cyma. 1995. Words and sentences: Event-related brain potential measures. *Psychophysiology*, vol. 32, pp. 511–525.

Van Petten, Cyma & Paul Bloom. 1999. Speech boundaries, syntax and the brain. *Nature Neuroscience*, vol. 2, no. 2, pp. 103–104.

Van Petten, Cyma & Marta Kutas. 1987. Ambiguous Words in Context: An Event-Related Potential Analysis of the Time Course of Meaning Activation. *Journal of Memory and Language*, vol. 26, pp. 188–208.

Van Riper, Charles. 1971/1982. *The Nature of Stuttering*. Englewood Cliffs, New Jersey: Prentice-Hall.

Van Turennout, Miranda, Peter Hagoort & Colin M. Brown. 1998. Brain Activity During Speaking: From Syntax to Phonology in 40 milliseconds. *Science*, vol. 280, pp. 572–574.

Van Wijk, Carel & Gerard Kempen. 1987. A Dual System for Producing Self-Repairs in Spontaneous Speech: Evidence from Experimentally Elicited Corrections. *Cognitive Psychology*, vol. 19, pp. 403–440.

Vaughan, Herbert G., Louis D. Costa & Walter Ritter. 1968. Topography of the human motor potential. *Electroencephalograpy and Clinical Neurophysiology*, vol. 25, pp. 1–10.

Velmans, Max. 1991a/1991b. Is human information processing conscious? *Behavioral and Brain Sciences*, vol. 14, pp. 651–726. Includes: Velmans, Max. 1991b. Consciousness from a first-person perspective. *Behavioral and Brain Sciences*, vol. 14, pp. 702–726.

Venditti, Jennifer J. 1997. Japanese ToBI Labelling Guidelines. *Ohio State University Working Papers in Linguistics*, vol. 50, pp. 127–162.

Veness, Thelma. 1962. An experiment on slips of the tongue and word association faults. *Language and Speech*, vol. 5, pt. 3, pp. 128–137.

Venkatagiri, Horabail S. 1981. Reaction Time for Voiced and Whispered /a/ in Stutterers and Nonstutterers. Journal of Fluency Disorders, vol. 6, pp. 265–271.

Verzeano, Marcel & Jacob E. Finesinger. 1949. An Automatic Analyzer for the Study of Speech in Interaction and in Free Association. *Science*, vol.110, pp. 45–46

von Hahn. See: Hahn.

Vos, Piet G., Jiři Mates & Noud W. van Kruysbergen. 1995. The Perceptual Centre of a Stimulus as the Cue for Synchronization to a Metronome: Evidence from Asynchronies. *The Quarterly Journal of Experimental Psychology*, vol. 48A, no. 4, pp. 1024–1040.

Voss, Bernd. 1979. Hesitation phenomena as sources of perceptual errors for non-native speakers. *Language and Speech*, vol. 22, pt. 2, pp. 129–144.

Wada, Juhn & Theodore Rasmussen. 1960. Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: experimental and clinical observation. *Journal of Neurosurgery*, vol. 17, pp. 266–282.

Walker, Stephen F. 2003. Misleading asymmetries of brain structure. *Behavioral and Brain Sciences*, vol. 26, no. 2, pp. 240–241.

Wall, Meryl J. 1980. A Comparison of Syntax in Young Stutterers and Nonstutterers. *Journal of Fluency Disorders*, vol. 5, pp. 345–352.

Waller, Steven J. 2002. Psychoacoustic influencies of the echoing environments of prehistoric art. *Journal of the Acoustical Society of America (JASA)*, vol. 112, no. 5, pt. 2, p. 2284.

Walter, W. Grey, R. Cooper, V. J. Aldridge, W. C. McCallum & A. L. Winter. 1964. Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. *Nature*, vol. 203, pp. 380–384.

Ward, Nigel. 2000. The Challenge of Non-lexical Speech Sounds. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2000*, 16–20 October 2000, Beijing, China, vol. 2, pp. 571–574.

Warren, Richard M. 1992. Global pattern perception and temporal order judgments. *Behavioral and Brain Sciences*, vol. 15, pp. 230–231.

Warren, Richard M. 1970. Perceptual Restoration of Missing Speech Sounds. *Science*, vol. 176, pp. 392–393.

Wasserman, Gerald S. 1985. Neural/mental chronometry and chronotechnology. *Behavioral and Brain Sciences*, vol. 8, pp. 556–557.

Waspwocz, Jan M., Ehud Yairi & Hugo H. Gregory. 1985. Acoustical and Perceptual Analysis of Stuttering and Nonstuttering Children's Sisfluency. *ASHA Reports* (American Speech–Language–Hearing Association), vol. 27, p. 186.

Watanabe, Michiko & Carlos Toshinori Ishi. 2001. The Usage of Fillers at Discourse Segment Boundaries in Japanese Lecture-style Monologues. *Proceedings of DiSS '01 Disfluency in Spontaneous Speech*, 29–31 August 2001, University of Edinburgh, Scotland, pp. 89–92.

Watson, Ben C. & Peter J. Alfonso. 1982. A comparison of LRT and VOT Values Between Stutterers and Nonstutterers. *Journal of Fluency Disorders*, vol. 7, pp. 219–241.

Wegner, Daniel M. 2002. *The Illusion of Conscious Will*. Cambridge, Massachusetts: Bradford Books.

Wegner, Daniel M. & Thalia Wheatley. 1999. Apparent Mental Causation. Source of the Experience of Will. *American Psychologist*, vol. 54, pp. 480–492.

Weeks, Gerald D. & Alphonse Chapanis. 1976. Cooperative versus conflictive problem solving in three telecommunication modes. *Perceptual and motor skills*, vol. 42, pp. 879–917.

Weeks, Gerald D., Michael J. Kelly & Alphonse Chapanis. 1974. Studies in interactive communication: V. Cooperative problem solving by skilled and unskilled typists in a teletypewriter mode. *Journal of Applied Psychology*, vol. 59, no. 6, pp. 665–674.

Weizenbaum, Joseph. 1967. Contextual Understanding by Computers. *Communications of the ACM*, vol. 10, no. 8, pp. 474–480.

Wells, Rulon. 1951/1973. Predicting Slips of the Tongue. In: Victoria A. Fromkin (ed.), *Speech Errors as Linguistic Evidence*. The Hague & Paris: Mouton, pp. 82–87. First published as: Wells, Rulon. 1951. Predicting Slips of the Tongue. *Yale Scientific Magazine* XXVI, no. 3, pp. 9–30.

Wengelin, Åsa. 2002. *Text Production in Adults with Reading and Writing Difficulties*. PhD thesis, Department of Linguistics, Göteborg University, Sweden.

Wengelin, Åsa. 2001. Disfluencies in Writing – are they Like in Speaking? *Proceedings of DiSS '01 Disfluency in Spontaneous Speech*, 29–31 August 2001, University of Edinburgh, Scotland, pp. 85–88.

Westby, Carol E. 1974. Language performance of stuttering and nonstuttering children. *Journal of Communications Disorders*, vol. 12, pp. 133–145.

Wexler, Karin B. 1982. Developmental disfluency in 2-, 4-, and 6-year-old boys in neutral and stress situations. *Journal of Speech and Hearing Research*, vol. 25, pp. 229–234.

Wexler, Karin B. & Edward D. Mysak. 1982. Disfluency Characteristics of 2-, 4-, and 6-Yr-Old Males. *Journal of Fluency Disorders*, vol. 7, pp. 37–46.

Whiteside, John, John Bennett & Karen Holtzblatt. 1988. Usability Engineering: Our Experience and Evolution. In: Martin Helander (ed.), *Handbook of Human–Computer Interaction*, Amsterdam: Elsevier (North-Holland), ch. 36, pp. 791–817.

Wightman, Colin. 2002. ToBI Or Not ToBI? *Proceedings of Speech Prosody 2002*, 11–13 April 2002, Aix-en-Provence, France, pp. 25–29.

Wightman, Colin, Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America (JASA)*, vol. 91, no. 3, pp. 1707–1717.

Wijk, Carel van. See: Van Wijk, Carel.

Wijnen, Frank. 1991. The Role of Language Formulation in Developmental Disfluency. *Proceedings of the XIIth International Congress of Phonetic Sciences*, 19–24 August 1991, Aix-en-Provence, France, pp. 146–149.

Williams, Dean E., Franklin H. Silverman & Joseph A. Kools. 1969a. Disfluency behavior of elementary-school stutterers and nonstutterers: The Consistency Effect. *Journal of Speech and Hearing Research*, vol. 12, pp. 301–307.

Williams, Dean E., Franklin H. Silverman & Joseph A. Kools. 1969b. Disfluency behavior of elementary-school stutterers and nonstutterers: loci of instances of disfluency. *Journal of Speech and Hearing Research*, vol. 12, pp. 308–318.

Williams, Dean E., Franklin H. Silverman & Joseph A. Kools. 1968. Disfluency behavior of elementary school stutterers and non-stutterers: The Adaptation Effect. *Journal of Speech and Hearing Research*, vol. 11, pp. 622–630.

Williams, Dean E., Michelle Wark & Fred D. Minifie. 1963. Ratings of Stuttering by Audio, Visual, and Audiovisual Cues. *Journal of Speech and Hearing Research*, vol. 6, pp. 91–100.

Wilson, David L. 1999. Mind–Brain Interaction and Violation of Physical Laws. *Journal of Consciousness Studies*, vol. 6, no. 8–9, pp. 185–200. Republished in: Benjamin Libet, Anthony Freeman & Keith Sutherland (eds.). 1999. *The Volitional Brain. Towards a Neuroscience of Free Will*, Thorverton, UK: Imprint Academic, pp. 185–200.

Wingate, Marcel E. 1994. Comments on Postma & Kolk's "The Covert Repair Hypothesis: Prearticulator Repair Processes in Normal and Stuttered Disfluencies" (1993). *Journal of Speech and Hearing Research*, vol. 37, no. 3, p. 581.

Wingate, Marcel E. 1987. Fluency and disfluency: illusion and identification. *Journal of Fluency Disorders*, vol. 12, pp. 79–101.

Wingate, Marcel E. 1984a. Pause loci in stuttered and normal speech. *Journal of Fluency Disorders*, vol. 9, pp. 227–235.

Wingate, Marcel E. 1984b. Fluency, disfluency, dysfluency, and stuttering. *Journal of Fluency Disorders*, vol. 17, pp. 163–168.

Wingate, Marcel E. 1984c. Definition *Is* the Problem. *Journal of Speech and Hearing Research*, vol. 49, pp. 429–431.

Wingate, M[arcel]. E. 1970. Effects on stuttering of changes in audition. Journal of Speech and Hearing Research, vol. 13, pp. 861–873.

Wingate, M[arcel]. E. 1969. Sound and pattern in "artificial" fluency. *Journal of Speech and Hearing Research*, vol. 12, pp. 677–686.

Wise, Richard J. S., Sophie K. Scott, S. Catrin Blank, Cath J. Mummery, Kevin Murphy & Elizabeth A. Warburton. 2001. Separate neural subsystems within 'Wernicke's area'. *Brain*, vol. 124, pt. 1, pp. 83–95.

Wohlert, Amy B. 1993. Event-Related Brain Potentials Preceding Speech and Nonspeech Oral Movements of Varying Complexity. *Journal of Speech and Hearing Research*, vol. 36, pp. 897–905.

Wohlert, Amy B. & Charles R. Larson. 1991. Cerebral Averaged Potentials Preceding Oral Movement. *Journal of Speech and Hearing Research*, vol. 34, pp. 1387–1396.

Wolfe Siegman, Aron. See: Siegman, Aron Wolfe.

Woll, Bencie & Jechil S. Sieratzki. 2003. Why homolaterality of language and hand dominance may not be the expression of a specific evolutionary link. *Behavioral and Brain Sciences*, vol. 26, no. 2, p. 241.

Wolpert, Lewis. 2003. Causal beliefs lead to toolmaking, which require handedness for motor control. *Behavioral and Brain Sciences*, vol. 26, no. 2, p. 242.

Wood, Charles C. 1985. Pardon, your dualism is showing. *Behavioral and Brain Sciences*, vol. 8, pp. 557–558.

Woodworth, R. S. 1900. The accuracy of voluntary movement. *Psychological Review*, vol. III, no. 4, pp. 1–114.

Yairi, Ehud. 1981. Disfluencies of normally speaking two-year-old children. *Journal of Speech and Hearing Research*, vol. 24, pp. 490–495.

Yairi, Ehud & Noel F. Clifton, Jr. 1972. Disfluent speech behavior of preschool children, high school seniors, and geriatric persons. *Journal of Speech and Hearing Research*, vol. 15, no. 4, pp. 714–719.

Yarmey, A. Daniel. 1973. I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Memory & Cognition*, vol. 1, no. 3, pp. 287–290.

Yarrow, Kielan, Patrick Haggard, Ron Heal, Peter Brown & Ron C. Rothwell. 2001. Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*, vol. 414, pt. 6861, pp. 302–305.

Yeni-Komshian, Grace, Richard Allen Chase & Richard L. Mobley. 1968. The development of auditory feedback monitoring: II. Delayed Auditory Feedback studies on the speech of children between two and three years of age. *Journal of Speech and Hearing Research*, vol. 11, pp. 307–315.

Young, Andy. 1992. Closing the Cartesian Theatre. *Behavioral and Brain Sciences*, vol. 15, p. 233.

Young, J. Z. 1962. The Thirty-Sixth Maudsley Lecture: Memory Mechanisms of the Brain. *The Journal of Mental Science*, vol. 108, no. 453, pp. 119–133.

Young, Martin A. 1985. Identification of Stuttering and Stutterers. In: Richard F. Curlee & William H. Perkins (eds.), *Nature and Treatment of Stuttering: New Directions*. San Diego, California: College-Hill Press, ch. 2, pp. 13–30.

Zebrowski, Patricia M. 1994. Duration of Sound Prolongations and Sound/Syllable Repetition in Children Who Stutter: Preliminary Observations. *Journal of Speech and Hearing Research*, vol. 37, pp. 254–263.

Zebrowski, Patricia M., Edward G. Conture & Edward A. Cudahy. 1985. Acoustic analysis of young stutterers' fluency: preliminary observations. *Journal of Fluency Disorders*, vol. 10, pp. 173–192.

Zeman, Adam. 2001. Consciousness. *Brain*, vol. 124, pp. 1263–1289.

Zimbardo, Philip G, George F. Mahl & James W. Barnard. 1987. Speech Disturbance in Anxious Children. In: George F. Mahl (ed.), *Explorations in Nonverbal and Vocal Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum, ch. 13, pp. 214–222.

Zimbardo, Philip G., George F. Mahl & James W. Barnard. 1963. The Measurement of Speech Disturbances in Anxious Children. *Journal of Speech and Hearing Disorders*, vol. 28, pp. 362–370.

Zimmerman, Gerald N. & J. R. Knott. 1974. Slow potentials of the brain related to speech processing in normal speakers and stutterers. *Electroencephalography and Clinical Neurophysiology*, vol. 37, pp. 599–607.

Zipf, George Kingsley. 1945. The meaning–frequency relationship of words. *The Journal of General Psychology*, vol. 33, pp. 251–256.

Zivin, Gail. 1986. Image or neural coding of inner speech and agency? *Behavioral and Brain Sciences*, vol. 9, no, pp. 534–535.

Zocchi, Luciano, Marc Estenne, Sharon Johnston, Leonardo Del Ferro, Michael E. Ward & Peter T. Macklem. 1990. Respiratory Muscle Incoordination in Stuttering Speech. *American Review of Respiratory Disease*, vol. 141, pp. 1510–1515.

APPENDICES

Appendix 1: WOZ-1

Appendix 1a. Summary statistics for WOZ-1, human–"machine–human" corpus. Number of dialogs, utterances and words given for all subjects, as well as the seven major categories of disfluencies, broken down and summarized for subjects and dialogs. Sums for individual subjects are shown in cells with 5% shading, and sum totals are shown in cells with 10% shading. Nota bene! Omitted subjects do not appear in sum total.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
01	F	1	11	57	3	0	4	0	0	0	0	7
		2	7	45	4	1	0	0	0	0	0	5
		3	12	73	4	0	1	0	0	0	1	6
		Σ	**30**	**175**	**11**	**1**	**5**	**0**	**0**	**0**	**1**	**18**
02	F	1	12	69	3	4	3	0	0	1	3	14
		2	13	47	1	1	0	0	0	0	0	2
		3	14	49	0	2	0	0	0	0	0	2
		4	8	21	0	0	0	0	0	0	0	0
		5	12	45	0	1	0	0	0	0	0	1
		6	11	41	0	1	0	0	0	0	0	1
		7	13	42	1	3	0	0	0	0	0	4
		Σ	**83**	**314**	**5**	**12**	**3**	**0**	**0**	**1**	**3**	**24**
03	M	1	21	140	14	15	7	3	0	3	6	48
		2	6	17	2	1	0	0	0	1	1	5
		3	16	77	8	6	7	1	1	1	0	24
		4	13	90	5	10	1	0	0	0	1	17
		5	15	95	3	7	1	0	1	0	0	12
		6	14	84	10	5	0	0	0	0	1	16
		7	13	66	3	3	1	0	0	0	0	7
		8	10	50	5	4	0	0	0	0	0	9
		9	8	36	2	3	0	0	0	0	0	5
		10	7	45	4	4	0	0	1	1	0	10
		11	14	89	4	1	0	0	0	1	1	7
		Σ	**137**	**789**	**60**	**59**	**17**	**4**	**3**	**7**	**10**	**160**
04	M	1	10	83	12	6	3	0	0	0	0	21
		2	8	76	12	4	1	0	0	0	3	20
		3	12	76	4	4	1	0	0	1	0	10
		4	7	53	6	3	0	0	0	1	1	11
		5	9	49	3	2	0	0	0	0	0	5
		6	8	42	2	3	0	0	0	0	0	5
		7	7	38	2	0	0	0	0	0	0	2
		8	7	32	2	0	1	0	0	0	0	3
		9	8	40	1	1	0	0	0	0	0	2
		10	8	60	2	1	0	0	0	0	0	3
		Σ	**84**	**549**	**46**	**24**	**6**	**0**	**0**	**2**	**4**	**82**
05	M	1	8	53	3	5	6	0	0	0	1	15
		2	9	46	1	5	1	0	0	0	0	7
		3	10	61	1	4	0	0	0	0	0	5
		4	6	29	0	1	0	0	0	0	0	1
		Σ	**33**	**189**	**5**	**15**	**7**	**0**	**0**	**0**	**1**	**28**
06	F	1	11	55	2	0	0	0	0	0	0	2
		2	17	78	1	0	0	0	0	0	0	1
		3	11	84	5	0	0	0	0	0	0	5
		4	11	53	4	0	0	0	3	0	0	7
		5	14	63	1	0	0	0	0	0	0	1
		6	14	57	0	0	0	0	0	0	0	0
		7	9	38	0	0	0	0	0	0	0	0
		8	12	44	1	0	0	0	0	0	0	1
		9	9	38	1	0	0	0	0	0	0	1
		10	14	66	2	0	0	0	0	1	1	4
		Σ	**122**	**576**	**17**	**0**	**0**	**0**	**3**	**1**	**1**	**22**

Appendix 1b. Summary statistics for WOZ-1, human–"machine–human" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
07	M	1	13	63	4	1	4	0	0	0	0	9
		2	8	40	2	2	0	0	0	0	0	4
		3	10	72	5	9	0	0	1	0	0	15
		4	9	48	2	4	0	0	0	1	1	8
		5	10	54	0	4	0	0	0	0	0	4
		6	5	34	1	2	0	0	0	0	0	3
		7	6	34	1	1	0	0	0	0	0	2
		8	7	35	3	5	0	0	0	0	0	8
		9	9	42	1	3	0	0	0	0	0	4
		10	9	46	0	3	0	0	0	0	0	3
		Σ	**86**	**468**	**19**	**34**	**4**	**0**	**1**	**1**	**1**	**60**
08	F	1	17	131	18	8	3	1	0	0	2	32
		2	16	150	19	7	0	0	0	0	2	28
		3	15	94	9	7	0	0	0	0	0	16
		4	21	126	8	11	0	0	0	2	0	21
		5	15	91	4	6	0	0	1	2	0	13
		6	13	86	7	5	0	0	0	0	0	12
		7	11	56	5	3	0	0	0	0	0	8
		8	11	55	3	4	0	0	0	0	0	7
		9	9	53	3	5	0	0	0	0	0	8
		10	7	56	2	5	0	0	0	0	0	7
		11	7	41	5	6	0	0	0	0	0	11
		12	7	37	2	0	0	0	0	0	0	2
		Σ	**149**	**976**	**85**	**67**	**3**	**1**	**1**	**4**	**4**	**165**
09	F	1	14	112	6	0	0	0	0	0	0	6
		2	14	99	5	1	0	0	0	0	0	6
		3	17	118	4	2	0	0	0	1	1	8
		4	12	86	4	1	0	0	0	0	0	5
		5	17	118	3	0	0	0	0	0	1	4
		6	14	101	6	0	1	0	0	1	0	8
		7	19	144	6	2	0	0	0	0	1	9
		8	12	82	2	0	0	1	1	0	1	5
		9	11	89	2	1	0	0	0	0	0	3
		10	17	127	4	0	1	0	0	1	1	7
		Σ	**147**	**1076**	**42**	**7**	**2**	**1**	**1**	**3**	**5**	**61**
10	M	1	5	28	2	0	0	0	0	0	0	2
		2	8	78	7	2	0	0	0	1	2	12
		3	8	71	8	3	1	0	0	1	1	14
		4	7	51	5	2	0	0	0	0	0	7
		5	7	34	5	1	0	0	0	0	0	6
		6	7	32	1	3	0	0	0	0	0	4
		7	9	61	6	1	0	0	0	0	0	7
		8	6	37	3	0	1	0	1	0	1	6
		9	5	40	6	0	0	0	0	1	1	8
		10	7	46	2	0	0	0	0	1	1	4
		Σ	**69**	**478**	**45**	**12**	**2**	**0**	**1**	**4**	**6**	**70**
11	M	Non-native speaker of Swedish. Omitted from analysis.										
		Σ	–	–	–	–	–	–	–	–	–	–
12	F	1	12	95	7	4	2	0	0	0	1	14
		2	15	119	6	2	0	1	0	0	1	10
		3	15	115	3	5	0	0	0	0	0	8
		4	14	91	1	5	0	0	0	0	0	6
		5	17	123	8	3	0	0	0	0	1	12
		6	14	90	1	1	0	0	0	0	0	2
		7	11	62	0	1	0	0	0	0	0	1
		8	11	72	6	2	0	0	0	0	0	8
		9	8	46	1	1	0	0	0	0	0	2
		10	10	54	0	0	0	0	0	0	0	0
		Σ	**127**	**867**	**33**	**24**	**2**	**1**	**0**	**0**	**3**	**63**

Appendix 1c. Summary statistics for WOZ-1, human–"machine–human" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
13	M	1	9	43	0	0	0	0	0	0	0	0
		2	8	75	0	2	3	0	0	0	1	6
		3	7	50	1	1	0	0	0	0	0	2
		4	10	72	0	0	0	0	0	1	2	3
		5	6	43	0	0	0	0	0	0	0	0
		6	6	49	0	0	0	0	0	0	0	0
		7	7	50	0	0	0	0	0	0	0	0
		8	7	43	0	1	0	0	0	0	0	1
		9	6	44	0	0	0	0	0	0	0	0
		10	7	43	1	0	0	0	0	0	0	1
		Σ	**73**	**512**	**2**	**4**	**3**	**0**	**0**	**1**	**3**	**13**
14	M	1	9	40	1	4	0	0	0	0	0	5
		2	10	67	2	8	0	0	0	1	1	12
		3	8	52	1	2	0	0	0	0	0	3
		4	9	48	0	3	0	0	0	0	0	3
		5	8	44	0	1	0	0	1	1	1	4
		6	7	60	2	2	0	0	0	0	0	4
		7	7	43	0	0	0	0	0	0	0	0
		8	8	39	0	0	0	0	0	0	0	0
		9	11	55	0	4	0	0	0	0	0	4
		10	7	38	0	0	1	0	0	0	0	1
		Σ	**84**	**486**	**6**	**24**	**1**	**0**	**1**	**2**	**2**	**36**
15	M	1	9	36	1	0	1	0	0	0	0	2
		2	9	49	1	1	0	0	0	0	0	2
		3	10	60	3	2	0	0	0	1	1	7
		4	8	35	1	0	0	0	0	0	0	1
		5	7	44	3	2	0	0	0	0	0	5
		6	9	49	2	2	0	0	0	0	0	4
		7	7	41	0	2	0	0	0	0	0	2
		8	8	55	1	3	0	0	0	0	0	4
		9	8	46	1	0	1	0	0	0	0	2
		10	8	42	1	0	0	0	0	0	0	1
		Σ	**83**	**457**	**14**	**12**	**2**	**0**	**0**	**1**	**1**	**30**
16	F	1	12	71	3	1	2	0	0	0	0	6
		2	9	25	0	0	0	0	0	0	0	0
		3	9	53	3	0	0	1	1	2	3	10
		4	9	63	6	0	0	0	0	0	0	6
		5	7	26	1	1	0	0	0	0	0	2
		6	11	75	1	1	0	0	0	0	0	2
		7	8	38	1	0	0	0	0	1	1	3
		8	8	47	3	0	0	0	0	0	0	3
		9	7	30	0	0	0	0	0	0	0	0
		Σ	**80**	**428**	**18**	**3**	**2**	**1**	**1**	**3**	**4**	**32**
17	M	1	29	303	14	16	2	0	0	3	4	39
		2	18	148	7	2	0	0	0	2	0	11
		3	12	89	7	3	1	0	0	0	0	11
		4	12	78	5	4	0	0	0	2	2	13
		5	11	88	9	4	0	0	2	7	1	23
		6	10	93	6	5	0	0	0	0	0	11
		7	9	34	2	1	0	0	0	0	0	3
		8	8	38	0	1	0	0	0	0	0	1
		9	7	41	2	2	0	0	0	0	0	4
		10	8	48	4	3	0	0	0	1	1	9
		Σ	**124**	**960**	**56**	**41**	**3**	**0**	**2**	**15**	**8**	**125**
18	M	1	17	91	7	1	2	0	0	0	1	11
		2	16	86	9	1	4	0	0	2	3	19
		3	8	29	1	0	0	0	0	0	0	1
		4	9	50	0	0	0	0	0	0	0	0
		5	8	61	2	0	0	0	0	2	1	5
		6	10	46	2	0	0	0	0	0	0	2
		7	7	29	1	0	0	0	0	0	0	1
		8	10	44	1	1	0	0	0	0	0	2
		9	7	41	0	1	0	0	0	0	0	1
		Σ	**92**	**477**	**23**	**4**	**6**	**0**	**0**	**4**	**5**	**42**

Appendix 1d. Summary statistics for WOZ-1, human–"machine–human" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
19	M	1	12	53	4	2	3	0	0	0	1	10
		2	13	49	3	0	1	0	0	1	1	6
		3	18	92	11	3	1	0	0	2	2	19
		4	8	45	4	1	0	0	1	0	1	7
		5	11	47	3	0	2	0	0	0	0	5
		6	9	49	8	1	1	0	0	0	1	11
		7	7	25	5	0	1	0	0	0	0	6
		8	8	37	2	0	0	0	0	0	0	2
		9	6	31	3	0	1	0	0	1	1	6
		10	6	44	1	0	2	0	0	0	0	3
		Σ	**98**	**472**	**44**	**7**	**12**	**0**	**1**	**4**	**7**	**75**
20	F	1	9	76	10	3	0	0	0	1	1	15
		2	12	105	5	2	0	2	0	1	2	12
		3	14	137	10	3	1	0	0	3	3	20
		4	9	91	3	0	1	0	0	0	1	5
		5	10	93	5	2	0	0	0	0	0	7
		6	13	106	6	0	0	0	0	0	0	6
		7	7	53	0	0	0	0	0	0	0	0
		8	9	93	3	1	2	0	1	0	1	8
		9	10	90	3	1	0	0	1	0	0	5
		10	10	125	5	2	0	0	0	1	1	9
		Σ	**103**	**969**	**50**	**14**	**4**	**2**	**2**	**6**	**9**	**87**
21	F	Non-native speaker of Swedish. Omitted from analysis.										
		Σ	–	–	–	–	–	–	–	–	–	–
22	M	1	11	62	5	0	0	0	0	0	0	5
		2	14	85	9	2	0	0	0	0	0	11
		3	11	76	10	2	0	0	0	0	0	12
		4	10	76	7	0	0	1	0	0	0	8
		5	10	68	6	1	0	0	0	0	0	7
		6	11	85	9	3	0	0	0	0	0	12
		7	11	64	4	1	0	0	0	0	0	5
		8	12	76	4	2	1	0	0	0	0	7
		9	11	68	3	2	0	0	1	0	1	7
		10	19	120	6	1	0	0	0	2	2	11
		Σ	**120**	**780**	**63**	**14**	**1**	**1**	**1**	**2**	**3**	**85**
23	M	1	6	45	3	1	0	0	0	0	0	4
		2	9	91	8	6	0	0	0	0	0	14
		3	8	55	9	1	0	0	0	0	0	10
		4	8	58	3	2	0	0	0	0	0	5
		5	9	60	5	3	1	0	0	0	0	9
		6	8	59	4	2	0	0	0	0	0	6
		7	6	50	3	2	1	0	0	0	0	6
		8	6	50	3	1	2	0	0	0	0	6
		9	6	48	3	2	0	0	0	0	0	5
		10	7	50	2	1	1	0	0	0	0	4
		Σ	**73**	**566**	**43**	**21**	**5**	**0**	**0**	**0**	**0**	**69**
24	M	1	7	46	6	0	1	0	0	0	1	8
		2	14	114	11	6	2	0	0	1	1	21
		3	10	77	8	3	0	0	0	0	0	11
		4	8	66	1	2	1	0	0	0	2	6
		5	10	84	6	3	1	0	0	0	0	10
		6	8	48	2	2	0	0	0	0	0	4
		7	9	67	1	2	1	0	0	1	1	6
		8	10	68	0	2	0	0	0	0	0	2
		9	9	59	2	2	1	0	0	0	0	5
		10	9	51	0	0	0	0	0	0	0	0
		Σ	**94**	**680**	**37**	**22**	**7**	**0**	**0**	**2**	**5**	**73**

Appendix 1e. Summary statistics for WOZ-1, human–"machine–human" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
25	F	1	13	82	5	0	0	0	0	0	0	5
		2	14	95	8	3	0	0	0	0	0	11
		3	7	86	8	1	1	0	0	0	0	10
		4	7	61	3	1	0	0	0	0	0	4
		5	9	69	4	1	0	0	0	0	0	5
		6	7	71	6	1	0	0	0	0	0	7
		7	9	53	2	0	0	0	0	0	0	2
		8	8	59	3	0	0	0	0	0	0	3
		9	6	43	3	1	0	0	0	0	0	4
		10	9	53	1	1	0	0	0	0	1	3
		Σ	**89**	**672**	**43**	**9**	**1**	**0**	**0**	**0**	**1**	**54**
26	M	1	9	34	3	0	0	0	0	0	0	3
		2	11	47	1	0	0	0	0	0	0	1
		3	8	38	0	0	0	0	0	0	0	0
		4	3	25	1	0	0	0	0	0	0	1
		5	4	40	3	0	0	0	0	0	1	4
		6	5	56	8	0	0	0	0	0	0	8
		7	6	31	1	0	0	0	0	0	0	1
		8	6	31	2	0	0	0	0	0	0	2
		9	4	29	3	0	0	0	0	0	0	3
		10	7	33	0	0	0	0	0	0	0	0
		Σ	**63**	**364**	**22**	**0**	**0**	**0**	**0**	**0**	**1**	**23**
27	M	1	11	58	5	0	2	0	0	0	0	7
		2	10	59	5	3	0	0	0	0	0	8
		3	10	70	3	2	0	0	0	0	1	6
		4	10	87	4	5	1	0	0	0	1	11
		5	10	91	4	2	0	0	0	0	0	6
		6	9	76	4	3	0	0	0	0	0	7
		7	13	92	0	1	1	0	0	0	0	2
		8	11	106	5	0	2	0	0	0	0	7
		9	13	128	2	2	1	0	0	0	0	5
		10	14	84	1	3	0	0	0	0	0	4
		Σ	**111**	**851**	**33**	**21**	**7**	**0**	**0**	**0**	**2**	**63**
28	M	1	7	17	1	2	1	0	0	0	0	4
		2	8	43	3	1	0	0	0	2	2	8
		3	9	50	0	1	0	0	0	1	0	2
		4	7	29	0	0	0	0	0	0	0	0
		5	9	30	0	1	0	0	0	0	0	1
		6	7	25	0	0	0	0	0	0	0	0
		7	7	34	0	0	1	0	0	1	1	3
		8	7	19	0	0	0	0	0	0	0	0
		9	8	24	0	1	0	0	0	0	0	1
		10	8	26	0	0	0	0	0	0	0	0
		Σ	**77**	**297**	**4**	**6**	**2**	**0**	**0**	**4**	**3**	**19**
29	M	1	11	32	3	1	1	0	0	0	0	5
		2	12	73	11	3	0	1	0	0	1	16
		3	9	41	0	1	0	0	0	0	0	1
		4	14	62	3	1	0	0	1	0	0	5
		5	10	45	0	1	0	0	0	0	0	1
		6	12	42	0	0	0	1	0	1	2	4
		7	8	43	0	1	0	1	0	1	1	4
		8	7	34	0	0	0	0	0	0	0	0
		9	6	33	0	0	0	0	0	0	0	0
		10	16	45	0	0	0	1	0	0	1	2
		Σ	**105**	**450**	**17**	**8**	**1**	**4**	**1**	**2**	**5**	**38**
30	M	No recording obtained.										
		Σ	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**

Appendix 1f. Summary statistics for WOZ-1, human–"machine–human" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
31	F	1	10	55	1	1	0	0	1	0	1	4
		2	9	44	1	1	0	0	0	0	0	2
		3	14	78	2	2	0	0	0	0	0	4
		4	10	58	1	2	0	0	0	3	1	7
		5	11	55	3	3	1	0	0	0	0	7
		6	11	64	0	3	0	0	0	0	0	3
		7	5	26	0	1	0	0	0	0	0	1
		8	8	47	0	0	0	0	0	0	0	0
		9	9	48	1	0	0	0	0	1	1	3
		10	9	48	1	2	1	0	0	0	0	4
		11	11	48	0	1	0	0	0	0	0	1
		Σ	**107**	**571**	**10**	**16**	**2**	**0**	**1**	**4**	**3**	**36**
32	F	1	9	62	1	5	0	1	0	0	1	8
		2	7	42	1	0	0	0	1	0	1	3
		3	7	60	4	0	0	0	0	0	1	5
		4	7	34	1	0	0	0	0	0	0	1
		5	8	43	1	0	0	0	0	0	0	1
		6	6	33	0	1	0	0	0	0	0	1
		7	5	26	0	1	0	0	0	1	1	3
		8	6	27	2	0	0	0	0	0	0	2
		9	4	22	1	0	0	0	0	0	0	1
		10	6	24	0	0	0	0	0	0	0	0
		Σ	**65**	**373**	**11**	**7**	**0**	**1**	**1**	**1**	**4**	**25**
33	M	No recording obtained.										
		Σ	–	–	–	–	–	–	–	–	–	–
34	M	1	12	38	0	2	0	0	0	1	1	4
		2	9	56	2	0	0	0	0	2	2	6
		3	10	64	0	0	0	0	0	0	0	0
		4	7	71	3	0	0	0	0	0	1	4
		5	10	60	0	0	0	0	0	0	0	0
		6	8	81	3	0	0	0	0	0	0	3
		7	9	51	3	0	0	0	0	0	0	3
		8	8	45	0	0	0	0	0	0	0	0
		9	9	52	1	0	0	0	0	0	0	1
		10	11	75	0	0	0	0	0	0	0	0
		Σ	**93**	**593**	**12**	**2**	**0**	**0**	**0**	**3**	**4**	**21**
35	F	1	9	78	5	3	6	0	2	5	5	26
		2	10	93	5	5	1	0	0	3	4	18
		3	9	58	0	3	1	0	0	0	0	4
		4	6	51	3	1	1	0	0	1	1	7
		5	7	42	2	0	0	0	0	0	0	2
		6	8	52	1	2	1	0	0	0	1	5
		7	6	50	2	3	0	0	0	0	0	5
		Σ	**55**	**424**	**18**	**17**	**10**	**0**	**2**	**9**	**11**	**67**
36	F	1	20	157	13	9	0	0	1	5	5	33
		2	21	151	6	10	2	0	0	2	2	22
		3	15	153	10	2	6	0	1	4	2	25
		4	12	120	7	6	0	0	0	3	1	17
		5	10	54	2	0	1	1	1	0	1	6
		6	11	94	7	6	1	0	0	3	1	18
		7	8	52	3	4	1	0	2	2	2	14
		8	8	66	5	2	1	0	0	3	3	14
		9	8	80	10	1	0	0	0	0	1	12
		10	16	181	11	5	0	0	0	2	0	18
		Σ	**129**	**1108**	**74**	**45**	**12**	**1**	**5**	**24**	**18**	**179**
37	M	1	12	66	1	3	1	0	0	1	1	7
		2	11	69	2	4	0	0	0	0	0	6
		3	8	62	5	2	0	0	0	0	0	7
		4	9	85	4	3	0	0	0	0	1	8
		5	8	60	5	2	0	0	0	1	1	9
		6	10	83	10	4	0	0	0	0	0	14
		7	6	48	5	0	0	0	0	0	0	5
		Σ	**64**	**473**	**32**	**18**	**1**	**0**	**0**	**2**	**3**	**56**

Appendix 1g. Summary statistics for WOZ-1, human–"machine–human" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
38	F	1	16	49	2	2	0	1	0	0	2	7
		2	11	84	2	3	0	0	0	0	0	5
		3	6	70	8	5	0	0	0	0	0	13
		4	8	39	2	2	0	0	0	0	0	4
		5	8	49	1	1	0	0	0	0	0	2
		6	11	118	11	5	0	0	0	0	0	16
		7	7	34	2	1	0	0	0	0	0	3
		8	7	49	5	2	0	0	0	0	0	7
		9	6	36	1	0	0	0	0	0	0	1
		10	8	44	1	1	0	0	0	1	0	3
		Σ	**88**	**572**	**35**	**22**	**0**	**1**	**0**	**1**	**2**	**61**
39	M	No recording obtained.										
		Σ	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**
40	F	1	5	59	6	0	1	0	0	0	0	7
		2	7	103	11	2	2	0	0	0	0	15
		3	8	96	10	5	1	0	0	2	3	21
		4	9	98	9	5	1	0	0	0	0	15
		5	7	59	0	0	0	0	0	0	0	0
		6	7	74	5	1	2	0	0	1	1	10
		7	6	45	2	1	0	0	0	0	0	3
		8	8	73	5	2	0	0	0	0	0	7
		9	5	49	4	1	0	0	0	1	0	6
		10	7	70	2	2	0	0	1	1	1	7
		Σ	**69**	**726**	**54**	**19**	**7**	**0**	**1**	**5**	**5**	**91**
41	F	1	5	86	10	6	0	0	0	0	1	17
		2	12	140	8	7	0	0	1	0	0	16
		3	9	114	12	8	0	0	0	3	4	27
		4	7	68	5	0	0	0	0	0	0	5
		5	10	117	9	4	0	0	0	2	2	17
		6	8	75	6	2	0	0	0	2	2	12
		7	9	61	3	2	0	0	0	0	0	5
		8	10	73	1	3	0	0	0	1	2	7
		9	8	67	4	2	0	0	0	1	0	7
		10	9	70	3	2	0	0	0	0	0	5
		Σ	**87**	**871**	**61**	**36**	**0**	**0**	**1**	**9**	**11**	**118**
42	M	1	8	26	2	2	2	0	0	0	1	7
		2	8	58	5	2	0	0	0	0	0	7
		3	9	75	6	0	0	0	0	1	1	8
		4	8	41	6	0	0	0	0	1	2	9
		Σ	**33**	**200**	**19**	**4**	**2**	**0**	**0**	**2**	**4**	**31**
43	M	1	18	161	19	16	1	0	0	2	4	42
		2	7	88	14	7	0	0	0	0	1	22
		3	10	100	19	6	0	0	0	0	0	25
		4	9	46	9	8	0	0	0	0	0	17
		5	7	58	9	6	0	0	0	0	0	15
		6	7	53	10	3	1	0	0	0	0	14
		7	10	64	8	3	0	0	0	1	1	13
		8	11	56	5	4	0	0	0	0	0	9
		9	6	42	7	2	0	0	0	0	0	9
		10	6	42	3	1	0	0	0	0	0	4
		Σ	**91**	**710**	**103**	**56**	**2**	**0**	**0**	**3**	**6**	**170**
44	M	Non-native speaker of Swedish. Omitted from analysis.										
		Σ	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**	**–**
45	M	1	4	38	3	5	1	0	0	2	2	13
		2	6	61	10	4	0	0	0	0	0	14
		3	4	36	3	1	0	0	0	1	9	5
		4	5	42	6	2	0	0	0	0	0	8
		5	6	38	3	3	0	0	0	0	0	6
		6	4	34	4	0	0	0	0	0	2	6
		7	4	37	1	0	0	0	0	0	0	1
		8	5	58	6	1	0	0	1	1	1	10
		9	4	35	4	2	0	0	0	0	0	6
		10	5	52	7	1	0	0	0	0	0	8
		Σ	**47**	**431**	**47**	**19**	**1**	**0**	**1**	**4**	**5**	**77**

Appendix 1h. Summary statistics for WOZ-1, human–"machine–human" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
46	M	1	5	29	0	1	0	0	0	0	0	1
		2	6	74	5	2	0	0	0	1	1	9
		3	6	70	0	3	0	0	0	0	0	3
		4	5	82	4	1	0	0	0	0	0	5
		5	9	80	2	0	0	0	0	1	2	5
		6	7	57	3	1	0	0	0	0	0	4
		7	5	37	0	0	0	0	0	0	0	0
		8	7	58	2	1	1	0	0	0	0	4
		9	2	15	0	0	0	0	1	0	0	1
		Σ	**52**	**502**	**16**	**9**	**1**	**0**	**1**	**2**	**3**	**32**
47	F	1	13	68	3	1	0	0	0	2	2	8
		2	5	47	4	0	0	0	0	2	2	8
		3	7	48	4	1	0	0	1	1	0	7
		4	8	63	7	1	0	0	0	0	0	8
		5	10	116	3	7	0	0	0	5	3	18
		6	9	50	2	0	0	0	0	1	1	4
		7	9	82	3	5	0	0	0	4	3	15
		8	7	58	2	2	0	0	0	1	2	7
		9	6	43	0	0	0	0	0	0	1	1
		10	7	68	5	0	0	0	0	0	0	5
		Σ	**81**	**643**	**33**	**17**	**0**	**0**	**1**	**16**	**14**	**81**
48	F	1	10	61	1	4	1	0	0	0	1	7
		2	9	60	2	3	0	0	0	0	0	5
		3	10	68	3	2	0	0	0	0	0	5
		4	8	48	2	3	0	0	0	0	0	5
		5	8	49	0	5	0	0	0	0	0	5
		6	9	73	3	6	0	0	0	0	0	9
		7	6	37	2	1	0	0	0	0	0	3
		8	8	57	2	4	0	0	0	0	0	6
		9	8	60	1	3	0	0	0	0	0	4
		10	9	58	1	0	0	0	0	0	0	1
		Σ	**85**	**571**	**17**	**31**	**1**	**0**	**0**	**0**	**1**	**50**
49	F	1	9	103	16	0	2	0	0	2	2	22
		2	11	146	23	2	1	0	0	0	0	26
		3	8	128	24	3	0	0	0	1	1	29
		4	9	90	10	3	0	0	0	1	1	15
		5	9	90	12	1	0	0	0	1	1	15
		6	8	99	9	4	0	0	1	0	1	15
		7	7	69	2	3	0	0	1	0	1	7
		8	9	107	15	3	1	0	0	0	0	19
		9	6	64	4	0	0	0	0	1	1	6
		10	9	100	7	3	0	0	0	1	0	11
		Σ	**85**	**996**	**122**	**22**	**4**	**0**	**2**	**7**	**8**	**165**
50	F	1	6	35	2	0	1	0	0	0	0	3
		2	5	41	3	0	0	0	0	0	0	3
		3	4	55	5	0	0	0	0	0	0	5
		4	4	39	2	0	0	0	0	0	0	2
		5	5	43	0	0	0	0	0	0	0	0
		6	5	67	6	0	0	0	0	0	0	6
		7	6	59	5	0	0	0	0	0	0	5
		8	5	55	3	0	0	0	0	0	0	3
		9	4	57	2	0	0	0	0	0	0	2
		10	4	36	2	0	0	0	0	0	0	2
		Σ	**48**	**487**	**30**	**0**	**1**	**0**	**0**	**0**	**0**	**31**
51	F	1	10	125	11	0	3	0	0	1	1	16
		2	13	140	10	1	0	2	0	2	4	19
		3	7	94	10	0	1	0	0	0	1	12
		4	9	90	5	0	0	0	1	0	1	7
		5	10	67	2	1	0	0	1	0	1	5
		6	8	75	3	0	0	0	1	0	1	5
		7	6	58	5	0	0	0	0	0	0	5
		8	8	45	4	2	0	0	0	0	0	6
		9	7	37	0	0	0	0	0	0	0	0
		10	9	49	3	1	0	0	0	0	0	4
		Σ	**87**	**780**	**53**	**5**	**4**	**2**	**3**	**3**	**9**	**79**

Appendix 1i. Summary statistics for WOZ-1, human–"machine–human" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
52	M	1	20	193	23	3	0	0	0	0	2	28
		2	13	71	1	0	0	0	0	1	1	3
		3	13	71	6	0	0	0	0	0	0	6
		4	11	73	0	0	0	0	0	0	0	0
		5	10	48	0	1	0	0	0	0	0	1
		6	14	80	0	0	0	0	0	0	0	0
		7	13	67	1	0	0	0	0	1	1	3
		8	7	42	1	0	0	0	0	0	0	1
		9	7	44	0	0	0	0	1	0	1	2
		10	13	66	0	1	0	0	2	1	1	5
		Σ	**121**	**755**	**32**	**5**	**0**	**0**	**3**	**3**	**6**	**49**
Σ	**46** **25M/21F**	**433**	**4023**	**27664**	**1622**	**815**	**156**	**20**	**41**	**167**	**215**	**3036**

Appendix 2: WOZ-2

Appendix 2a. Summary statistics for WOZ-2, human–"machine" corpus. Number of dialogs, utterances and words given for all subjects, as well as the seven major categories of disfluencies, broken down and summarized for subjects and dialogs. Sums for individual subjects are shown in cells with 5% shading, and sum totals are shown in cells with 10% shading. Omitted subjects do not appear in sum total.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
01	F	1	28	194	4	1	2	0	0	0	0	7
		2	23	172	15	2	0	1	1	0	2	21
		3	26	167	4	1	1	0	0	0	0	6
		Σ	**77**	**533**	**23**	**4**	**3**	**1**	**1**	**0**	**2**	**34**
02	M	1	40	134	21	9	2	0	1	0	1	34
		2	34	193	48	9	1	0	0	3	4	65
		3	41	210	57	5	1	1	0	3	3	70
		Σ	**115**	**537**	**126**	**23**	**4**	**1**	**1**	**6**	**8**	**169**
03	M	1	25	154	7	8	4	0	0	2	1	22
		2	39	310	27	26	10	1	0	1	4	69
		3	39	266	8	11	12	0	0	1	1	33
		Σ	**103**	**730**	**42**	**45**	**26**	**1**	**0**	**4**	**6**	**124**
04	F	1	30	265	29	8	4	1	0	2	3	47
		2	22	155	12	2	0	0	0	0	1	15
		3	24	165	19	5	0	0	0	0	1	25
		Σ	**76**	**585**	**60**	**15**	**4**	**1**	**0**	**2**	**5**	**87**
05	M	1	34	241	30	18	1	0	0	0	1	50
		2	17	128	10	7	2	0	0	0	0	19
		3	28	233	15	17	2	0	0	0	0	34
		Σ	**79**	**602**	**55**	**42**	**5**	**0**	**0**	**0**	**1**	**103**
06	M	1	36	127	7	1	1	0	0	2	2	13
		2	27	76	2	0	0	0	0	0	0	2
		3	45	149	5	0	0	0	0	0	0	5
		Σ	**108**	**352**	**14**	**1**	**1**	**0**	**0**	**2**	**2**	**20**
07	M	1	24	99	8	1	0	0	0	1	1	11
		2	11	90	10	4	1	0	0	1	1	17
		Σ	**35**	**189**	**18**	**5**	**1**	**0**	**0**	**2**	**2**	**28**
08	M	1	21	98	8	4	0	0	0	0	0	12
		2	40	216	14	4	10	1	0	0	1	30
		3	26	132	7	3	2	0	0	0	0	12
		Σ	**87**	**446**	**29**	**11**	**12**	**1**	**0**	**0**	**1**	**54**
09	M	1	22	128	10	6	1	0	0	0	1	18
		2	42	248	25	6	0	3	0	4	4	42
		3	25	155	10	1	0	0	0	1	1	13
		Σ	**89**	**531**	**45**	**13**	**1**	**3**	**0**	**5**	**6**	**73**
10	M	1	35	138	4	13	0	0	0	0	1	18
		2	37	145	10	7	0	0	0	2	5	24
		3	41	222	13	4	2	0	0	3	3	25
		Σ	**113**	**505**	**27**	**24**	**2**	**0**	**0**	**5**	**9**	**67**
11	M	1	31	287	37	24	3	0	0	2	4	70
		2	21	184	30	12	5	1	0	0	4	52
		3	36	320	27	21	0	0	0	1	3	52
		Σ	**88**	**791**	**94**	**57**	**8**	**1**	**0**	**3**	**11**	**174**
12	M	1	17	134	3	7	0	0	0	0	0	10
		2	26	244	23	7	1	2	0	2	4	39
		3	21	132	6	2	1	0	0	0	1	10
		Σ	**64**	**510**	**32**	**16**	**2**	**2**	**0**	**2**	**5**	**59**
13	M	1	46	205	10	17	1	1	0	0	2	31
		2	33	103	5	3	1	0	0	1	2	12
		3	49	189	18	4	3	0	0	0	0	25
		Σ	**128**	**497**	**33**	**24**	**5**	**1**	**0**	**1**	**4**	**68**
14	M	1	35	252	33	7	1	0	0	2	3	46
		2	29	238	36	5	2	0	0	3	5	51
		3	37	327	31	10	1	0	1	1	7	51
		Σ	**101**	**817**	**100**	**22**	**4**	**0**	**1**	**6**	**15**	**148**
15	–	No recording obtained.										
		Σ	–	–	–	–	–	–	–	–	–	–

Appendix 2b. Summary statistics for WOZ-2, human–"machine" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
16	F	Non-native speaker of Swedish. Omitted from analysis.										
		Σ	–	–	–	–	–	–	–	–	–	–
17	F	1	22	170	12	2	1	0	1	1	2	19
		2	33	369	36	1	0	0	0	1	4	42
		3	25	218	20	0	0	0	0	0	1	21
		Σ	**80**	**757**	**68**	**3**	**1**	**0**	**1**	**2**	**7**	**82**
18	F	1	19	110	9	3	2	0	0	0	1	15
		2	29	214	17	3	3	0	0	0	0	23
		3	20	130	7	0	0	0	0	0	0	7
		Σ	**68**	**454**	**33**	**6**	**5**	**0**	**0**	**0**	**1**	**45**
19	M	1	20	151	22	14	2	0	0	2	2	42
		2	13	76	8	3	3	0	0	0	0	14
		3	19	92	14	5	0	0	0	0	0	19
		Σ	**52**	**319**	**44**	**22**	**5**	**0**	**0**	**2**	**2**	**75**
20	M	1	27	132	4	12	1	0	0	0	0	17
		2	23	107	1	1	0	0	0	0	0	2
		3	36	160	7	1	2	0	0	0	0	10
		Σ	**86**	**399**	**12**	**14**	**3**	**0**	**0**	**0**	**0**	**29**
21	M	1	27	300	29	0	0	0	0	0	0	29
		2	16	162	24	3	0	1	0	1	2	31
		3	25	253	29	1	0	0	0	0	0	30
		Σ	**68**	**715**	**82**	**4**	**0**	**1**	**0**	**1**	**2**	**90**
22	F	1	16	110	8	8	0	0	0	0	1	17
		2	18	119	8	2	0	0	0	0	0	10
		3	22	141	8	2	0	0	0	0	4	14
		Σ	**56**	**370**	**24**	**12**	**0**	**0**	**0**	**0**	**5**	**41**
23	F	1	21	153	10	8	1	0	1	1	1	22
		2	32	282	18	15	1	0	0	1	1	36
		3	23	143	14	15	0	0	0	0	1	30
		Σ	**76**	**578**	**42**	**38**	**2**	**0**	**1**	**2**	**3**	**88**
24	M	1	12	23	0	2	0	0	0	0	0	2
		2	20	40	1	0	0	0	0	1	1	3
		3	24	61	4	2	0	0	0	3	3	12
		4	41	84	3	1	0	0	0	4	5	13
		Σ	**97**	**208**	**8**	**5**	**0**	**0**	**0**	**8**	**9**	**30**
25	M	1	16	145	18	7	0	0	1	2	2	30
		2	24	151	15	1	0	0	0	1	1	18
		3	17	120	6	1	1	0	0	0	0	8
		Σ	**57**	**416**	**39**	**9**	**1**	**0**	**1**	**3**	**3**	**56**
26	F	1	44	701	52	52	0	0	2	7	16	129
		2	17	230	26	25	0	0	0	3	6	60
		Σ	**61**	**931**	**78**	**77**	**0**	**0**	**2**	**10**	**22**	**189**
27	F	1	21	205	22	5	3	0	0	1	3	34
		2	24	258	27	2	1	1	0	1	2	34
		3	7	64	6	1	0	0	0	0	0	7
		Σ	**52**	**527**	**55**	**8**	**4**	**1**	**0**	**2**	**5**	**75**
28	M	1	16	154	16	7	0	1	0	0	2	26
		2	24	215	22	5	0	1	0	0	3	31
		3	17	152	20	3	0	0	0	0	0	23
		Σ	**57**	**521**	**58**	**15**	**0**	**2**	**0**	**0**	**5**	**80**
29	M	1	19	85	7	4	0	0	0	0	0	11
		2	18	87	7	3	0	0	0	0	0	10
		3	10	57	8	0	0	0	0	0	0	8
		Σ	**47**	**229**	**22**	**7**	**0**	**0**	**0**	**0**	**0**	**29**
30	M	1	25	181	27	7	0	0	0	0	0	34
		2	18	131	16	5	1	0	0	0	0	22
		3	14	108	9	0	0	0	0	0	0	9
		Σ	**57**	**420**	**52**	**12**	**1**	**0**	**0**	**0**	**0**	**65**
31	M	1	31	165	13	5	0	0	0	3	1	22
		2	27	117	2	2	0	1	0	1	1	7
		3	28	121	3	5	0	0	0	1	2	11
		Σ	**86**	**403**	**18**	**12**	**0**	**1**	**0**	**5**	**4**	**40**

Appendix 2c. Summary statistics for WOZ-2, human–"machine" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
32	F	1	22	213	28	5	4	0	0	0	1	38
		2	18	173	25	9	0	1	0	3	2	40
		3	17	155	12	2	2	0	0	0	0	16
		Σ	**57**	**541**	**65**	**16**	**6**	**1**	**0**	**3**	**3**	**94**
33	M	1	16	113	7	0	0	0	0	1	1	9
		2	16	78	7	0	1	0	0	0	0	8
		3	16	92	4	0	0	0	0	0	1	5
		Σ	**48**	**283**	**18**	**0**	**1**	**0**	**0**	**1**	**2**	**22**
34	M	1	29	182	19	17	0	2	0	2	9	49
		2	26	223	18	16	2	0	0	1	7	44
		3	32	181	11	11	1	0	0	2	6	31
		Σ	**87**	**586**	**48**	**44**	**3**	**2**	**0**	**5**	**22**	**124**
35	M	1	12	198	19	2	1	0	2	1	2	27
		2	11	201	26	3	0	0	0	0	3	32
		3	34	428	40	11	1	0	1	7	4	64
		Σ	**57**	**827**	**85**	**16**	**2**	**0**	**3**	**8**	**9**	**123**
36	M	1	14	215	15	14	0	0	0	2	2	33
		2	24	336	27	14	0	0	0	2	5	48
		3	14	171	15	8	0	0	0	3	3	29
		Σ	**52**	**722**	**57**	**36**	**0**	**0**	**0**	**7**	**10**	**110**
37	F	1	22	280	14	7	1	0	0	1	1	24
		2	26	280	7	3	1	0	0	0	0	11
		3	15	182	4	0	1	0	0	0	1	6
		Σ	**63**	**742**	**25**	**10**	**3**	**0**	**0**	**1**	**2**	**41**
38	F	1	19	98	1	1	0	0	1	0	0	3
		2	11	92	3	0	0	0	0	0	1	4
		3	12	107	5	2	0	0	1	0	0	8
		Σ	**42**	**297**	**9**	**3**	**0**	**0**	**2**	**0**	**1**	**15**
39	M	1	27	181	15	4	0	0	0	0	0	19
		2	22	148	13	5	1	0	0	3	4	26
		3	25	174	14	2	1	0	1	0	1	19
		Σ	**74**	**503**	**42**	**11**	**2**	**0**	**1**	**3**	**5**	**64**
40	F	1	19	269	17	8	1	0	0	2	0	28
		2	30	373	18	5	1	1	1	4	6	36
		3	17	175	6	3	0	0	0	0	0	9
		Σ	**66**	**817**	**41**	**16**	**2**	**1**	**1**	**6**	**6**	**73**
41	F	1	45	650	56	33	2	1	0	2	4	98
		2	24	322	45	16	2	0	0	2	3	68
		3	17	152	13	8	1	0	0	0	0	22
		Σ	**86**	**1124**	**114**	**57**	**5**	**1**	**0**	**4**	**7**	**188**
42	M	1	33	372	44	29	2	0	0	1	1	77
		2	50	484	47	54	0	0	3	2	5	111
		3	48	412	34	45	4	0	0	4	6	93
		Σ	**131**	**1268**	**125**	**128**	**6**	**0**	**3**	**7**	**12**	**281**
43	M	1	34	221	12	17	0	3	0	2	3	37
		2	30	228	13	19	1	1	0	2	5	41
		3	40	261	19	28	1	1	0	0	5	54
		Σ	**104**	**710**	**44**	**64**	**2**	**5**	**0**	**4**	**13**	**132**
44	–	No recording obtained.										
		Σ	–	–	–	–	–	–	–	–	–	–
45	M	1	19	182	12	15	0	0	1	2	3	33
		2	24	288	19	15	0	0	0	2	2	38
		3	17	141	9	7	0	0	0	0	0	16
		Σ	**60**	**611**	**40**	**37**	**0**	**0**	**1**	**4**	**5**	**87**
46	M	1	21	174	9	3	0	0	0	0	1	13
		2	19	115	6	3	0	0	0	0	0	9
		3	16	95	2	4	0	0	0	0	0	6
		Σ	**56**	**384**	**17**	**10**	**0**	**0**	**0**	**0**	**1**	**28**
47	M	1	20	383	16	13	0	0	2	5	8	44
		2	14	148	5	5	0	0	0	1	0	11
		3	16	138	7	4	0	1	1	0	1	14
		Σ	**50**	**669**	**28**	**22**	**0**	**1**	**3**	**6**	**9**	**69**

Appendix 2d. Summary statistics for WOZ-2, human–"machine" corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
48	F	1	16	175	9	2	0	0	0	0	0	11
		2	15	129	2	0	0	0	0	0	0	2
		3	34	317	18	1	0	0	0	0	1	20
		Σ	**65**	**621**	**29**	**3**	**0**	**0**	**0**	**0**	**1**	**33**
49	M	1	32	304	22	11	0	3	0	2	4	42
		2	22	181	15	7	0	0	0	0	0	22
		3	23	199	22	3	0	0	0	0	0	25
		Σ	**77**	**684**	**59**	**21**	**0**	**3**	**0**	**2**	**4**	**89**
Σ	**46** **32M/14F**	**137**	**3438**	**26261**	**2179**	**1040**	**132**	**31**	**22**	**134**	**257**	**3795**

Appendix 3: Nymans

Appendix 3. Summary statistics for Nymans, human–human corpus. Number of dialogs, utterances and words given for all subjects, as well as the seven major categories of disfluencies, broken down and summarized for subjects and dialogs. Sums for individual subjects are shown in cells with 5% shading, and sum totals are shown in cells with 10% shading.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
01	M	1	27	286	12	8	3	0	1	7	7	38
		2	54	603	29	21	7	0	0	17	25	99
		3	54	467	24	8	8	0	2	15	9	66
		Σ	**135**	**1356**	**65**	**37**	**18**	**0**	**3**	**39**	**41**	**203**
02	M	1	54	190	17	14	8	0	0	0	0	39
		2	97	371	25	29	24	0	0	0	2	80
		3	61	251	22	14	14	0	0	1	2	53
		Σ	**212**	**812**	**64**	**57**	**46**	**0**	**0**	**1**	**4**	**172**
03	M	1	79	342	27	10	5	0	1	11	10	64
		2	44	134	6	2	2	0	0	6	2	18
		3	90	494	30	10	7	0	0	15	7	69
		Σ	**213**	**970**	**63**	**22**	**14**	**0**	**1**	**32**	**19**	**151**
04	F	1	80	420	31	12	8	1	0	6	6	64
		2	107	712	37	11	8	0	0	10	15	81
		3	118	664	31	9	9	1	1	4	17	72
		Σ	**305**	**1796**	**99**	**32**	**25**	**2**	**1**	**20**	**38**	**217**
05	M	1	82	456	28	4	2	0	0	14	10	58
		2	90	438	21	0	0	1	0	5	10	37
		3	82	410	16	5	1	0	0	7	11	40
		Σ	**254**	**1304**	**65**	**9**	**3**	**1**	**0**	**26**	**31**	**135**
06	F	1	43	205	12	13	2	1	0	3	6	37
		2	69	311	12	7	3	0	0	0	1	23
		3	82	303	26	16	2	1	0	1	2	48
		Σ	**194**	**819**	**50**	**36**	**7**	**2**	**0**	**4**	**9**	**108**
07	M	1	82	461	31	4	1	0	0	4	4	44
		2	96	490	13	0	0	1	2	4	6	26
		3	43	218	6	1	1	0	1	0	0	9
		Σ	**221**	**1169**	**50**	**5**	**2**	**1**	**3**	**8**	**10**	**79**
08	M	1	78	366	34	4	5	2	1	8	13	67
		2	65	338	33	0	4	0	1	2	4	44
		3	57	320	39	1	5	1	0	1	3	50
		Σ	**200**	**1024**	**106**	**5**	**14**	**3**	**2**	**11**	**20**	**161**
Σ	**8** 6M/2F	**24**	**1734**	**9250**	**562**	**203**	**129**	**9**	**10**	**141**	**172**	**1226**

Appendix 4: Bionic

Appendix 4a. Summary statistics for Bionic, human–machine corpus. Number of dialogs, utterances and words given for all subjects, as well as the seven major categories of disfluencies, broken down and summarized for subjects and dialogs. Sums for individual subjects are shown in cells with 5% shading, and sum totals are shown in cells with 10% shading.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
01	M	1	23	137	3	12	11	0	0	4	5	35
		2	31	217	20	19	8	0	0	7	7	61
		3	21	161	8	12	10	0	0	4	6	40
		4	23	178	4	12	6	2	0	2	2	28
		5	17	126	8	5	4	0	0	0	0	17
		Σ	**115**	**819**	**43**	**60**	**39**	**2**	**0**	**17**	**20**	**181**
02	F	1	29	79	1	4	1	0	0	1	0	7
		2	30	116	7	4	1	0	0	0	0	12
		3	40	188	9	4	2	1	0	1	2	19
		4	27	112	14	4	2	0	0	0	0	20
		Σ	**126**	**495**	**31**	**16**	**6**	**1**	**0**	**2**	**2**	**58**
03	F	1	29	123	14	4	4	0	0	0	0	22
		2	38	250	28	8	3	1	0	3	7	50
		3	29	160	12	7	5	0	0	0	4	28
		4	63	367	24	2	2	0	0	5	6	39
		Σ	**159**	**900**	**78**	**21**	**14**	**1**	**0**	**8**	**17**	**139**
04	M	1	24	149	19	7	5	0	0	1	2	34
		2	27	137	14	2	1	0	0	1	1	19
		3	23	124	21	0	0	0	1	0	0	22
		4	6	59	10	1	0	0	0	0	0	11
		Σ	**80**	**469**	**64**	**10**	**6**	**0**	**1**	**2**	**3**	**86**
05	M	1	33	183	28	14	14	0	0	2	2	60
		2	27	125	17	12	0	0	0	0	0	29
		3	11	49	11	9	1	0	0	2	1	24
		4	27	158	24	17	3	0	2	1	4	51
		Σ	**98**	**515**	**80**	**52**	**18**	**0**	**2**	**5**	**7**	**164**
06	F	1	20	226	16	15	5	1	1	6	5	49
		2	54	468	26	36	2	0	3	2	2	71
		3	29	198	1	11	3	0	0	1	1	17
		4	27	168	7	7	1	0	0	2	2	19
		Σ	**130**	**1060**	**50**	**69**	**11**	**1**	**4**	**11**	**10**	**156**
07	M	1	26	167	21	2	1	0	0	1	1	26
		2	26	126	19	2	1	0	0	0	0	22
		3	28	113	12	1	0	0	0	0	0	13
		4	32	100	9	4	1	0	0	0	0	14
		Σ	**112**	**506**	**61**	**9**	**3**	**0**	**0**	**1**	**1**	**75**
08	F	1	34	342	43	8	7	1	0	3	2	64
		2	35	210	35	4	10	0	0	5	5	59
		3	26	149	21	3	2	0	0	0	0	26
		4	25	90	10	1	1	0	0	1	0	13
		Σ	**120**	**791**	**109**	**16**	**20**	**1**	**0**	**9**	**7**	**162**
09	M	1	35	98	16	2	2	0	0	4	4	28
		2	40	132	26	6	1	1	0	0	1	35
		3	22	70	20	2	0	0	0	0	0	22
		4	44	203	41	2	1	0	0	2	3	49
		5	13	74	15	1	0	0	0	0	2	18
		Σ	**154**	**577**	**118**	**13**	**4**	**1**	**0**	**6**	**10**	**152**
10	M	1	14	52	12	5	0	0	0	0	1	18
		2	15	75	13	6	3	0	0	0	0	22
		3	13	48	9	3	0	0	0	0	0	12
		4	24	135	27	15	3	0	0	0	0	45
		5	24	126	17	12	0	0	0	0	0	29
		Σ	**90**	**436**	**78**	**41**	**6**	**0**	**0**	**0**	**1**	**126**
11	F	1	28	321	33	2	3	1	1	11	11	62
		2	59	577	50	2	13	1	1	13	16	96
		3	33	406	28	3	4	5	0	7	7	54
		4	43	527	53	2	3	2	0	7	11	78
		Σ	**163**	**1831**	**164**	**9**	**23**	**9**	**2**	**38**	**45**	**290**

Appendix 4b. Summary statistics for Bionic, human–machine corpus.

Subject	Sex	Dialog	Utts	Words	UPs	FPs	PRs	EETs	MPs	TRs	REPs	Σ DFs
12	M	1	43	311	44	31	2	1	5	3	6	92
		2	23	259	34	16	3	1	9	7	8	78
		3	68	315	30	26	5	1	1	6	9	78
		4	18	114	8	8	1	0	0	1	2	20
		Σ	**151**	**999**	**116**	**81**	**11**	**3**	**15**	**17**	**25**	**268**
13	F	1	32	318	14	24	4	2	1	6	9	60
		2	32	347	19	21	3	2	1	3	5	54
		3	26	393	20	19	2	2	2	1	7	53
		4	30	314	11	15	4	2	2	2	5	41
		Σ	**120**	**1372**	**64**	**79**	**13**	**8**	**6**	**12**	**26**	**208**
14	M	1	24	262	19	11	4	0	0	6	7	47
		2	31	244	29	17	5	0	0	3	5	59
		3	33	330	21	17	2	0	0	3	3	46
		4	37	258	30	5	4	0	0	2	4	45
		Σ	**125**	**1094**	**99**	**50**	**15**	**0**	**0**	**14**	**19**	**197**
15	M	1	29	79	1	1	4	0	0	1	2	9
		2	44	152	1	0	1	0	0	0	0	2
		3	24	108	4	1	1	1	0	0	1	8
		4	32	122	1	0	0	0	0	0	0	1
		Σ	**129**	**461**	**7**	**2**	**6**	**1**	**0**	**1**	**3**	**20**
16	F	1	32	140	18	2	1	0	0	2	2	25
		2	36	180	18	9	0	0	1	1	2	31
		3	29	122	21	2	1	0	0	0	1	25
		4	16	82	7	2	0	0	0	0	1	10
		Σ	**113**	**524**	**64**	**15**	**2**	**0**	**1**	**3**	**6**	**91**
Σ	**16** **9M/7F**	**67**	**1985**	**12849**	**1226**	**543**	**197**	**28**	**31**	**146**	**202**	**2373**

Appendix 5: Transcription sample

Appendix 5.1. An example of a fully transcribed dialog. The first dialog from **`sentences_vertical`**, Bionic corpus, subject 5.

```
57.440682                   D1
58.557013                   S-09\DF-06
59.357937           ff<
59.904427                   öh
59.917040           >f
60.111288                   jag
60.207151                   vill
60.769716           p<
60.991715           {-a}
61.294440           >p
61.300000                   boka
61.307049           p<
61.433184           {u-}
61.534093           >p
62.550000                   [
62.596154                   umeå
62.613811           u<
63.118353           >u
63.120000                   +
63.136009           f<
63.610279                   eh
63.622893           >f
64.150000           i1
64.170320                   resa
64.500000           i2
64.571432                   till
65.000000           r1
65.091109                   umeå
65.300000                   ]
65.308061                   E


76.132981                   S-01\DF-00
76.889795                   stockholm
77.134498                   E


86.440752                   S-11\DF-09
86.653599           ff<
87.152156                   uhh
87.167293           >f
87.507858                   jag
87.732379                   vill
88.370623           p<
88.481622           {-a}
88.557304           >p
88.567395                   åka
88.587574           u<
89.026526           >u
89.356999                   den
89.508361           p<
89.642064           {f-}
89.717746           >p
90.630967                   fjärde
```

```
90.643578            u<
91.095144            >u
91.125415            p<
91.246505            {m-}
91.405436            >p
91.930160                   maj
91.945295            u<
92.459928            >u
93.421079                   alternativt
93.948324                   den
93.960934            u<
94.223296            >u
94.922087                   tredje
94.934700            u<
95.081018            >u
95.580513                   maj
95.787375                   E


103.539642                  S-07\DF-00
103.991207                  jag
104.162752                  vill
104.425114                  vara
105.123905                  framme
106.132985                  klockan
106.768706                  tio
107.530565                  nollnoll
107.729856                  E


113.847412                  S-01\DF-00
114.417544                  ja
114.692520                  E


116.034596                  S-01\DF-00
116.561843                  ja
116.761137                  E


148.839843                  S-05\DF-05
148.950842    _ingr<
149.156122                  a(ingr)
149.187977    >ingr_
149.203113           u<
149.394839           >u
149.407452           f<
150.699081                  öhhn
150.719262           >f
150.731873           p<
150.825213           {j-}
150.916031           >p
151.382732                  ja
151.392822           u<
151.690502           >u
152.762653                  alternativet
152.775263           u<
152.901399           >u
153.431169                  tåg
153.562350                  E
```

```
179.536109                  S-06\DF-03
179.832206          ff<
180.580512                  öh
180.595648          >f
180.953870                  den
181.793933                  tredje
182.131974                  i
182.152156          u<
182.270722          >u
183.085558                  femte
183.095648          u<
183.226829          >u
184.160233                  nattåg
184.359524                  E


245.350943                  S-16\DF-05
245.624335          ff<
246.231370                  öh
246.243983          >f
246.715730                  jag
246.952865                  vill
247.780312                  boka
247.792921          u<
248.158714          >u
248.171328          p<
248.289895          {f-}
248.383236          >p
249.165277                  flyg
249.175363          u<
249.642065          >u
250.068403                  som
250.300491                  är
250.752056                  framme
250.767192          p<
250.885760          {i}
250.963964          >p
251.138031                  i
251.614821                  umeå
252.063864                  noll
252.386768                  nio
252.969514                  nollnoll
253.236919                  den
253.577485                  fjärde
254.059323                  femte
254.306546                  E


284.869108                  S-07\DF-00
285.805034                  ja
285.958919                  jag
286.087577                  vill
286.501299                  boka
287.598678                  transfer
288.597669                  stockholm
289.435204                  arlanda
289.599180                  E
```

```
307.005841                  S-15\DF-12
307.168230          ff<
307.939240                  eh
307.951854          >f
308.320169          p<
308.484145          {-k-}
308.572440          >p
309.238436                  okej
309.258614          f<
310.171836                  eh
310.181927          >f
310.207149          u<
310.888281          >u
312.611292                  återresa
312.633993          u<
313.796963          >u
313.814619          f<
314.427638                  eh
314.440251          >f
315.108767                  från
315.777285                  umeå
315.789896          u<
316.072439          >u
316.264165                  den
316.276777          p<
316.347413          {sj-}
316.413003          >p
317.184953                  sjätte
317.563358                  i
317.575971          u<
317.742470          >u
318.368100                  femte
318.378188          u<
318.663255          >u
318.860026          p<
318.925617          {-f-}
319.001298          >p
319.344387                  efter
319.354477          u<
319.505840          >u
320.197062                  klockan
320.923602                  sjutton
321.425619                  nollnoll
321.574460                  E


353.918049                  S-02\DF-00
354.561341                  svar
354.818657                  ja
355.093633                  E


385.625918                  S-02\DF-00
385.984143                  svar
386.352459                  ja
386.491208                  E
```

```
404.599177                  S-05\DF-00
405.197060                  ja
405.358513                  jag
405.487171                  vill
405.731875                  boka
406.254076                  hotell
406.614824                  E


417.848423                  S-02\DF-01
418.453873                  ndj/
418.829757                  ja
418.923098                  E


431.188480                  S-10\DF-02
431.345828          ff<
432.391814                  eh
432.409473          >f
432.422085          p<
432.517948          {n-}
432.583538          >p
433.226831                  natt
433.943276                  mot
434.389796                  den
435.154175                  femte
435.358515                  och
435.706648                  natt
436.102712                  mot
436.261643                  den
436.793935                  sjätte
436.957909                  E


463.244488                  S-01\DF-00
463.789394                  ja
464.071938                  E


492.550740                  S-04\DF-01
493.188986                  ligger
493.580007                  strand
494.225821                  hotell
494.235910          u<
494.407454          >u
495.260131                  centralt
495.361039                  E


514.831268                  S-03\DF-00
515.378696                  finns
515.734396                  där
516.546709                  restaurant
516.698072                  E
```

```
527.891306                  S-14\DF-05
528.257099                  jag
528.648119                  måste
528.900391                  ha
528.973549                  [
529.036617                  h/
529.175366                  +
529.566387          r1
529.584046                  hotell
529.599182                  ]
529.609270          f<
529.765678                  eh
529.775769          >f
530.186971                  med
531.102713                  restaurant
531.112400          u<
531.995494          >u
532.473731                  finns
532.602779                  det
533.303689                  alternativ
533.313816          u<
533.412500          >u
533.784462                  till
534.093165                  strand
534.475248                  hotell
534.774718                  E


557.445902                  S-03\DF-00
558.043066                  ligger
558.377072                  det
559.186786                  centralt
559.449943                  E


580.707542                  S-03\DF-00
581.190839                  finns
581.514725                  det
582.488911                  restaurant
582.698930                  E


598.609870                  S-04\DF-00
599.103289                  då
599.475251                  bokar
599.614420                  jag
599.935775                  det
600.120491                  E


620.763210                  S-06\DF-00
621.213613                  jag
621.350252                  vill
621.651364                  ha
622.073933                  rum
622.607840                  för
623.389723                  rökare
623.561787                  E
```

```
649.834560                  S-01\DF-00
650.674642                  tack
650.775857                  E


651.426164                  S-04\DF-03
651.638462          ff<
652.395290                  ehm
652.415533          >f
652.428187          u<
652.815332          >u
652.827983          p<
652.921606          {d-}
653.022821          >p
653.475754                  då
653.802175                  vill
654.257639                  jag
654.460178                  E


656.980426                  S-14\DF-01
657.137958          ff<
657.633976                  eh
657.641567          >f
658.183065                  svar
658.560087                  nej
658.754929                  jag
658.848552                  vill
658.916871                  ha
658.985191                  en
659.799965                  bekräftelse
660.055535                  på
660.475574                  vad
660.753913                  som
661.006951                  nu
661.232152                  är
661.801482                  bokat
661.971016                  E


667.297441                  S-05\DF-01
667.430591          ff<
668.134988                  eh
668.147640          >f
668.785290                  martin
669.286301                  nio
669.923955                  rudolf
670.478102                  fyrtiotvå
670.602096                  E


686.404222                  S-01\DF-00
687.019098                  ja
687.188632                  E
```

```
733.205841                  S-10\DF-04
733.899158                  ja
733.906751          u<
734.091467          >u
734.192681                  jag
734.344502                  vill
734.493793                  ha
734.503916          p<
734.638025          {b-}
734.731648          >p
735.250371                  buss
735.260494          u<
735.442680          >u
736.366259                  stockholm
736.373857          u<
736.596529          >u
737.520108                  arlanda
738.221021                  arlanda
739.063629                  stockholm
739.835390                  bussbiljett
740.022637                  E


771.006955                  S-06\DF-02
771.409281                  (ja)ha
771.424463          u<
771.753409          >u
771.834381                  då
771.884988                  var
772.044400                  jag
772.317678                  klar
772.327802          u<
772.479623          >u
772.841463                  tack
772.993283                  E


795.862724                  S-01\DF-00
796.161306                  nej
796.303006                  E


808.221022                  S-03\DF-00
808.663834                  hm
808.987719                  tack
809.445713                  själv
809.579822                  E
```

Appendix 5.2. An example of a fully transcribed dialog. The first dialog from **`sentences_horizontal`**, Bionic corpus, subject 5. Not that line breaks here are due to the A4 format of this thesis. In the original file, each utterance was represented by one line of characters.

```
D1

S-09\DF-06 ff< öh >f jag vill p< {-a} >p boka p< {u-} >p [ umeå u< >u + f<
eh >f i1 resa i2 till r1 umeå ] E


S-01\DF-00 stockholm E


S-11\DF-09 ff< uhh >f jag vill p< {-a} >p åka u< >u den p< {f-} >p fjärde
u< >u p< {m-} >p maj u< >u alternativt den u< >u tredje u< >u maj E


S-07\DF-00 jag vill vara framme klockan tio nollnoll E


S-01\DF-00 ja E


S-01\DF-00 ja E


S-05\DF-05 _ingr< a(ingr) >ingr_ u< >u f< öhhn >f p< {j-} >p ja u< >u
alternativet u< >u tåg E


S-06\DF-03 ff< öh >f den tredje i u< >u femte u< >u nattåg E


S-16\DF-05 ff< öh >f jag vill boka u< >u p< {f-} >p flyg u< >u som är
framme p< {i} >p i umeå noll nio nollnoll den fjärde femte E


S-07\DF-00 ja jag vill boka transfer stockholm arlanda E


S-15\DF-12 ff< eh >f p< {-k-} >p okej f< eh >f u< >u återresa u< >u f< eh
>f från umeå u< >u den p< {sj-} >p sjätte i u< >u femte u< >u p< {-f-} >p
efter u< >u klockan sjutton nollnoll E


S-02\DF-00 svar ja E


S-02\DF-00 svar ja E


S-05\DF-00 ja jag vill boka hotell E


S-02\DF-01 ndj/ ja E


S-10\DF-02 ff< eh >f p< {n-} >p natt mot den femte och natt mot den sjätte
E
```

S-01\DF-00 ja E

S-04\DF-01 ligger strand hotell u< >u centralt E

S-03\DF-00 finns där restaurant E

S-14\DF-05 jag måste ha [h/ + r1 hotell] f< eh >f med restaurant u< >u finns det alternativ u< >u till strand hotell E

S-03\DF-00 ligger det centralt E

S-03\DF-00 finns det restaurant E

S-04\DF-00 då bokar jag det E

S-06\DF-00 jag vill ha rum för rökare E

S-01\DF-00 tack E

S-04\DF-03 ff< ehm >f u< >u p< {d-} >p då vill jag E

S-14\DF-01 ff< eh >f svar nej jag vill ha en bekräftelse på vad som nu är bokat E

S-05\DF-01 ff< eh >f martin nio rudolf fyrtiotvå E

S-01\DF-00 ja E

S-10\DF-04 ja u< >u jag vill ha p< {b-} >p buss u< >u stockholm u< >u arlanda arlanda stockholm bussbiljett E

S-06\DF-02 (ja)ha u< >u då var jag klar u< >u tack E

S-01\DF-00 nej E

S-03\DF-00 hm tack själv E

Postlude

The following work:

Wilkes, Kathleen V. 1988. —, yìshì, duh, um, and consciousness. In: A. J. Marcel & E. Bisiach (eds.), *Consciousness in Contemporary Science*, Oxford: Clarendon Press, ch. 2, pp. 16–41.

… is not about unfilled pauses, interjections, filled pauses, and their relation to consciousness, which I hoped when I came across this reference—after all, that is exactly the kind of study I would like to see done, which should be obvious to anyone who managed to get through chapter two, especially the section on consciousness research.

Wilkes treats the vocabulary for the concept ‘consciousness’ in Greek (no word, hence —), Mandarin (*yìshì*), Slovenian (*duh* and *um*) and English (*consciousness*). For those of you who are interested in the philosophy of consciousness—especially from a diachronic perspective—it is recommended reading, so long as the potential reader is aware of the fact that it has nothing to do with disfluency at all.
