Distilling dialogues - A method using natural dialogue corpora for dialogue systems development

Arne Jönsson and Nils Dahlbäck
Department of Computer and Information Science
Linköping University
S-581 83, LINKÖPING
SWEDEN
nilda@ida.liu.se, arnjo@ida.liu.se

Abstract

We report on a method for utilising corpora collected in natural settings. It is based on distilling (re-writing) natural dialogues to elicit the type of dialogue that would occur if one the dialogue participants was a computer instead of a human. The method is a complement to other means such as Wizard of Oz-studies and un-distilled natural dialogues. We present the distilling method and guidelines for distillation. We also illustrate how the method affects a corpus of dialogues and discuss the pros and cons of three approaches in different phases of dialogue systems development.

1 Introduction

It has been known for quite some time now, that the language used when interacting with a computer is different from the one used in dialogues between people, (c.f. Jönsson and Dahlbäck (1988)). Given that we know that the language will be different, but not *how* it will be different, we need to base our development of natural language dialogue systems on a relevant set of dialogue corpora. It is our belief that we need to clarify a number of different issues regarding the collection and use of corpora in the development of speech-only and multimodal dialogue systems. Exchanging experiences and developing guidelines in this area are as important as, and in some sense a necessary pre-requisite to, the development of computational models of speech, language, and dialogue/discourse. It is interesting to note the difference in the state of art in the field of natural language dialogue systems with that of corpus linguistics, where issues of the usefulness of different samples, the necessary sampling size, representativeness in corpus design and other have been discussed for quite some time (e.g. (Garside et al., 1997; Atkins et al., 1992; Crowdy, 1993; Biber, 1993)). Also the neighboring area of evaluation of NLP systems (for an overview, see Sparck Jones and Galliers (1996)) seems to have advanced further.

Some work have been done in the area of natural language dialogue systems, e.g. on the design of Wizard of Oz-studies (Dahlbäck et al., 1998), on measures for inter-rater reliability (Carletta, 1996), on frameworks for evaluating spoken dialogue agents (Walker et al., 1998) and on the use of different corpora in the development of a particular system (The Carnegie-Mellon Communicator, Eskenazi et al. (1999)).

The question we are addressing in this paper is how to collect and analyse relevant corpora. We begin by describing what we consider to be the main advantages and disadvantages of the two currently used methods; studies of human dialogues and Wizard of Oz-dialogues, especially focusing on the ecological validity of the methods. We then describe a method called 'distilling dialogues', which can serve as a supplement to the other two.

2 Natural and Wizard of Oz-Dialogues

The advantage of using real dialogues between people is that they will illustrate which tasks and needs that people actually bring to a particular service provider. Thus, on the level of the users' general goals, such dialogues have a high validity. But there are two drawbacks here. First; it is not self-evident that users will have the same task expectations from a computer system as they have with a person. Second, the language used will differ from the language used when interacting with a computer.

These two disadvantages have been the major force behind the development of Wizard of Oz-methods. The advantage here is that the setting will be human-computer interaction. But there are important disadvantages, too. First, on the practical side, the task of setting up a high quality simulation environment and training the operators ('wizards') to use this is a resource consuming task (Dahlbäck et al., 1998). Second, and probably even more important, is that we cannot then observe real users using a system for real life tasks, where they bring their own needs, motivations, resources, and constraints to bear. To some extent this problem can be overcome using well-designed so called 'scenarios'. As pointed out in Dahlbäck (1991), on many levels of analysis the artificiality of the situation will not affect the language used. An example of this is the pattern of pronoun-antecedent relations. But since the tasks given to the users are often pre-described by the researchers, this means that this is not a good way of finding out *which* tasks the users actually want to perform. Nor does it provide a clear enough picture on how the users will act to find something that satisfies their requirements. If e.g. the task is one of finding a charter holiday trip or buying a TV-set within a specified set of constraints (economical and other), it is conceivable that people will stay with the first item that matches the specification, whereas in real life they would probably look for alternatives. In our experience, this is primarily a concern if the focus is on the users' goals and plans, but is less a problem when the interest is on lower-level aspects, such as, syntax or patterns of pronoun-antecedent relationship (c.f. Dahlbäck (1991)).

To summarize; real life dialogues will provide a reasonably correct picture of the way users' approach their tasks, and what tasks they bring to the service provider, but the language used will not give a good approximation of what the system under construction will need to handle. Wizard of Oz-dialogues, on the other hand, will give a reasonable approximation of some aspects of the language used, but in an artificial context.

The usual approach has been to work in three steps. First analyse real human dialogues, and based on these, in the second phase, design one or more Wizard of Oz-studies. The final step is to fine-tune the system's performance on real users. A good example of this method is presented in Eskenazi et al. (1999). But there are also possible problems with this approach (though we are not claiming that this was the case in their particular project). Eskenazi et al. (1999) asked a human operator to act 'computer-like' in their Wizard of Oz-phase. The advantage is of course that the human operator will be able to perform all the tasks that is usually provided by this service. The disadvantage is that it puts a heavy burden on the human operator to act as a computer. Since we know that lay-persons' ideas of what computers can and cannot do are in many respects far removed from what is actually the case, we risk introducing some systematic distortion here. And since it is difficult to perform consistently in similar situations, we also risk introducing non-systematic distortion here, even in those cases when the 'wizard' is an NLP-professional.

Our suggestion is therefore to supplement the above mentioned methods, and bridge the gap between them, by post-processing human dialogues to give them a computer-like quality. The advantage, compared to having people do the simulation on the fly, is both that it can be done with more consistency, and also that it can be done by researchers that actually know what human-computer natural language dialogues can look like. A possible disadvantage with using both Wizard of Oz-and real computer dialogues, is that users will quickly adapt to what the system can provide them with, and will therefore not try to use it for tasks they know it cannot perform. Consequently, we will not get a full picture of the different services they would like the system to provide.

A disadvantage with this method is, of course, that post-processing takes some time compared to using the natural dialogues as they are. There is also a concern on the ecological validity of the results, as discussed later.

3 Distilling dialogues

Distilling dialogues, i.e. re-writing human interactions in order to have them reflect what a human-computer interaction could look like involves a number of considerations. The main issue is that in corpora of natural dialogues one of the interlocutors is not a dialogue system. The system's task is instead performed by a human and the problem is how to anticipate the behaviour of a system that does not exist based on the performance of an agent with different performance characteristics. One important aspect is how to deal with human features that are not part of what the system is supposed to be able to handle, for instance if the user talks about things outside of the domain, such as discussing an episode of a recent TV show. It also involves issues on how to handle situations where one of the interlocuters discusses with someone else on a different topic, e.g. discussing the up-coming Friday party with a friend in the middle of an information providing dialogue with a customer.

It is important for the distilling process to have at least an outline of the dialogue system that is under development: Will it for instance have the capacity to recognise users' goals, even if not explicitly stated? Will it be able to reason about the discourse domain? What services will it provide, and what will be outside its capacity to handle?

In our case, we assume that the planned dialogue system has the ability to reason on various aspects of dialogue and properties of the application. In our current work, and in the examples used for illustration in this paper, we assume a dialogue model that can handle any relevant dialogue phenomenon and also an interpreter and speech recogniser being able to understand any user input that is relevant to the task. There is is also a powerful domain reasoning module allowing for more or less any knowledge reasoning on issues that can be accomplished within the domain (Flycht-Eriksson, 1999). Our current system does, however, not have an explicit user task model, as opposed to a system task model (Dahlbäck and Jönsson, 1999), which is included, and thus, we can not assume that the 'system' remembers utterances where the user explains its task. Furthermore, as our aim is system development we will not consider interaction outside the systems capabilities as relevant to include in the distilled dialogues.

The context of our work is the development a multi-modal dialogue system. However, in our current work with distilling dialogues, the abilities of a multi-modal system were not fully accounted for. The reason for this is that the dialogues would be significantly affected, e.g. a telephone conversation where the user always likes to have *the next connection, please* will result in a table if multi-modal output is possible and hence a fair amount of the dialogue is removed. We have therefore in this paper analysed the corpus assuming a speech-only system, since this is closer to the original telephone conversations, and hence needs fewer assumptions on system performance when distilling the dialogues.

4 Distillation guidelines

Distilling dialogues requires guidelines for how to handle various types of utterances. In this section we will present our guidelines for distilling a corpus of telephone conversations between a human information provider on local buses[1] to be used for developing a multimodal dialogue system (Qvarfordt and Jönsson, 1998; Flycht-Eriksson and Jönsson, 1998; Dahlbäck et al., 1999; Qvarfordt, 1998). Similar guidelines are used within another project on developing Swedish Dialogue Systems where the domain is travel bureau information.

We can distinguish three types of contributors: 'System' (i.e. a future systems) utterances, User utterances, and other types, such as moves by other speakers, and noise.

4.1 Modifying system utterances

The problem of modifying 'system' utterances can be divided into two parts: *how* to change and *when* to change. They are in some respects intertwined, but as the how-part affects the when-part more we will take this as a starting point.

- *The 'system' provides as much relevant information as possible at once.* This depends on the capabilities of the systems output modalities. If we have a screen or similar output device we present as much as possible which normally is all relevant information. If we, on the other hand, only have spoken output the amount of information that the hearer can interpret in one utterance must be considered when distilling. The system might in such cases provide less information. The principle of providing all relevant information is based on the assumption that a computer system often has access to all relevant information when querying the background system and can also present it more conveniently, especially in a multimodal system (Ahrenberg et al., 1996). A typical example is the dialogue fragment in figure 1. In this fragment the system provides information on what train to take and how to change to a bus. The result of distilling this fragment provides the revised fragment of figure 2. As seen in the fragment of figure 2 we also remove a number of utterances typical for human interaction, as discussed below.
- *System utterances are made more computer-like and do not include irrelevant information.* The latter is seen in S9 in the dialogue in figure 3 where the provided information is not relevant. It could also be possible to remove S5 and respond with S7 at once. This, however, depends on if the information grounded in S5-U6 is needed for the 'system' in order to know the arrival time or if that could be concluded from U4. This in turn depends on the system's capabilities. If we assume that the dialogue system has a model of user tasks, the information in S5-U6 could have been concluded from that. We will, in this case, retain S5-U6 as we do not assume a user task model (Dahlbäck and Jönsson, 1999) and in order to stay as close to the original dialogue as possible.

The next problem concerns the case when 'system' utterances are changed or removed.

- *Dialogue contributions provided by something or someone other than the user or the 'system' are removed.* These are regarded as not being part of the interaction. This means that if someone interrupts the current interaction, say that the telephone rings during a face-to-face interaction, the interrupting interaction is normally removed from the corpus.

  Furthermore, 'system' interruptions are removed. A human can very well interrupt another human interlocuter, but a computer system will not do that.

  However, this guideline could lead to problems, for instance, when users follow up such interruptions. If no information is provided or the interrupted sequence does not affect the dialogue, we have no problems removing the interruption. The problem is what to do when information from the 'system' is used in the continuing dialogue. For such cases we have no fixed strategy,

[1] The bus time table dialogues are collected at Linköping University and are available (in Swedish) on http://www.ida.liu.se/~arnjo/kfb/dialoger.html

U4: yes I wonder if you have any mm buses or (.) like express buses leaving from Linköping to Vadstena (.) on sunday
ja ville undra om ni hade några öh bussar eller (.) typ expressbussar som åkte från Linköping till Vadstena (.) på sönda

S5: no the bus does not run on sundays
nej bussen går inte på söndagar

U6: how can you (.) can you take the train and then change some way (.) because (.) to Mjölby 'n' so
hur kan man (.) kan man ta tåg å sen byta på nått sätt (.) för de (.) till mjölby å så

S7: that you can do too yes
de kan du göra också ja

U8: how (.) do you have any such suggestions
hur (.) har du nåra några såna förslag

S9: yes let's see (4s) a moment (15s) now let us see here (.) was it on the sunday you should travel
ja ska se här (4s) ett ögonblick (15s) nu ska vi se här (.) va de på söndagen du skulle åka på

U10: yes right afternoon preferably
ja just de eftermidda gärna

S11: afternoon preferable (.) you have train from Linköping fourteen twenty nine
eftermidda gärna (.) du har tåg från Linköping fjorton å tjugonie

U12: mm
mm

S13: and then you will change from Mjölby station six hundred sixty
å så byter du från Mjölby station sexhundrasexti

U14: sixhundred sixty
sexhundrasexti

S15: fifteen and ten
femton å tie

Figure 1: Dialogue fragment from a real interaction on bus time-table information

U4: I wonder if you have any buses or (.) like express buses going from Linköping to Vadstena (.) on sunday

S5: no the bus does not run on sundays

U6: how can you (.) can you take the train and then change some way (.) because (.) to Mjölby and so

S7: you can take the train from Linköping fourteen and twenty nine and then you will change at Mjölby station to bus six hundred sixty at fifteen and ten

Figure 2: A distilled version of the dialogue in figure 1

the dialogue needs to be rearranged depending on how the information is to be used (c.f. the discussion in the final section of this paper).

- *'System' utterances which are no longer valid are removed.* Typical examples of this are the utterances S7, S9, S11 and S13 in the dialogue fragment of figure 1.
- *Remove sequences of utterances where the 'system' behaves in a way a computer would not do.* For instance jokes, irony, humor, commenting on the other dialogue participant, or dropping the telephone (or whatever is going on in S7 in figure 4). A common case of this is when the 'system' is talking while looking for information, S5 in the dialogue fragment of figure 4 is an example of this. Related to this is when the system provides its own comments. If we can assume that it has such capabilities they are included, otherwise we remove them.
- *The system does not repeat information that has already been provided unless explicitly asked to do so.* In human interaction it is not uncommon to repeat what has been uttered for purposes other than to provide grounding information or feedback. This is for instance common during

U4: 'n' I must be at Resecentrum before fourteen and thirty five (.) 'cause we will going to the interstate buses
å ja ska va på rececentrum innan fjorton å trettifem (.) fö vi ska till långfärdsbussarna
S5: aha (.) 'n' then you must be there around twenty past two something then
jaha (.) å då behöver du va nere strax efter tjuge över två nånting då
U6: yes around that
ja ungefär
S7: let's see here (11s) two hundred and fourteen Ryd end station leaves forty six (.) thirteen 'n' forty six then you will be down fourteen oh seven (.)
då ska vi se här (11s) tvåhundrafjorton Ryd ändhållplatsen går förtisex (.) tretton å förtisex då är du nere fjorton noll sju (.)
U8: aha
jaha
S9: 'n' (.) the next one takes you there (.) fourteen thirty seven (.) but that is too late
å (.) nästa är du nere (.) fjorton å trettisju (.) men de ä ju för sent

Figure 3: Dialogue fragment from a real interaction on bus time-table information

U2: Well, hi (.) I am going to Ugglegatan eighth
ja hej (.) ja ska till Ugglegatan åtta
S3: Yes
ja
U4: and (.) I wonder (.) it is somewhere in Tannefors
och (.) jag undrar (.) det ligger nånstans i Tannefors
S5: Yes (.) I will see here one one I will look exactly where it is one moment please
ja (.) jag ska se här ett ett jag ska titta exakt var det ligger ett ögonblick bara
U6: Oh Yeah
jarå
S7: (operator disconnects) (25s) mm (.) okey (5s) what the hell (2s)
(operator connects again) hello yes
((Telefonisten kopplar ur sig)) (25s) ähh (.) okey (5s) de va som faan (2s)
((Telefonisten kopplar in sig igen)) hallå ja
U8: Yes hello
ja hej
S9: It is bus two hundred ten which runs on old tannefors road that you have to take and get off at the bus stop at that bus stop named vetegatan
det ä buss tvåhundratio som går gamla tanneforsvägen som du får åka å gå av vid den hållplatsen vid den hållplatsen som heter vetegatan.

Figure 4: Dialogue fragment from a natural bus timetable interaction

search procedures as discussed above.

- *The system does not ask for information it has already achieved.* For instance asking again if it is on Sunday as in S9 in figure 1. This is not uncommon in human interaction and such utterances from the user are not removed. However, we can assume that the dialogue system does not forget what has been talked about before.

4.2 Modifying user utterances

The general rule is to change user utterances as little as possible. The reason for this is that we do not want to develop systems where the user needs to restrict his/her behaviour to the capabilities of the dialogue system. However, there are certain changes made to user utterances, in most cases as a consequence of changes of system utterances.

- *Utterances that are no longer valid are removed.* The most common cases are utterances whose request has already been answered, as seen in the distilled dialogue in figure 2 of the dialogue in figure 1.

S11: sixteen fifty five
sexton femtifem
U12: sixteen fifty five (.) aha
sexton femtifem (.) jaha
S13: bus line four hundred thirty five
linje fyrahundra trettifem

Figure 5: Dialogue fragment from a natural bus timetable interaction

- *Utterances are removed where the user discusses things that are in the environment.* For instance commenting the 'systems' clothes or hair. This also includes other types of communicative signals such as laughter based on things outside the interaction, for instance, in the environment of the interlocuters.
- *User utterances can also be added in order to make the dialogue continue.* In the dialogue in figure 5 there is nothing in the dialogue explaining why the system utters S13. In such cases we need to add a user utterance, e.g. *Which bus is that?*. However, it might turn out that there are cues, such as intonation, found when listening to the tapes. If such detailed analyses are carried out, we will, of course, not need to add utterances. Furthermore, it is sometimes the case that the telephone operator deliberately splits the information into chunks that can be comprehended by the user, which then must be considered in the distillation.

5 Applying the method

To illustrate the method we will in this section try to characterise the results from our distillations. The illustration is based on 39 distilled dialogues from the previously mentioned corpus collected with a telephone operator having information on local bus time-tables and persons calling the information service.

The distillation took about three hours for all 39 dialogues, i.e. it is reasonably fast. The distilled dialogues are on the average 27% shorter. However, this varies between the dialogues, at most 73% was removed but there were also seven dialogues that were not changed at all.

At the most 34 utterances where removed from one single dialogue and that was from a dialogue with discussions on where to find a parking lot, i.e. discussions outside the capabilities of the application. There was one more dialogue where more than 30 utterances were removed and that dialogue is a typical example of dialogues where distillation actually is very useful and also indicates what is normally removed from the dialogues. This particular dialogue begins with the user asking for the telephone number to 'the Lost property office' for a specific bus operator. However, the operator starts a discussion on what bus the traveller traveled on before providing the requested telephone number. The reason for this discussion is probably that the operator knows that different bus companies are utilised and would like to make sure that the user really understands his/her request. The interaction that follows can, thus, in that respect be relevant, but for our purpose of developing systems based on an overall goal of providing information, not to understand human interaction, our dialogue system will not able to handle such phenomenon (Jönsson, 1996).

The dialogues can roughly be divided into five different categories based on the users task. The discussion in twenty five dialogues were on bus times between various places, often one departure and one arrival but five dialogues involved more places. In five dialogues the discussion was one price and various types of discounts. Five users wanted to know the telephone number to 'the Lost property office', two discussed only bus stops and two discussed how they could utilise their season ticket to travel outside the trafficking area of the bus company. It is interesting to note that there is no correspondence between the task being performed during the interaction and the amount of changes made to the dialogue. Thus, if we can assume that the amount of distillation indicates something about a user's interaction style, other factors than the task are important when characterising user behaviour.

Looking at what is altered we find that the most important distilling principle is that the 'system' provides all relevant information at once, c.f. figures 1 and 2. This in turn removes utterances provided by both 'system' and user.

Most added utterances, both from the user and the 'system', provide explicit requests for information that is later provided in the dialogue, e.g. utterance S3 in figure 6. We have added ten utterances in all 39 dialogues, five 'system' utterances and five user utterances. Note, however, that we utilised the transcribed dialogues, without information on intonation. We would probably not have needed to add this many utterances if we had utilised the tapes. Our reason for not using information on intonation is that we do not assume that our system's speech recogniser can recognise intonation.

Finally, as discussed above, we did not utilise the full potential of multi-modality when distilling the dialogues. For instance, some dialogues could be further distilled if we had assumed that the system had presented a time-table. One reason for this is that we wanted to capture as many interesting aspects intact as possible. The advantage is, thus, that we have a better corpus for understanding human-

U2: Yees hi Anna Nilsson is my name and I would like to take the bus from Ryd center to Resecentrum in Linköping
jaa hej Anna Nilsson heter jag och jag vill åka buss från Ryds centrum till resecentrum i Linköping.

S3: mm **When do you want to leave?**
mm **När vill du åka?**

U4: 'n' I must be at Resecentrum before fourteen and thirty five (.) 'cause we will going to the interstate buses
å ja ska va på rececentrum innan fjorton å trettifem (.) fö vi ska till långfärdsbussarna

Figure 6: Distilled dialogue fragment with added utterance

computer interaction and can from that corpus do a second distillation where we focus more on multimodal interaction.

6 Discussion

We have been presenting a method for distilling human dialogues to make them resemble human computer interaction, in order to utilise such dialogues as a knowledge source when developing dialogue systems. Our own main purpose has been to use them for developing multimodal systems, however, as discussed above, we have in this paper rather assumed a speech-only system. But we believe that the basic approach can be used also for multi-modal systems and other kinds of natural language dialogue systems.

It is important to be aware of the limitations of the method, and how 'realistic' the produced result will be, compared to a dialogue with the final system. Since we are changing the dialogue moves, by for instance providing all required information in one move, or never asking to be reminded of what the user has previously requested, it is obvious that what follows after the changed sequence would probably be affected one way or another. A consequence of this is that the resulting dialogue is less accurate as a model of the entire dialogue. It is therefore not an ideal candidate for trying out the systems over-all performance during system development. But for the smaller sub-segments or sub-dialogues, we believe that it creates a good approximation of what will take place once the system is up and running. Furthermore, we believe distilled dialogues in some respects to be more realistic than Wizard of Oz-dialogues collected with a wizard acting as a computer.

Another issue, that has been discussed previously in the description of the method, is that the distilling is made based on a particular view of what a dialogue with a computer will look like. While not necessarily being a detailed and specific model, it is at least an instance of a class of computer dialogue models. One example of this is whether the system is meant to acquire information on the user's underlying motivations or goals or not. In the examples presented, we have not assumed such capabilities, but this assumption is not an absolute necessity. We believe, however, that the distilling process should be based on one such model, not the least to ensure a consistent treatment of similar recurring phenomena at different places in the corpora.

The validity of the results based on analysing distilled dialogues depends partly on how the distillation has been carried out. Even when using natural dialogues we can have situations where the interaction is somewhat mysterious, for instance, if some of the dialogue participants behaves irrational such as not providing feedback or being too elliptical. However, if careful considerations have been made to stay as close to the original dialogues as possible, we believe that distilled dialogues will reflect what a human would consider to be a natural interaction.

Acknowledgments

This work results from a number of projects on development of natural language interfaces supported by The Swedish Transport & Communications Research Board (KFB) and the joint Research Program for Language Technology (HSFR/NUTEK). We are indebted to the participants of the Swedish Dialogue Systems project, especially to Staffan Larsson, Lena Santamarta, and Annika Flycht-Eriksson for interesting discussions on this topic.

References

Lars Ahrenberg, Nils Dahlbäck, Arne Jönsson, and Åke Thurée. 1996. Customizing interaction for natural language interfaces. *Linköping Electronic articles in Computer and Information Science, also in Notes from Workshop on Pragmatics in Dialogue, The XIV:th Scandinavian Conference of Linguistics and the VIII:th Conference of Nordic and General Linguis-*

tics, Göteborg, Sweden, 1993, 1(1), October, 1. http://www.ep.liu.se/ea/cis/1996/001/.

Sue Atkins, Jeremy Clear, and Nicholas Ostler. 1992. Corpus design criteria. *Literary and Linguistic Computing*, 7(1):1–16.

Douglas Biber. 1993. Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4):244–257.

Jean Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

Steve Crowdy. 1993. Spoken corpus design. *Literary and Linguistic Computing*, 8(4):259–265.

Nils Dahlbäck and Arne Jönsson. 1999. Knowledge sources in spoken dialogue systems. In *Proceedings of Eurospeech'99, Budapest, Hungary.*

Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1998. Wizard of oz studies – why and how. In Mark Maybury & Wolfgang Wahlster, editor, *Readings in Intelligent User Interfaces.* Morgan Kaufmann.

Nils Dahlbäck, Annika Flycht-Eriksson, Arne Jönsson, and Pernilla Qvarfordt. 1999. An architecture for multi-modal natural dialogue systems. In *Proceedings of ESCA Tutorial and Research Workshop (ETRW) on Interactive Dialogue in Multi-Modal Systems, Germany.*

Nils Dahlbäck. 1991. *Representations of Discourse, Cognitive and Computational Aspects.* Ph.D. thesis, Linköping University.

Maxine Eskenazi, Alexander Rudnicki, Karin Gregory, Paul Constantinides, Robert Brennan, Christina Bennett, and Jwan Allen. 1999. Data collection and processing in the carnegie mellon communicator. In *Proceedings of Eurospeech'99, Budapest, Hungary.*

Annika Flycht-Eriksson and Arne Jönsson. 1998. A spoken dialogue system utilizing spatial information. In *Proceedings of ICSLP'98, Sydney, Australia.*

Annika Flycht-Eriksson. 1999. A survey of knowledge sources in dialogue systems. In *Proceedings of IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems, August, Stockholm.*

Roger Garside, Geoffrey Leech, and Anthony McEnery. 1997. *Corpus Annotation.* Longman.

Arne Jönsson and Nils Dahlbäck. 1988. Talking to a computer is not like talking to your best friend. In *Proceedings of the First Scandinavian Conference on Artificial Interlligence, Tromsø.*

Arne Jönsson. 1996. Natural language generation without intentions. In *Proceedings of ECAI'96 Workshop Gaps and Bridges: New Directions in Planning and Natural Language Generation*, pages 102–104.

Pernilla Qvarfordt and Arne Jönsson. 1998. Effects of using speech in timetable information systems for www. In *Proceedings of ICSLP'98, Sydney, Australia.*

Pernilla Qvarfordt. 1998. Usability of multimodal timetables: Effects of different levels of domain knowledge on usability. Master's thesis, Linköping University.

Karen Sparck Jones and Julia R. Galliers. 1996. *Evaluating Natural Language Processing Systems.* Springer Verlag.

Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella. 1998. Paradise: A framework for evaluating spoken dialogue agents. In Mark Maybury & Wolfgang Wahlster, editor, *Readings in Intelligent User Interfaces.* Morgan Kaufmann.