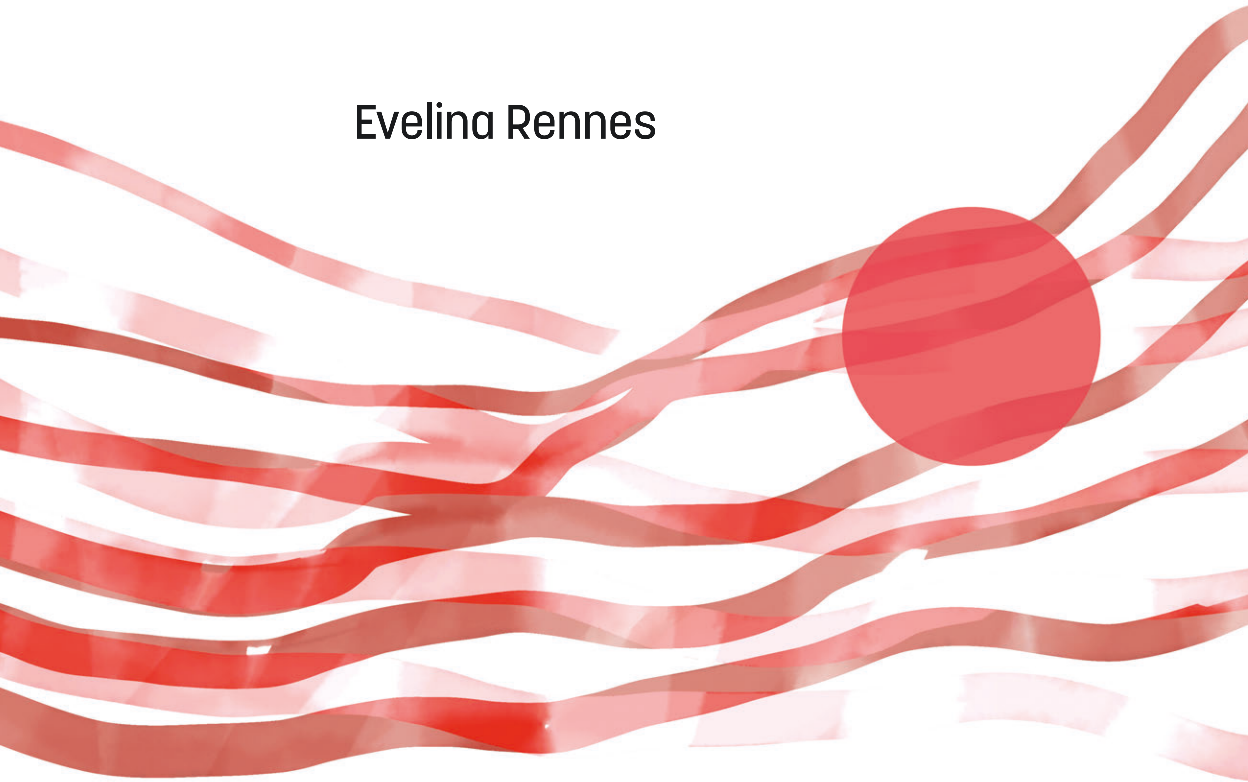
Linköping Studies in Science and Technology
Dissertation No. 2219

Automatic Adaptation of Swedish Text for Increased Inclusion

Evelina Rennes



LINKÖPING UNIVERSITY

Linköping Studies in Science and Technology
Dissertations, No. 2219

Automatic Adaptation of Swedish Text for Increased Inclusion

Evelina Rennes

Linköping University
Department of Computer and Information Science
Division of Human-Centered Systems
SE-581 83 Linköping, Sweden

Linköping 2022

Typeset using XeLaTeX

Printed by LiU-Tryck, Linköping 2022

Edition 1:1

ISBN 978–91–7929–269–0 (print)
ISBN 978–91–7929–270–6 (PDF)
https://doi.org/10.3384/9789179292706
ISSN 0345-7524

ABSTRACT

Imagine what life would be if words did not make any sense to you. If you had to use a dictionary, or Google Translate, to check the meaning of each word. If you forgot the first part of the sentence when you got to the last word. Working would be hard, studying perhaps even more difficult. Reading is a skill essential for participation in many parts of the modern society, but there are many individuals incapable of assimilating text. This could be for various reasons, for instance due to dyslexia, cognitive disabilities, or as a consequence of having a different first language. In order to automate the adaptation process, more knowledge about the properties of Easy Language texts is needed. There are several collections of guidelines aimed at writers of Easy Language texts. Such guidelines are useful for getting an idea of what Easy Language should be, but they are often rather vague and rely on the expert knowledge and intuition of the writer. By analysing texts produced by writers of Easy Language, it is possible to operationalise the guidelines and obtain data for building models for automatic text adaptation. However, this approach requires large corpora of high quality, and the results are highly dependent on the training data. This doctoral thesis adopts an end-user approach to automatic text adaptation. The application of a reading comprehension perspective on Easy Language text adaptation differs from the text-based perspectives commonly used within the field and this position implies that the strengths and weaknesses of individuals in the different groups of poor readers are important to consider. We present and evaluate a variety of techniques for automatic text adaptation, as well as work on text complexity visualisation. Assessment of text complexity can be useful for both readers and writers of Easy Language texts. However, common complexity features are not easily interpretable, and this thesis presents work on enhancing interpretability of such features, using clustering and visualisation methods. The main contributions of this thesis are 1) a mapping of Easy Language guidelines to a theoretical model of reading comprehension, 2) a corpus of simple and standard documents, aligned at the sentence level, 3) clustering and visualisation of text complexity measures, and 4) a number of tools and services for the automatic adaptation of Swedish text.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Fundera över hur ditt liv skulle se ut om du inte kunde tillgodogöra dig skriven text. Om du behövde slå upp näst intill varje ord i en mening, eller om du hade glömt vad meningen du läste handlade om när du kom till det sista ordet. Det är verkligheten för många människor som—av någon anledning—har svårt att ta till sig skriven text. Det kan exempelvis handla om personer med dyslexi, afasi, intellektuell funktionsnedsättning och personer som har svenska som andraspråk. Dessa olika typer av läsare har behov av texter skrivna på lätt svenska, men att anpassa text manuellt är en resurskrävande aktivitet, både vad gäller tid och pengar. Om denna process kunde, helt eller delvis, automatiseras skulle det göra att fler människor kan tillgodogöra sig skriven text, vilket skulle vara fördelaktigt för individen, men också för samhället i stort.

I dag används riktlinjer för att skriva text på lätt svenska. I Sverige har bland annat Myndigheten för Tillgängliga Medier (MTM) utvecklat riktlinjer för att skriva lättläst. I dessa riktlinjer ges råd gällande lingvistiska konstruktioner så som ordval eller grammatik, men även andra faktorer så som vilken typ av bilder som bör användas ihop med en text samt råd om radavstånd och val av typsnitt. Eftersom texter skrivna på lätt svenska i regel riktar sig mot en bred publik och inte är specifikt målgruppsanpassade, så kan dessa anpassningar te sig ganska generella. Detta har naturligtvis fördelarna att det finns ett tydligt schema för hur text bör anpassas, vilket underlättar för nya skribenter att lära sig lättlästskrivande, men nackdelen är att det inte nödvändigtvis blir bra för alla läsare. Vi vet att personer med dyslexi har andra läsutmaningar än, exempelvis, personer med svenska som andraspråk, och de kan behöva olika typer av textanpassningar.

Språkteknologi är ett forskningsområde som syftar till att på olika sätt bearbeta, tolka och analysera mänskligt språk med hjälp av teknik. Inom detta område ryms flera tekniker som skulle kunna användas för att anpassa text till lätt svenska på automatisk väg. Ett exempel är automatisk textförenkling som syftar till att förenkla texten lexikalt, syntaktiskt eller semantiskt, utan att förlora för mycket av textens ursprungliga betydelse. Automatisk textsammanfattning är ett annat exempel, där målet är att på automatisk väg sammanfatta det viktigaste innehållet i en text. En gemensam nämnare hos samtliga tekniker är att vi behöver korpusar (samlingar av text) med texter skrivna på lätt svenska. Utifrån sådana korpusar kan vi få viktig information om hur lätt text ska skrivas eller för att mäta komplexiteten hos en text, och storskaliga korpusar kan användas för att träna upp språkmodeller för automatisk textanpassning

Denna avhandling tar avstamp i det mänskliga målgruppsperspektivet och beskriver olika egenskaper och utmaningar hos de olika målgrupperna för anpassad text. Vidare kartläggs riktlinjer för att manuellt anpassa text, och dessa riktlinjer kopplas till en teoretisk modell av läsförståelse samt värderas utifrån dess applicerbarhet. Sedan följer ett kapitel om textkomplexitet. Tidigare forskning om textkomplexitet har resulterat i ett antal textkomplexitetsmått som kan användas för att få fram numeriska värden för att beskriva en text. Problemet med dessa är att de inte är speciellt enkla att förstå. Vi har försökt att göra dessa mått lite mer förståeliga genom att gruppera måtten och vi presenterar också förslag för hur dessa mått ska visualiseras. Därefter beskrivs arbetet med att samla in korpusar för forskning om textanpassning till lätt svenska, samt hur dessa parallellställdes och analyserades. Avhandlingen beskriver sedan arbetet med att utveckla och utvärdera olika tekniker för automatisk textanpassning, vilket innefattar tekniker för automatiskt synonymutbyte, syntaktisk förenkling samt automatisk textsammanfattning. Slutligen samlas samtliga tekniker i en uppsättning verktyg som kan användas för automatisk textanpassning.

Avhandlingens huvudsakliga bidrag är 1) en kartläggning av riktlinjer för lätt svenska till en teoretisk modell för läsförståelse, 2) en korpus av texter skrivna på lätt svenska, parallellställda på både dokument- och meningsnivå, 3) arbetet med att göra textkomplexitet mer förståeligt genom att gruppera dem och visualisera dem och 4) ett antal verktyg och tjänster för automatisk textanpassning för att skriva och läsa lätt svenska.

Acknowledgments

They say that doing a PhD is a journey. I always thought it was somewhat of a cliche, but approaching the finish line, I understand completely. It has really been a journey for me, both professionally and personally.

First and foremost, I would like to thank my supervisor Arne Jönsson. You have provided me with much-needed inspiration and motivation and you have helped me to believe in myself. Where I see problems, you see solutions and opportunities. I have always felt your support, knowing that you were only two doors away (or lately, only a Teams call away), always taking time for me when needed.

Thank you, Marina Santini, my secondary supervisor, for your tireless flow of ideas and inspiration. You always take time for me, answering my e-mails without any delay, and bring new perspectives to every discussion.

Thank you Magnus Merkel, for being my secondary supervisor at first, but also for giving me the chance to become part of the Fodina family, providing me with a well-needed time-out from work at university.

I have truly enjoyed getting to know so many intelligent and kind people at the Division for Human-Centered Systems at the Department of Computer and Information Science at Linköping University. Thank you, and especially to all the current and former members of NLPLAB, for supporting me and giving me valuable feedback throughout the years. Lars Ahrenberg, Jody Foo, Jalal Maleki, Julius Monsen, Marco Kuhlmann, Robin Kurtz, Michael Jonasson, and many, many others. An extra thanks to Daniel Holmer for patiently answering all my questions about the current system implementation and status, as well as for being a great discussion partner during my final PhD year. And obviously, thanks to the rest of the members of the knights of the round fika table, for taking the Swedish fika to a whole new level.

Thank you, project leader Åsa Elwér, always inspiringly structured with the flow charts and detailed project plans, for sharing your great knowledge and for the discussions and laughs.

Thank you, Anne Moe. For all the help and support with the administration, and for being a great support during the final months of my PhD. But also, for being a friend and someone to talk to about everything and anything.

Thank you, Henrik Danielsson, for the helpful and insightful comments on the final thesis draft.

Thank you, Gustav Häger, for always being up for a coffee and providing me with motivational(?) llama GIFs every time I needed cheering up.

It has been a journey professionally, but it wouldn't have been possible if I didn't have my life outside of university, where I can recharge my batteries. Thank you, all my friends at Stenholmen Islandshästcenter, for always being there for me.

Thank you, my best friend Ylva. You have been my vent, sharing both frustration and laughs, and you make me believe in myself and remind me of my value and competence. You are the best.

Thank you, mum and dad, for the constant encouragement and for always letting me go my own way and make my own mistakes. And thanks mum, for the awesome thesis cover design.

I wouldn't be me if I didn't acknowledge my four-legged companions, those who keep me healthy, happy, and sane (enough). My lovely Labradors Calla and Vilde (always by my feet while writing this thesis) and my horse Gyllir. Thanks for making sure I get my daily dose of laughter.

And last but not least, Daniel. You are and have been everything to me during these years. You are my rock.

Evelina Rennes
Linköping, March 23, 2022

Contents

Papers and Author's Contributions

This chapter provides an overview of the papers included in this thesis on a per-chapter basis, and a description of the authors' contributions (according to the CRediT taxonomy[1]).

Chapters 2, 4, 5, 6 are based on a literature review where the author was the only contributor. The literature review was first conducted in connection with work on the author's master's thesis, and has been extended with further material throughout the work on this doctoral thesis.

Chapters 3 and 7 are based on a submitted article:

Paper I Evelina Rennes and Åsa Elwér (2022). "Evidence-based Text Adaptations for Poor Readers." Manuscript submitted for publication

Contributions

- Evelina Rennes: *conceptualisation, resources, investigation, writing – original draft, writing – review and editing*
- Åsa Elwér: *conceptualisation, methodology, investigation, writing – original draft, writing – review and editing*

The initial idea was based on an earlier, unpublished article where the first author was the only contributor, and the new literature search was conducted by the second author. The article was written and revised by the first author and second author in collaboration.

Chapter 8 is mainly built on two published papers and one unpublished series of design workshops.

[1] https://casrai.org/credit/

Paper II Simon Jönsson, Evelina Rennes, Johan Falkenjack, and Arne Jönsson (2018). “A component based approach to measuring text complexity.” In: *Proceedings of The Seventh Swedish Language Technology Conference 2018 (SLTC-18)*

Contributions

- Simon Jönsson: *methodology, formal analysis, writing – original draft, writing – review and editing*
- Evelina Rennes: *conceptualisation, supervision, visualisation, writing – original draft, writing – review and editing*
- Johan Falkenjack: *conceptualisation, supervision*
- Arne Jönsson: *conceptualisation, supervision*

The second author proposed the visualisation format and participated in discussions about the method and study design, as well as contributing to the process of writing, revising and proof-reading the article.

Paper III Marina Santini, Arne Jönsson, and Evelina Rennes (2020). “Visualizing Facets of Text Complexity across Registers.” In: *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI@LREC). Marseille, France*, pp. 49–56

Contributions

- Marina Santini: *investigation, formal analysis, writing – original draft, writing – review and editing*
- Arne Jönsson: *investigation, formal analysis, writing – original draft, writing – review and editing*
- Evelina Rennes: *visualisation, writing – original draft, writing – review and editing*

The third author proposed the visualisation format and participated in discussions about the method and study design, as well as contributing to the process of writing, revising and proof-reading the article.

Regarding the series of design workshops, the author contributed in a supervisory role and was involved in discussions about the workshop planning and results. The design suggestions and workshop arrangement were conducted by Simon Cavedoni, Emil Fritz and Jakob Säll.

Chapter 9 is based on the following papers:

Paper IV Sarah Albertsson, Evelina Rennes, and Arne Jönsson (2016). “Similarity-Based Alignment of Monolingual Corpora for Text Simplification.” In: *Coling 2016 Workshop on Computational Linguistics for Linguistic Complexity (CL4LC), Osaka, Japan*

Contributions

- Sarah Albertsson: *conceptualisation, software, investigation, formal analysis, writing – original draft, writing – review and editing*
- Evelina Rennes: *supervision, software, investigation, formal analysis, writing – original draft, writing – review and editing*
- Arne Jönsson: *supervision, writing – review and editing*

The work on implementing the algorithm was conducted jointly by the first and second author, and the evaluation setup and article planning, writing, and proof-reading were shared between the authors.

Paper V Evelina Rennes and Arne Jönsson (2016). “Towards a Corpus of Easy to Read Authority Web Texts.” In: *Proceedings of the Sixth Swedish Language Technology Conference (SLTC-16), Umeå, Sweden*

Contributions

- Evelina Rennes: *conceptualisation, software, data curation, writing – original draft, writing – review and editing*
- Arne Jönsson: *supervision, writing – review and editing*

The first author formulated the idea behind the paper and was the main contributor regarding the execution of corpus collection and text processing. The paper was written by the author and revised together with the co-author.

Paper VI Evelina Rennes (2018). “An Aligned Resource of Swedish Complex-Simple Sentence Pairs.” In: *Proceedings of the Seventh Swedish Language Technology Conference (SLTC2018), Stockholm, Sweden*

Contributions

- Evelina Rennes: *conceptualisation, software, resources, investigation, formal analysis, writing – original draft, writing – review and editing*

The author was the only contributor.

Paper VII Evelina Rennes (2020). "Is it simpler? An Evaluation of an Aligned Corpus of Standard-Simple Sentences." In: *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI@LREC), Marseille, France.* Pp. 6–13

Contributions

- Evelina Rennes: *conceptualisation, software, resources, investigation, formal analysis, writing – original draft, writing – review and editing*

The author was the only contributor.

Paper VIII Daniel Holmer and Evelina Rennes (2022). "NyLLex: A Novel Resource of Swedish Words Annotated with Reading Proficiency Level." Manuscript submitted for publication

Contributions

- Daniel Holmer: *software, data curation, formal analysis, writing – original draft, writing – review and editing*
- Evelina Rennes: *conceptualisation, writing – original draft, writing – review and editing, supervision*

The second author formulated the idea behind the paper. The procedure was planned by the authors in collaboration but conducted by the first author. Article planning, writing, and proofreading were shared between the authors.

Chapter 10 is based on the following papers:

Paper IX Evelina Rennes and Arne Jönsson (2015). "A Tool for Automatic Simplification of Swedish Texts." In: *Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa-2015), Vilnius, Lithuania*

Contributions

- Evelina Rennes: *conceptualisation, software, writing – original draft, writing – review and editing*
- Arne Jönsson: *supervision, writing – review and editing*

The first author was the main contributor regarding the idea behind the article and development of the tool. The paper was written and revised by the first and second author in collaboration.

Paper X Vida Johansson and Evelina Rennes (2016). "Automatic Extraction of Synonyms from an Easy-to-Read Corpus." In: *Proceedings of the Sixth Swedish Language Technology Conference (SLTC-16), Umeå, Sweden*

Contributions

- Vida Johansson: *investigation, formal analysis, software, writing – original draft, writing – review and editing*
- Evelina Rennes: *conceptualisation, supervision, writing – original draft, writing – review and editing*

The second author contributed by formulating the main idea and study design. Regarding the algorithm implementation, the author contributed in a supervisory role and was involved in discussions about the results. The paper was written and revised by the first author and second author in collaboration. The main work on implementing the algorithms was performed by the first author and the rest of the student group of the course *Tillämpad kognitionsvetenskap*, for which the second author was supervisor: Linnea Fornander, Marc Friberg, Viktor Lind-Håård, and Pontus Ohlsson. However, they were not interested in contributing to writing a research paper. Contributions: *Data curation, formal analysis, software.*

Paper XI Evelina Rennes and Arne Jönsson (2021). "Synonym Replacement based on a Study of Basic-level Nouns in Swedish Texts of Different Complexity." In: *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pp. 259–267

Contributions

- Evelina Rennes: *conceptualisation, software, investigation, formal analysis, writing – original draft, writing – review and editing*
- Arne Jönsson: *supervision, writing – review and editing*

The first author formulated the main idea behind the paper and was responsible for implementation of the synonym replacement algorithm. The paper was written by the first author and revised together with the second author.

Chapter 11 is based on the following papers:

Paper XII Evelina Rennes and Arne Jönsson (2014). "The impact of cohesion errors in extraction based summaries." In: *9th International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association, pp. 1575–1582

Contributions

- Evelina Rennes: *conceptualisation, investigation, formal analysis, writing – original draft, writing – review and editing*
- Arne Jönsson: *supervision, writing – review and editing*

The first author formulated the main idea behind the paper and was responsible for the evaluation design and execution. The paper was written by the first author and revised together with the second author.

Paper XIII Julius Monsen and Evelina Rennes (2022). "Perceived Text Quality and Readability in Extractive and Abstractive Summaries." Manuscript submitted for publication

Contributions

- Julius Monsen: *investigation, formal analysis, writing – original draft, writing – review and editing*
- Evelina Rennes: *conceptualisation, supervision, writing – original draft, writing – review and editing*

The second author mainly contributed in a supervisory role and was involved in discussions about the planning of the study design and results. The paper was written and revised by the first author and second author in collaboration.

Chapter 12 is based on Paper IX and the following papers:

Paper XIV Daniel Fahlborg and Evelina Rennes (2016). "Introducing SAPIS - an API service for text analysis and simplification." In: *The second national Swe-Clarin workshop: Research collaborations for the digital age, Umeå, Sweden*

- Daniel Fahlborg: *software, writing – original draft*
- Evelina Rennes: *software, writing – original draft, writing – review and editing, supervision*

Paper XV Johan Falkenjack, Evelina Rennes, Daniel Fahlborg, Vida Johansson, and Arne Jönsson (2017). "Services for text simplification and analysis." In: *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pp. 309–313

- Johan Falkenjack: *resources, writing – original draft*
- Evelina Rennes: *software, writing – original draft, writing – review and editing, supervision*

- Daniel Fahlborg: *software*
- Vida Johansson: *software*
- Arne Jönsson: *supervision*

The different services described have different contributors. The main contributions of the second author concern the description of the simplification tool and the API. The paper was written by the second author and revised and proof-read by the second author in collaboration with the co-authors.

The current tools and services developed in our research projects are presented in this chapter, and parts of the text included can be found in the publications listed. However, as the tools have been developed further after the publications, the chapter has been supplemented with updated information and descriptions of modules added after the publication of the articles.

Regarding the development of the tools and services, there are many colleagues and students that have contributed to the different parts. The author developed the original version of StilLett. Later revisions of StilLett were developed by Vida Johansson, Daniel Fahlborg, Hampus Arvå, Erik Kindberg, Maja Lindberg, Michael Jonasson, and Daniel Holmer. TeCST and FriendlyReader have been developed in collaboration with Hampus Arvå, David Holmstedt, Michael Jonasson and Daniel Holmer. The first version of SAPIS was developed by the author in collaboration with Daniel Fahlborg. Later versions have been developed by Hampus Arvå, Michael Jonasson, Wiktor Strandqvist and Daniel Holmer. JuliuSum was developed by Julius Monsen and integrated in the tools by Julius Monsen and Daniel Holmer.

CHAPTER 1

Introduction

“Everything should be made as simple as possible, but no simpler.”

The quote above is often attributed to the theoretical physicist Albert Einstein. When entering the quote into Google, the search engine returns many hits for the phrase *”What did Einstein really mean by...”*, and similarly phrased questions, and the different explanations, interpretations and speculations are numerous. As is the case with many catchy quotations attributed to celebrities, it seems likely Einstein never actually uttered those words. Instead, the statement that claims to be the original version of the simplified (and erroneous) quote at the start of this thesis reads:

> *”It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.”* (Einstein, 1934)

Although it can be argued that the two statements carry more or less the same semantic meaning, the first quote has been adapted to a more appealing and engaging version, but in the process, it seems to have lost some of its meaning.

The loss of information is one of the many parameters that must be taken into account when transforming text to a version that is supposedly easier to read. Since words and other linguistic elements carry meaning, albeit some more than others, tweaking a piece of text will inevitably result in some shift in meaning or information loss. The challenge is to find the sweet spot where the text is simple enough to be easy to decode and comprehend, while preserving the original content.

But if meaning risks being lost or altered, what is the purpose of reformulating text at all? Why should we strive toward simple text?

1.1 Motivation

Think of all the text you encounter every day of your life: all the e-mails you read, the notes you take during each meeting, the news you read on newspaper websites. Books, signs, information leaflets. Think of all the texts you have encountered going through the years of elementary school, then secondary school, perhaps university, and all the text you have to go through when getting an insurance policy, making bank transfers, filling out applications, or simply communicating with friends and family on social media. In this Information Age, a huge amount of textual data has been made publicly available on the Internet. Today, it is possible to carry around basically all knowledge from all of history in your pocket!

Now, imagine what life would be if words did not make any sense to you. If you had to use a dictionary, or Google Translate, to look up the meaning of each word. If you forgot the first part of a sentence by the time you got to the last word. Working would be hard, studying perhaps even more difficult. Reading this doctoral thesis would be almost impossible. Now, since you *are* reading this thesis, I assume that you are one of the lucky ones who do not struggle with reading, or if you do, you have found a way of overcoming this issue. Good for you! But the number of people who struggle with reading is large, which is obviously a significant problem since reading is essential for participation in many parts of modern society. Reading allows us to communicate, study and work, and engage in democratic processes, in unions and politics. And reading offers even more than that. Reading is a pleasure that enriches life.

As a matter of fact, literacy has been recognised in numerous declarations of human rights. The UN Convention on the Rights of Persons with Disabilities, Article 21 states:

> *States Parties shall take all appropriate measures to ensure that persons with disabilities can exercise the right to freedom of expression and opinion, including the freedom to seek, receive and impart information and ideas on an equal basis with others and through all forms of communication of their choice, as defined in article 2 of the present Convention, including by:*
>
> (a) *Providing information intended for the general public to persons with disabilities in accessible formats and technologies appropriate to different kinds of disabilities in a timely manner and without additional cost; . . .*

There are some keywords and phrases that we should note here: *disabilities*, *accessible formats*, *technologies*, and *timely manner and without additional cost*. All these keywords are relevant to this thesis.

The first keyword is *disabilities*. An international survey conducted within The Programme for the International Assessment of Adult Competencies (PIAAC) (2012; 2013) maps the proficiency of adults in literacy, numeracy, and problem solving. Adults in Sweden generally scored high in all skills (top five in literacy and numeracy and number one in problem solving), but despite this, relatively large proportions of the adult population exhibit lower scores. Regarding literacy, 58% of the Swedish adult population scored 3 or higher (on a scale of 6 literacy proficiency levels), which was the threshold for being a good reader. Thus, 42% of the Swedish adult population are below this level. About 13% of the population are at a literacy proficiency level of 1 or lower, which indicates a very low level of literacy.

When talking about reading disabilities, dyslexia is naturally the first disability that comes to mind. This makes sense, since dyslexia is a learning disability that affects the skill of reading, writing, and spelling. Since there are natural variations in the complex skills of reading and writing, it is somewhat difficult to estimate the number of people suffering from dyslexia. The NHS estimates that 1 in every 10 to 20 people have dyslexia (NHS, n.d.), and according to the Swedish Dyslexia Association (*Dyslexiföreningen*), around 5–8 percent of the population suffer from difficulties with writing and reading (Dyslexiföreningen, n.d.).

But there are other audiences that, for various reasons, struggle with reading. Such audiences include, for instance, individuals with aphasia, people with intellectual disabilities, the hearing-impaired or the deaf, and those learning Swedish as a second language.

Aphasia denotes a language impairment caused by some acquired brain damage. In Sweden, about 12,000 individuals are diagnosed with some kind of aphasia every year, and 35% of the diagnosed patients are in working age (*Vad är Afasi?* N.d.). About 1% of the Swedish population have an intellectual disability of some sort (*Utvecklings-störning* n.d.), and 18.5% of the Swedish adult population have a hearing impairment (*Hörselskadade i siffror 2017. Statistik om hörselskadade och hörapparatutprovningar i Sverige från Hörselskadades Riksförbund (HRF)* 2017). About 0.1% of the Swedish population are congenitally deaf (or deaf since very early in childhood) (*Hur många döva finns det i Sverige?* N.d.). The range of impairment is obviously very large, and while many individuals with hearing impairment might not require adapted text, this is nevertheless a group of readers that should not be overlooked. Second-language learners of Swedish denote persons who do not speak Swedish as their first language, and thus might struggle with reading. The number of Swedish inhabitants born abroad amounted to over 2 million people on December 31, 2020 (*Inrikes och utrikes födda efter region, ålder och kön. År 2000–2020* 2020). It comes as no surprise that the level of literacy is lower in this population than the national average, but according to the PIAAC report (The Programme for the International Assessment of Adult Competencies (PIAAC), 2012), the difference in literacy skills between foreign-language immigrants and native-born persons was the largest observed difference of all the countries participating in the survey.

Aside from the target audiences mentioned above, there are other groups of people that could benefit from simple text, such as children, the elderly or individuals with autism spectrum disorder. To complicate things further, an individual could obviously have several impairments, with different or even conflicting needs. Taken together, the number of people in need of accessible text is significant.

The second keyword from the UN convention Article 21 is *accessible format*. This could be interpreted in various ways. From a linguistic point of view, it often means that the text conforms to a certain language standard, often defined using terms such as easy-to-read or plain language. Such guidelines are meant to ensure that the information is presented to the reader in an accessible way. The guidelines give linguistic advice, but also guidance about fonts, line length, etc.

The third keyword taken from the UN convention is *technologies*. Although expert writers of accessible material are preferable to technological solutions as they are likely to produce a high-quality end-product, there are advantages to using technology to automate the process. For example, technological solutions could allow for easier, on-demand access to information, and let the reader adapt the text to their own specific needs and preferences. Such solutions could also be a way of overcoming the issues connected to the manual creation of such material: shortage of time and money (fourth keyword: *timely manner and without additional cost*!).

Thus, using technology to assist the process of either writing simple text, or, as an end user, reading and comprehending text is one way of ensuring that information is provided to persons with disabilities in accordance with the UN declaration Article 21. The question is then, how should this be done? That is exactly this that this thesis is concerned with. The ability to read—whether with eyes, ears or hands—is clearly an important matter for inclusion in society, but it also fulfils another purpose. Reading is a pleasure of life that should be made available to as many people as possible. And that is my personal motivation for conducting the research that has resulted in this thesis.

1.2 Simplification or Adaptation

Automatic simplification of text is an established subfield of natural language processing commonly referred to as *Automatic Text Simplification* (ATS). It covers adaptations that transform a text either syntactically, lexically, or semantically so that the complexity[1] is reduced while the important content is preserved. In this doctoral thesis, I have, to a large extent, focused on automatic text simplification. But I have also tried to look beyond the notion of simplification, to include other methods and techniques that do not fit within this label but could be helpful for creating accessible texts. Thus, I have chosen to use the term *Automatic Text Adaptation* (ATA), as a wider term that includes all possible modifications and transformations. In this thesis, aside from automatic text simplification, methods used mainly include automatic text summarisation, but in theory other techniques could be covered by this term. Such techniques include other natural language processing methods, as well as purely design-related adaptations such as text design (line spacing, line length etc.).

1.3 Research Questions

This thesis was built on the following research questions:

RQ1 What linguistic adaptations are needed in an automatic adaptation system for simple Swedish?

RQ2 How can automatic text adaptation be implemented and conceived in order to meet the needs of different target audiences?

The first research question focuses on specific adaptations used by professional writers of simple text, which have been codified into existing guidelines. This question aims to address how texts should be adapted according to guidelines for Easy Language, and which of these guidelines are applicable in an automatic adaptation system.

The second research question relates to the actual application of automatic adaptation operations, and how those operations should be realised to address the needs and challenges of specific reader audiences. In order to answer this question, it is necessary to 1) map the different reader audiences according to objective and subjective reading difficulties, and 2) investigate the extent to which the text adaptation operations can be implemented in a software system.

1.4 Contributions

The main contributions of this thesis include discussions of various techniques for automatic text adaptation and how these techniques might be applicable for different target audiences. More specifically, the main research contributions of this thesis can be narrowed down into a few items:

[1] The notion of *text complexity* is not straightforward. Is this a feature of the text or a function of the text and the reader? In ATS papers, both definitions seem to be used. This issue is further explored in Chapter 6.

- We map Easy Language guidelines to a theoretical model of reading comprehension, with a goal of obtaining information about if and how Easy Language text adaptations are useful for different groups of poor readers.
- We provide a palette of tools and services for analysis and adaptation of texts to target readers.
- We present work on creating parallel corpora that can be used for automatic text adaptation purposes.
- We present work on how to make sense of text complexity features, by clustering and visualising them.

1.5 Delimitations

The thesis is delimited concerning the following areas:

- The thesis does not try to find all possible adaptation and simplification techniques. For example, it does not include machine translation approaches to automatic text simplification, or all possible ways of simplifying a text lexically. There is a limited number of adaptation guidelines, but there are countless techniques that could be applied, and this thesis does not claim to find the perfect combination of adaptation techniques. On the contrary, one claim in this thesis is that such perfect combination is not possible.
- This thesis does not cover the design of the suggested tools and services. Such design and usability issues have been addressed by other contributions made in our research projects (e.g., Bergström Kousta (2016) and Ihs Håkansson (2018)), but design of tools and services is not within the scope of this thesis.

1.6 Outline

This doctoral thesis is a monograph divided into four parts. Each part is divided into a series of chapters, and each chapter covers a specific area of study.

Part I (chapters 2–6) provides the theoretical background relevant to the thesis.

Chapter 2 introduces the various reader audiences often described in the literature as target audiences for Easy Language text. The reading-related issues and challenges faced by individuals from the various groups are described in this chapter. Chapter 3 presents Easy Language initiatives as collections of guidelines, and explores the effect Easy Language texts has on poor readers.

Chapter 4 describes earlier work on automatic text adaptation. It provides an introduction to the automatic text simplification, and presents text simplification for various target audiences, and text simplification efforts in Swedish. Finally, the chapter introduces the research field of automatic text summarisation. Chapter 5 explains the role of corpora in automatic text adaptation research, and provides an overview of available resources. In Chapter 6, the notions of text complexity and readability are explained, and both traditional metrics and more modern data-driven complexity measures are described.

Part II (chapters 7–9) comprises the thesis contributions of exploring texts written in Easy Language. Chapter 7 discusses how Easy Language guidelines can and should be implemented in a system for automatic text adaptation, on the basis of a theoretical model of reading comprehension. Chapter 8 presents our work on how to make sense of text complexity measures. This issue was approached from two directions: is it possible to cluster text complexity features in an intuitive and comprehensible way, and what is the best way to visualise such features? Chapter 9 presents our work on constructing resources useful for research on automatic text adaptation. The main resources described are an aligned corpus of complex-simple sentence pairs, and a lexical resource of words extracted from Easy Language books, annotated with reading proficiency level.

Part III (chapters 10–12) describes the various text adaptation techniques explored in our research projects. Chapter 10 presents our work on syntactic and lexical simplification. First, the syntactic simplification module is presented, and the work on constructing simplification rules is described. Then, our work on lexical simplification is described: finding synonyms based on a corpus analysis of Easy Language texts and by exploiting a lexical resource to find near-synonymous words at a higher and presumably more basic taxonomic level. Chapter 11 presents work on evaluation of automatic summarisation techniques as a means for achieving Easy Language texts. Two types of automatic summarisation techniques are evaluated with respect to cohesion and text quality. Chapter 12 presents the various tools and services developed for reading and writing Easy Language text, and describes the various sub-modules involved in the process.

Part IV (chapters 13–16) sums up the thesis work, discusses the strengths and weaknesses of different parts and orients the thesis work in a wider context. Chapter 13 summarises the whole thesis at a per-chapter level. Chapter 14 discusses the chosen research approach, and discusses an alternative method. This chapter also describes some implications related to ethics and sustainability. In Chapter 15, we return to the initial research questions of this thesis and the results and insights from the previous chapters are discussed. Chapter 16 concludes the thesis work.

Part I

Background

CHAPTER 2

Target Audiences

Adapting texts to use an easier and more concise language, with the purpose of making them available to a broader audience, is an often-addressed aim of automatic text adaptation. This is a demanding task, since the output must be of consistently high quality, as even minor flaws might lower the perceived quality of the text (Saggion, 2017). Groups that might benefit from adapted text are, for instance, individuals with different kinds of cognitive disabilities, the deaf or hard-of-hearing, second language learners, and children.

The issue of target group adaptation has been raised by numerous authors before, but many of the contemporary approaches have focused on general adaptations rather than targeting specific audiences. This is a problem, since the needs vary between the potential user groups, but also within the groups, imposing a great challenge. For instance, simple text provided by public authorities could be more difficult to read and comprehend for individuals with intellectual disabilities or second language learners of Swedish, than for individuals with dyslexia (Falk and Johansson, 2006).

This chapter presents some of the reader populations that could benefit from adapted text and describes the specific needs of the groups. The idea is not to map the disabilities in detail, but rather to point out the diversity of challenges associated with being a poor reader.

2.1 Dyslexia

Since the term *dyslexia* first appeared, it has been defined in various ways. Many attempts have been made to divide dyslexia into sub-types, but there is no clear consensus of how this division should be defined. Dyslexia was first explained as a disorder related to visual perception, but is now rather recognised as a deficit in the phonological processing (Vellutino, 1987). Some researchers claim that dyslexia should be described as a multi-deficit disorder, involving deficits in various abilities, such as working memory and auditory temporal processing (Fostick and Revah, 2018), which could provide an explanation to the differences that have been described within the reader audience, as well as the various suggested subtypes. Regardless of the exact causes of dyslexia, the variety of conditions covered by the term indicates that dyslexia includes a range of different conditions affecting the ability to read and comprehend text. Some criticism has been raised concerning the tendency to create assistive technologies targeting dyslexia as one single disability, and it has been suggested that such technology should consider the variation of difficulties experienced by individuals with dyslexia (Alsobhi, Khan, and Rahanu, 2014).

Individuals with dyslexia struggle with decoding words. Depressed decoding skills are associated with problems establishing the grapheme-phoneme correspondences that are the basis of decoding (Vellutino, Fletcher, Snowling, and Scanlon, 2004), and hence individuals with dyslexia struggle immensely with reaching automatic decoding. In general, reading is not automatised in the adult population with dyslexia, even though many individuals with dyslexia learn to decode. The lack of automatisation influences reading comprehension. When decoding takes too much effort, it is very difficult to simultaneously perform other mental operations (Høien and Lundberg, 2013), and it has been found that individuals with dyslexia have limited ability to make various types of inferences (Simmons and Singleton, 2000). Individuals with dyslexia are often better at listening comprehension compared to reading comprehension (Elwér, Keenan, Olson, Byrne, and Samuelsson, 2013), as their language comprehension is commonly good. Commonly referred reading challenges include difficulties concerning long words and low-frequent words (Rello, Baeza-Yates, Dempere-Marco, and Saggion, 2013), homophones, words that are orthographically similar, new words, and non-words (Rello, Baeza-Yates, Bott, and Saggion, 2013).

2.2 Aphasia

Aphasia is a language impairment caused by brain damage acquired by for example stroke, trauma to the head, neuro-degenerative diseases, or brain surgery. The term origins from the Greek word *aphatos*, meaning speechless, and describes a condition which negatively affects an individual's ability to read, write, and understand spoken words and produce speech. Since aphasic people often had a normal language function prior to the brain damage, the sudden inability of language use often results in a feeling of isolation in the affected individual (Canning and Tait, 1999).

There are many kinds of aphasia, but they all—in some way—affect the ability to comprehend or formulate language. People with aphasia may, among other symptoms, encounter problems with speaking, understanding speech, writing and/or reading. The type of aphasia that relates to the ability to read is called *alexia*.

There are several classifications of aphasia, and there are divergent views of what division to apply, but it is common to separate non-fluent from fluent aphasia.

- *Non-fluent aphasia* terms a condition where the aphasic individual struggles with words and the speech is slow and halting, with incomplete sentences and missing words. The ability to write is affected in the same way as the ability to speak. Language comprehension can be either impaired or close to intact.
- *Fluent aphasia* describes a condition where the aphasic individual can produce speech that sounds normal (fluent), but uses irrelevant or made-up words, causing the speech to not make sense. The writing of the aphasic person is difficult or not possible to understand as the words do not make any sense. Language comprehension can be either impaired or close to intact.

Although a classification of aphasia is useful, it is noteworthy that not all individuals with the different types of aphasia have the same symptoms. The concept of *fluency* is, in this context, more of a continuum rather than two opposite poles and patients might, for example, have non-fluent speech but produce some very specific perfectly fluent phrases (Hillis, 2007).

Common difficulties experienced by individuals with aphasia include high information density, long sentences, long sequences of adjectives, passive voice and noun compounds (Carroll, Minnen, Pearce, Canning, Devlin, and Tait, 1999). Other difficulties described in the literature are sentences with object relative clauses and comparisons of word meaning (*"is x larger than y?"*) (Hillis, 2007).

2.3 Intellectual Disabilities

Individuals with intellectual disabilities (ID) have a delay in reading as compared to typical readers. This is apparent in their decoding and reading comprehension skills (Nilsson, Danielsson, Elwér, Messer, Henry, and Samuelsson, 2021a; Nilsson, Danielsson, Elwér, Messer, Henry, and Samuelsson, 2021b). ID is characterised by low IQ and limitations in many cognitive abilities, such as working memory and executive functions (Danielsson, Henry, Messer, and Rönnberg, 2012), and language skills such as vocabulary and grammatical understanding (Nilsson, Danielsson, Elwér, Messer, Henry, and Samuelsson, 2021a).

Generally, individuals with ID encounter problems associated with literacy learning, although there are individual variations, and children with ID are capable of enhancing their literacy skills (Ratz and Lenhard, 2013).

Research has shown that intrinsic motivation is a key factor for understanding a text (Guthrie, Wigfield, Humenick, Perencevich, Taboada, and Barbosa, 2006; Wigfield and Guthrie, 1997), that text engagement correlates with enhanced reading comprehension and that the motivation for reading increases with the amount of reading (Wigfield and Guthrie, 1997). For individuals with ID, the issue of motivation is especially challenging, since this group of people are often met with low expectations and prejudices of incompetence. Adapting texts to the abilities, needs, and interests of the reader enables easier access to other cognitive abilities and strategies linked to reading. This, in turn, leads to a greater chance of extracting meaning from a text, at the same time as it has a positive effect on the motivation and engagement of the reader (Morgan and Moni, 2008).

2.4 Deaf and Hard-of-Hearing

People who are deaf or hard-of-hearing may also struggle with reading. Studies have compared the deaf with second language learners since a deaf person's preferred language is generally sign language. As sign language is a purely visual language, deaf individuals must learn to read without the support of phonological aspects of language. It is established that childhood hearing loss deeply affects language development, and the language deficits may also affect the development of literacy negatively (Lederberg, Schick, and Spencer, 2013).

Children who are deaf and hard-of-hearing especially struggle with grammar (Lederberg, Schick, and Spencer, 2013), and syntactically complex sentences (Siddharthan, 2006), as well as a limited vocabulary and difficulties generalising a word's meaning to other contexts (Fabbretti, Volterra, and Pontecorvo, 1998).

2.5 Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that is distinguished by, among other things, impairment in social and communicative behaviours. Individuals with ASD may have an associated language impairment, although this is not required for an ASD diagnosis according to the diagnostic criteria. Thus, with respect to reading comprehension skills, the ASD audience is diverse. Subgroups of children with ASD have been described in the context of their language abilities (Kjelgaard and Tager-Flusberg, 2001), where the language skills range from normal in some children, to not reaching basal skills in others. Some children with ASD can receive relatively high scores on vocabulary tests, but experience difficulties in understanding concepts related to emotions (Hobson and Lee, 1989). Difficulties with understanding figurative language is one of the most prominent problems. According to a meta-analysis of the research on figurative language for individuals with ASD (Kalandadze, Norbury, Nærland, and Næss, 2018), the difficulties seem to be related to basic language skills, and that enhanced general language skills might improve the comprehension of figurative language. The authors highlighted that it is important for individuals with ASD to be exposed to figurative language, and that it is beneficial to provide explanations to such constructions instead of avoiding them.

A meta-analysis of reading comprehension skills of individuals with ASD found that the performance on reading comprehension of individuals with ASD depends on text type (Brown, Oram-Cardy, and Johnson, 2013). Generally, individuals with ASD perform better when reading texts that require little social knowledge. However, the ASD diagnosis covers a variety of symptoms and deficits, and the diagnosis in itself does not imply any reading comprehension difficulties.

2.6 Second Language Learners

One often addressed target audience for automatic text simplification is second language (L2) learners. This audience differs from the other groups, since learners of a new language do not need to have any impairment that hinders reading or comprehension, but rather experience difficulties related to a poor vocabulary, unfamiliarity of specific cultural phenomena, or a lack of knowledge about the grammar of the language that is being learnt. The L2 audience is very diverse in nature. Factors like the reader's knowledge in the particular language being learnt and how similar the second language is to the reader's mother tongue affect the experienced reading difficulties.

Knowing a language's vocabulary has proved to be an important factor for learning a new language (Krashen, 1989). It has been argued that L2 learners, just like children, develop most of their vocabulary and spelling skills by reading. The Input hypothesis (Krashen, 1989) states that a language is best learnt when receiving comprehensible input while reading. According to this theory, it is important to be exposed to text of the language you are trying to learn in order to acquire knowledge about the vocabulary, but it also stresses the importance of the comprehensibility of a text. The input should be slightly more difficult than the current competence of the reader to allow for progress in language development.

2.7 Children

Children could benefit from adapted text, although not having any physical or cognitive disability. There is a developmental aspect of children's reading that should not be disregarded. The text should not be too simple, since reading encourages learning of new words, and the reading level should thus be adapted to the reading level of the certain reader (De Belder and Moens, 2010).

2.8 Chapter Summary

This chapter has provided a brief introduction to the various reader audiences often mentioned in research on automatic text adaptation. The intention of the chapter was not to describe the various difficulties associated with each audience in detail, but rather to highlight the wide range of challenges that each group faces. However, even though it is possible to say something general about the skills of each target audience, it is important to remember that each reader is an individual, and that the challenges and difficulties might vary even within the groups. There are also other aspects to consider, such as the developmental aspects of reading. Providing texts that are too simple might inhibit further enhancement in reading ability, and lower the motivation of the reader.

CHAPTER 3

Easy Language and Text Adaptation

This chapter describes the characteristics of *Easy Language*, and examines various initiatives that provide guidelines for writing Easy Language texts. The chapter also reviews the effects that adapted texts may have on the reader.

3.1 Easy Language

Reading is a very complex task that involves a combination of cognitive processes (Rayner, 1998). Poor readers come in many forms, from those with clinical issues, such as individuals with an intellectual disability or aphasia, to second language learners who struggle to comprehend texts in a language in which they are not fluent. Poor readers from these different target groups have a range of cognitive and language difficulties that selectively impair different aspects of reading comprehension.

The Swedish Language Act (SFS 2009:600) was adopted in 2009 and Section 11 states that *"The language of the public sector is to be cultivated, simple and comprehensible"*. The Swedish Language Council (*Swedish: Språkrådet*) regularly evaluates the use of plain language in Swedish public agencies through a survey distributed to all such agencies and interviews with writers responsible for plain language at the public agencies. The 2020 version of this evaluation revealed that time shortage is the most common obstacle to providing plain language texts (Hansson, 2020).

A number of initiatives have attempted to adapt texts and make them more comprehensible so that they are accessible to poor readers. Examples of such initiatives in Sweden include the recommendations issued by Myndigheten för Tillgängliga Medier (MTM, 2021) and Begriplig text (Begriplig Text, 2019). Internationally, the most influential set of guidelines for Easy Language is Plain Language (PLAIN, 2011). The guidelines commonly have a one-size-fits-all approach, which suggests that the resulting text will improve comprehension of readers regardless of the cognitive and language factors that underlie an individual's reading difficulties.

Wengelin (2015) conducted a review of syntactic and lexical guidelines, in order to evaluate the validity of the recommendations offered in the guidelines. She found that many of the guidelines were based on outdated literature and theories, and that in many cases the studies were based on research conducted on the English language. Moreover, some guidelines are based on research that focused on how linguistic units are processed, rather than the actual text *comprehension*. Wengelin (2015) pointed out that isolated units are important for processing text, but that they work in the context of a combination of syntax, semantics, and properties of the reader, and it is important to consider the combination of variables. The author concludes that the most valid guidelines, based on research, recommend using words that are frequent and familiar for a given reader and avoiding subordinate phrases.

3.2 Easy Language Initiatives

Provision of Easy Language text is part of the solution to increased inclusion in society. Historically, texts have been manually adapted to meet the demands of groups in need of simplified texts. To ensure uniform adaptations, and to facilitate the process, guidelines for simple writing have been developed.

There are many terms describing simple language in the literature: Easy Language, Easy-to-Read, Easy Reading, Easy to Understand, Simplified language, Clear Writing, and Plain Language. These terms are often confused or used interchangeably. No matter the term used, all these notions describe texts that are clear and comprehensible, and include advice about content, language, layout, images, and other graphical elements.

In this thesis, the term *Easy Language* is used to describe text written to be easier to read for readers with different kinds of impairing conditions. The Swedish Language Act points out a distinction between two slightly different perspectives: *Easy Language* and *Comprehensible Language* (Wengelin, 2015). The former, which could be described as a text-based perspective, suggests that there are some generic traits that make a text easier to read for all, and the latter includes the reader perspective. This is consistent with the two definitions of easy-to-read provided by the IFLA guidelines (2010), where the first describes text adapted to be easier to read, but not necessarily to comprehend, whereas the second includes both reading and comprehension. Although this distinction is useful, especially when describing the characteristics of a text, this thesis uses *Easy Language* as an umbrella term covering both notions.

A number of different guidelines have been developed to guide Easy Language production. Below, we present the most influential efforts internationally and in a Swedish context. The focus is on guidelines for Swedish and English, since the number of guidelines available for Swedish Easy Language is rather limited. Some guidelines are language-independent, using common words and transparent organisation of the text, whereas other guidelines are language-dependent and only applicable to a specific language's structure.

Plain Language is one of the most commonly mentioned ways of adapting text to make it more accessible. It is defined as follows: *"A communication is in plain language if its wording, structure, and design are so clear that the intended audience can easily find what they need, understand what they find, and use that information"* (International Plain Language Federation, n.d.). Plain Language guidelines (PLAIN, 2011) propose various suggestions for simple writing. For example, it is important to avoid passive voice, abstract words and superfluous words. Plain Language guidelines overlap with Easy Language guidelines in many ways. However, the guidelines do not claim to be *easy-to-read* in the same sense as Easy Language texts targeting readers with different types of impairments, but rather clear and comprehensible for all readers. Nevertheless, the common denominator of Easy Language and Plain Language is that they both intend to provide clear and accessible texts by taking different linguistic and stylistic factors into account.

Inclusion Europe is an organisation that speaks up for the rights of people with intellectual disabilities and their families. Among other initiatives, they provide guidelines on how to write Easy Language texts for people with intellectual disabilities (Inclusion Europe, 2020). The Inclusion Europe guidelines include, among other things, advice about word selection, when to write numbers out, and ordering the content in an easy, intuitive and chronological way.

The International Federation of Library Association and Institutions (IFLA) published guidelines for writing Easy Language texts (Misako Nomura and Tronbacke, 2010). They identified two main audiences for such texts: 1) individuals with some kind of disability affecting reading skills who have a permanent need for Easy Language texts, and 2) readers with a limited reading proficiency, but whose limitation is not permanent, and will likely improve in skill. The guidelines are intended to cover a wide range of reader audiences, and consist of general advice, such as *write concretely* and *be concise*.

Easy Language and Plain Language in Sweden

The project *Begriplig Text* was an initiative by the Swedish membership organisations for persons with aphasia, autism, and dyslexia (*Afasiförbundet*, *Autism- och Aspergerförbundet*, *Dyslexiförbundet*) and the Swedish National

Association for People with Intellectual Disability (*Riksförbundet FUB*). The project set out to describe the features that make texts easy to read and comprehend. In the project, a total of 19 guidelines were compiled, directed towards journalists, writers and other people who produce Easy Language (Begriplig Text, 2019). The guidelines were collected through an extensive web survey and cover both the design of documents, as well as offering specific advice regarding the comprehensibility of the textual content. Aside from the 19 guidelines, the project also highlighted the importance of reflecting the intended reader audience and what the writer wants to communicate with the text. The first impression is also important: if a text looks difficult, it might be avoided, even though the content might be comprehensible.

LL-förlaget is a part of the Swedish Agency for Accessible Media (MTM), and publishes, produces and distributes accessible texts such as literature and newspapers. *LL-förlaget* and MTM provide guidelines for writing in Easy Language, and separate the guidelines into advice concerning content, language, and form and graphical content. The more general guidelines about the content highlight the importance of keeping the reader in mind, keeping a common thread throughout the text, keeping a straightforward chronology in the text, and avoiding presenting too many different persons in the text.

Myndigheternas skrivregler (Språkrådet, 2014) is a guide from The Institute for Language and Folklore, a Swedish authority that, among other missions, provides the government and other authorities with information and recommendations. The guide targets any writer for the Swedish public authorities or any government-controlled business, and contains advice on how to write *Klarspråk*, or Plain Language. The guide highlights general advice about, for example, writing with a specific audience in mind, to carefully consider the specific aim and goal with the text, and to use a clear text outline. More specific and straightforward guidelines cover word choices, capitalisation, abbreviations, etc.

Table 3.1: Overview of the referred Easy Language guidelines.

Guidelines	Type	Language
MTM/LL-förlaget	Easy-to-Read	Swedish
Begriplig Text	Easy-to-Read	Swedish
Myndigheternas skrivregler	Plain Language	Swedish
IFLA	Easy-to-Read	English
Inclusion Europe	Easy-to-Read	English
Plain Language Handbook	Plain Language	English

The guidelines we refer to in this thesis are presented in Table 3.1.

3.3 Effects of Easy Language for Poor Readers

The goal of Easy Language is to improve reading comprehension for different types of poor readers. The most valid measurement of such effects are comparisons of reading comprehension for authentic texts and adapted texts in the target groups. However, other outcome measures are also relevant as they may be sub-targets in improvement of reading comprehension or relate to creating a positive reading experience. Therefore, the studies reviewed below include outcome measures like reading speed, ability to search the texts, number of reading errors, fixation time, perceived reading comprehension and actual reading comprehension. We start by reviewing studies which have examined Easy Language texts in target groups, and continue with studies on specific adaptations to match deficits of the target groups.

Easy Language Texts for Poor Readers

One approach to study whether Easy Language adaptations work in target groups is to examine if individuals from the targeted audience understand text that is modified using Easy Language principles. This was done in a study by Fajardo, Avila, Ferrer, Tavares, Gomez, and Hernandez (2014). Twenty-eight students with Intellectual Disabilities (ID) read three easy-to-read news texts per week over the course of sixteen weeks. They answered literal and inference questions (multiple choice) after reading each text. The students showed high levels of corrects answers (87% for literal questions and 76% for inferential questions) which indicates that the text was not too difficult for them, and that the Easy Language adaptations were useful. However, using another text type considered complex in terms of content (Falk and Johansson, 2006), one minor study in Swedish explored how Easy Language texts provided by public authorities were experienced by individuals with ID. Questionnaires were used and the results indicated that the Easy Language texts were still difficult to read for persons with ID since the information was too abstract in nature.

A different approach to examining efficiency in Easy Language guidelines in poor readers is to compare comprehension of authentic texts and adapted texts. This was done in a study by Karreman, Van Der Geest, and Buursink (2007). A website was adapted with Easy Language guidelines and the effect on individuals with ID was evaluated. The adapted website was validated by an expert in web communication and a specialist in care for people with ID. The websites were evaluated by letting the participants perform a number of tasks and answer questions, measuring searching and reading time and comprehension. Regarding the searching and reading time, there were no difference between performance on the authentic and adapted versions, but the guidelines were proven effective for improving comprehension for individuals with ID.

In a study by Yaneva, Temnikova, and Mitkov (2015) a group of English participants with Autism Spectrum Disorder (ASD) was compared to a control group on various text adaptations including pictures and simple language. The 9 texts were collected from websites and books. The main finding on text processing was that the group with ASD rated the nine separate Easy Language texts very differently, ranging from ”very difficult” to ”very easy”. In the control group the texts were generally rated as ”very easy”. The authors argue that personal interests may play a bigger role in how a text is perceived in the group with ASD.

There is evidence for an enhanced comprehensibility for simplified texts read by second language (L2) readers. It has been found that L2 students were more successful on comprehension-focused multiple-choice questions after reading a simplified version of a text, than those who read the original version (Long and Ross, 1993). In one study, forty-eight Japanese college students that were L2 speakers of English read thirteen texts and answered comprehension questions (Yano, Long, and Ross, 1994). The set of texts consisted of three versions of a text: one original version, one simplified version using Easy Language guidelines and one enhanced version. In the enhanced version, the texts were enhanced based on principles of how natives talk to foreigners. Such texts often have clear explicit connections, extra words to improve comprehension and are therefore commonly longer than the original. The study showed that the level of understanding was significantly better when using the Easy Language versions compared to the original, however the students performed at similar levels when reading the enhanced texts. These results suggest that a different approach to text adaptations may be more relevant to L2 learners.

Text structure is another factor that can affect reading comprehension. Text that is organised in logical patterns of cause and effect, compare and contrast, and problem and solution seem to be easier for L2 readers than text that is more loosely organised (Carrell, 1984).

Reader-specific Adaptations

Individuals with ID experience severely depressed reading comprehension skills resulting from compromised construction and integration of text ideas and world knowledge (Nilsson, Danielsson, Elwér, Messer, Henry, and Samuelsson, 2021a; Danielsson, Henry, Messer, and Rönnberg, 2012). As this group has widespread difficulties using broad spectrum adaptations may be a good idea. The few studies that have studied Easy Language adaptations on groups with ID suggest that Easy Language is helpful in this group of readers (Karreman, Van Der Geest, and Buursink, 2007; Fajardo, Avila, Ferrer, Tavares, Gomez, and Hernandez, 2014). However, based on this research we cannot determine if some aspects of adaptations are more important than others, and also whether different adaptations, such as elaborations would be even more helpful. This is a group where assessments of how texts are experienced is especially difficult (Huenerfauth, Feng, and Elhadad, 2009) as their meta cognitive ability is very limited (Soto and Poblete, 2018).

In the group with dyslexia, empirically evaluated text adaptations are much more specific. Text adaptations have focused on lexical simplifications. Some studies have found effects on lexical adaptations in groups with dyslexia, however they mainly report effects on other reading-related skills such as reduction in reading errors, increased reading speed and positive assessments of having simplifications on demand (Rello, Baeza-Yates, Dempere-Marco, and Saggion, 2013; Rello, Baeza-Yates, Bott, and Saggion, 2013; Gala and Ziegler, 2016). Maybe these results indicate that more extensive adaptations, which involve higher levels of processing, are necessary to improve reading comprehension in dyslexic readers. But again, the empirical evidence is very limited.

Two other groups with reading difficulties are deaf/hard of hearing and L2 learners. Despite reported grammatical and lexical deficits in deaf/hard of hearing (Siddharthan, 2006; Fabbretti, Volterra, and Pontecorvo, 1998), adaptations to reduce grammatical and lexical complexity have not been found to improve comprehension by itself in the two studies presented (Alonzo, Seita, Glasser, and Huenerfauth, 2020; Vettori and Mich, 2011). For second language learners, the study by Yano, Long, and Ross (1994) indicated improved comprehension when using Easy Language guidelines, however elaborated versions were comprehended equally well. For L2 learners it could be argued that elaborated texts are more positive for their language growth compared to Easy Language texts. Possibly this type of adaptations could be useful for other poor reader groups as well, particularly in an educational setting.

3.4 Chapter Summary

Audience of Easy Language texts comes in many forms, and the one-size-fits-all approach to Easy Language might not be the best in order to enhance comprehension for a specific reader audience. Arguments come from the fact that poor readers differ within and across groups, and likely require different types of adaptations related to the specific difficulties they experience. It has also been suggested that the guidelines in general lack empirical evidence (Wengelin, 2015), and the literature examining effectiveness of text adaptations in groups of poor readers is sparse. Although the adapted texts appear to be useful for readers groups in some studies, the underlying reasons are not well understood. Researchers in the field deal with a number of complicated factors when attempting to gauge the effectiveness of adapted text. Easy Language texts appear to be helpful to some readers with extensive cognitive and language deficits, like individuals with intellectual disability, but they are not necessarily more comprehensible for all readers. For instance, for individuals with dyslexia, studies show that lexical adaptations may be helpful to improve decoding performance, but these adaptations are not sufficient to improve reading comprehension. For the deaf or hard of hearing, no effects of Easy Language adaptations were found. To sum up, we know very little about if and how text adaptations are useful for poor readers. Although the texts appear to be useful for readers groups in some studies, the underlying reasons are not understood.

CHAPTER 4

Automatic Text Adaptation

This chapter includes the relevant theoretical background of automatic text adaptation research. It provides an introduction to the automatic text simplification process and describes the efforts made on text simplification for various target audiences. In this chapter we also describe the research conducted on text simplification in Swedish and introduce the research field of automatic text summarisation.

4.1 Automatic Text Simplification

Automatic text simplification is often referred to as the process of reducing the syntactic, lexical, and/or semantic complexity of a text, while preserving the core meaning.

Historically, automatic text simplification has had two different aims. On the one hand, it has been used as a pre-processing step prior to various other natural language processing tasks, such as machine translation or text summarisation, under the assumption that a simpler syntactic structure would improve text processing quality due to reduced degree of ambiguity (Chandrasekar, Doran, and Srinivas, 1996). On the other hand, automatic text simplification has been developed as an aid for individuals that benefit from simpler texts. For a long time, texts have been manually simplified to suit the needs of various reader audiences, but it is a time-consuming and expensive task, which drives efforts to make this process automatic. Currently, technical and computational development is rapidly increasing, which enables even faster and more advanced natural language processing techniques.

Automatic text simplification is usually divided into two main tasks: syntactic simplification and lexical simplification (Saggion, 2017).

Reducing the complexity by making changes to the syntactic structure of a text is called *syntactic simplification*. Most commonly, syntactic simplification operations involve breaking up long sentences into several shorter sentences, changing the voice of a sentence from passive to active voice, or realigning the word order.

The first approaches to syntactic simplification aimed to simplify input data of other natural language processing applications (Chandrasekar, Doran, and Srinivas, 1996), and did this by making use of hand-crafted simplification rules. Soon, the field instead turned to the automatic deduction of such rules from annotated corpora (Chandrasekar and Srinivas, 1997). Some approaches identified simplification rules by the manual analysis of corpora of Easy Language texts. Brouwers, Bernhard, Ligozat, and François (2014) aligned and explored parallel corpora of French texts to identify simplification operations. Three types of rules were implemented in a rule-based simplification tool: deletion rules, modification rules and splitting rules. Decker (2003) performed a similar analysis for Swedish, which resulted in a set of simplification rules that were implemented in the simplification tool CogFlux (Rybing, Smith, and Silvervarg, 2010; Abrahamsson, 2011).

Lexical simplification denotes the process of finding and replacing difficult words or phrases with simpler alternatives, without altering the meaning of the sentence. It can be divided into subtasks. Typically, the lexical simplification pipeline is described as in Figure 4.1. Research on lexical simplification does not generally cover all parts of the pipeline, but most research covers at least some parts.

The first step, *Complex Word Identification*, aims to identify words in need of simplification, that is, words that are considered to be complex by a given target audience. The definition of word complexity can differ, but one way to identify appropriate words is to use any kind of lexical complexity measures, such as word frequencies, or to make use of lexical resources of base vocabulary. The second step, *Substitution Generation*, describes the task of generating possible synonyms for the input word. The third step, *Substitution Selection*, is the task of selecting the most appropriate synonym candidates among the generated suggestions, for instance with regards to keeping the input word's sense. The final step, *Substitution Ranking*, ranks the synonym candidates according to simplicity.

Several approaches to lexical simplification have been proposed. The most common approach is to use frequency measures when choosing between substitution candidates, since a common word is assumed to be easier to comprehend. However, as highlighted by Alfter (2021), more frequent words could also be more polysemous, and thus, more complex. The complexity of a word could also be defined by other factors, such as how familiar it is to the

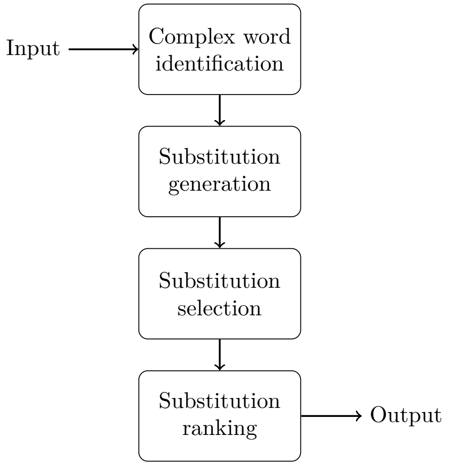


Figure 4.1: The lexical simplification pipeline. Figure adapted from Shardlow (2014) and Paetzold and Specia (2017).

reader (Anderson and Freebody, 1979; Gernsbacher, 1984), which means that the reader's background knowledge is a factor when assessing word difficulty.

Lexical simplification has proven successful for enhancing comprehension for different reader audiences, such as individuals with dyslexia (Rello, Baeza-Yates, Dempere, and Saggion, 2013) and second language learners (Gardner and Hansen, 2007). But it is far from a simple task. The conventional definition of a *synonym* is a word that has the same or nearly the same meaning as another word. However, in reality, there are very few words that are synonymous in this strict sense, and the substitution candidates should, if possible, preserve the original word's connotations, semantic meaning, as well as the grammatical form. For instance, synonyms to the word *woman* could be *girl*, *female*, *lass*, *gal* or *lady*, and although it could be argued that the various words define a somewhat similar concept (a female human being), they carry differing meaning and connotations. The slight shift in meaning across various synonymous (or near-synonymous) words can imply that the words are only synonymous within specific contexts. However, for the impaired reader, also words with a slight shift in connotation or meaning could also be useful for lexical simplification, such as near-synonyms or other semantically similar words. One such example is basic-level words and hyponymy, concepts which are further explored in Section 10.2.

Data-driven Text Simplification

The vast majority of text simplification research has been devoted to simplification at the sentence level (Alva-Manchego, Scarton, and Specia, 2020). Limiting the task in this way makes it possible to reuse methods from related natural language processing tasks. For instance, simplification is often regarded as a task that is analogous to monolingual machine translation (Specia, 2010; Coster and Kauchak, 2011c; Coster and Kauchak, 2011a; Wubben, Bosch, and Krahmer, 2012; Xu, Napoles, Pavlick, Chen, and Callison-Burch, 2016; Nisioi, Štajner, Ponzetto, and Dinu, 2017; Zhang and Lapata, 2017; Zhang, Ye, Feng, Zhao, and Yan, 2017). Alva-Manchego, Scarton, and Specia (2020) pointed out that methods relying on statistical machine translation do indeed perform lexical simplification and reordering and deletion of parts of sentences, but that such methods do not in themselves handle splitting of sentences in a good way. Rule-based approaches can handle split operations, but come with other problems: which rules should be applied and in what order? Alva-Manchego, Scarton, and Specia (2020) considered sequence labelling methods (using machine learning to predict simplification operations) to be a promising way forward for data-driven simplification, but the lack of properly annotated corpora limits the exploration of such methods.

Text Simplification for Different Reader Audiences

Automatic text simplification research is often motivated by highlighting the beneficial effects that text simplification could have on individuals belonging to the different target audiences, but the actual needs of the target audiences are not always reflected in the resources or evaluation methods that are used. The various reader audiences are not homogeneous, and to complicate the issue further, there are large individual differences within the different reader groups. However, some approaches have taken on an audience-centred perspective in their efforts to develop techniques to simplify text, and this section will describe some efforts with a specific focus on the reader audience.

Simplification methods targeted at individuals with dyslexia have primarily been focused on making individual words easier to decode. The substitution of difficult words with simpler synonyms is a method often suggested for text simplification for this group of readers. However, this does not necessarily mean that simply replacing difficult words with synonyms is a good solution. In a study, Rello, Baeza-Yates, Bott, and Saggion (2013) investigated whether simplified texts were helpful for individuals with dyslexia by measuring reading time and fixation duration with an eye-tracking device, and by using text comprehension questionnaires to measure comprehension (Rello, Baeza-Yates, Bott, and Saggion, 2013). They found that a system that presented synonyms on demand was preferred over a system that automatically replaced difficult words with simpler synonyms, but that the simplification strategy did not have any detectable effect on text comprehension.

In another study, eye-tracking was used to investigate how word frequency and word length affects reading for individuals with dyslexia. Reading comprehension was tested using comprehension tests (Rello, Baeza-Yates, Dempere-Marco, and Saggion, 2013). They found that using frequent words improved readability for individuals with dyslexia and that shorter words made the text more understandable. Similar patterns of results were found in a study by Gala and Ziegler (2016). They examined the effects of lexical complexity in 10 students with dyslexia. The students were between 8 and 12 years of age. In the reading material lexical substitution was performed according to the following substitution principles: rare words to more frequent words, irregular words to regular words and longer words to shorter ones. In addition, adjectives and adverbs were removed if they were assessed as unnecessary for comprehension. The lexically adapted versions improved reading speed and reduced the number of reading errors, and comprehension was at the same level as with the original texts. These results suggest that the simplification made the reading a more pleasant experience, without a loss in comprehension.

Another eye tracking study explored the readability of number representation for individuals with dyslexia (Rello, Bautista, Baeza-Yates, Gervás, Hervás, and Saggion, 2013). The authors concluded that expressing numbers with digits instead of words, and using percentages instead of fractions improved the readability of a text for individuals with dyslexia

Simplext (Saggion, Štajner, Bott, Mille, Rello, and Drndarevic, 2015) is a text simplification system for Spanish. The target audience was people with cognitive disabilities, but the system was also extended to also cover more general text simplification issues. In order to identify simplification operations, a corpus of 200 news texts in Spanish and their expert-made simplified counterparts were compiled (Bott and Saggion, 2011b). The corpus was analysed, and a module-based system was constructed containing three modules: a rule-based lexical simplification module, a synonym-based simplification module, and a module with syntactic simplification grammar. The lexical simplification module handled operations such as transformation of numerical expressions, normalisation of reporting verbs, and elimination of parenthetical expressions. The synonym-based simplification module, LexSiS (Bott, Rello, Drndarević, and Saggion, 2012), used a thesaurus to look up synonym candidates and used a vector space model to distinguish between different senses of the candidate words. The candidates were then ranked using a measure that combined word frequency and word length, and the word with the highest score was chosen for substitution. The syntactic simplification module was built on a manually created grammars that modified the syntactic dependency trees, and included operations such as sentence splitting of subordinate and coordinate structures.

Simplext was evaluated using readability metrics and a web survey that included participants that did not belong to any target audience, measuring grammaticality, simplicity and meaning preservation. The results of the readability metrics evaluation revealed that the simplified texts scored significantly lower than the standard texts on all readability metrics used, but also that the simplified versions generally did not reach the simplicity level of expert-made simplifications. The results of the web survey revealed that the automatically simplified sentences were perceived as simpler, whereas the non-simplified sentences were perceived as more grammatical. A majority of participants considered the meaning to have been preserved in the simplification process.

In 1998, the project PSET (*Practical Simplification of English Text*) published some work on simplifying newspaper text for people suffering from aphasia (Carroll, Minnen, Canning, Devlin, and Tait, 1998). The system was divided into a syntactic and a lexical simplifier. The syntactic simplification component SYSTAR (Canning and Tait, 1999) was rule-based and aimed to handle passive-to-active rewriting, splitting of sentences into shorter ones, and extracting embedded clauses. The lexical simplifier was based on WordNet and used the Kuččera-Francis frequencies (Kucera and Francis, 1967) to choose between synonym candidates. The tools and resources of the PSET project were further refined and used in the project HAPPI (*Helping Aphasiac People Process Information*) (Devlin and Unthank, 2006), which aimed to improve comprehension of web texts by the use of automatic text simplification techniques.

The Able to Include project[1] (Saggion, Marimon, and Ferrés, 2015) developed text simplification for Spanish and English, targeting individuals with intellectual or developmental disabilities. The text simplification process was based on the techniques and resources used in the Simplext project.

The effect of lexical simplification on adult readers who were deaf or hard-of-hearing has been explored in a study by Alonzo, Seita, Glasser, and Huenerfauth (2020). The study revealed no enhancement on the performance of text comprehension questions, but the perceived comprehension was enhanced when the participants were presented with the lexically simplified texts. The study also found that it is preferable that the simplified words be presented on request. In another study, eighteen deaf Italian students read stories that had been lexically simplified so that uncommon words were replaced with simpler synonyms or paraphrased (Vettori and Mich, 2011). At the syntactic level, the number of subordinate clauses was also reduced. These simplifications alone did not improve the participants' comprehension of the stories, however simplified versions with pictures improved comprehension for the deaf students.

[1] http://able-to-include.com

Daelemans, Höthker, and Sang (2004) used sentence simplification through summarisation to generate subtitles automatically for people who are hard-of-hearing. They tested two approaches for reducing sentence length: one knowledge-based approach, and one machine-learning approach learning sentence reduction generalisations from a monolingual corpus of transcripts of TV programs, aligned with their corresponding subtitles. When the machine learning approach did not perform as expected (probably, they claim, due to the humble size of the corpus), they tried a rule-based approach where they manually constructed rules based on the same syntactic information as the machine learning approach was given. The combination of the manually constructed rules and the word replacements given by the machine learner proved to be the best solution.

Inui, Fujita, Takahashi, Iida, and Iwakura (2003) focused on text simplification for the deaf, by providing paraphrases to a given sentence. They identified readability assessment as a sub-task to find the difficult parts of a text before generating paraphrases for the difficult parts of the text. To assess readability for this specific target audience, teachers at a school for the deaf were asked to compare the readability of given sentences to the readability of manually created paraphrases by filling in a questionnaire. The intuition was to take advantage of the expert knowledge possessed by the teachers, and the collected readability ratings were used to train a support vector machine classification model, with promising results. For the paraphrasing sub-task, they used a rule-based approach, feeding the paraphrase engine with more than 28,000 transfer rules, which resulted in a system that could perform both lexical and structural paraphrasing.

Another study presented work on text simplification for deaf people by using manually constructed transformation rules made especially for this specific target group on newspaper articles (Chung, Min, Kim, and Park, 2013). Except for a graphical representation, they implemented a rule-based approach that identified structurally complex sentences, and relocated the clauses according to a set of rules tailored for Korean, in order to simplify the syntactic structure. The system was evaluated through two user studies including six users with normal hearing ability. In the first study, the participants were asked to restore the relocated clauses to their original position, and in the second study they compared their system to a similar baseline system with respect to adequacy. The results showed that participants were able to restore the clauses to a high extent, and their system scored a higher mean adequacy value when compared to similar systems.

The FIRST project (Barbu, Martín-Valdivia, and Ureña-López, 2013) focused on reading comprehension difficulties in individuals with Autism Spectrum Disorder (ASD) with an IQ level above 70. Within the project, they performed literature reviews and a number of studies on participants of varying ages, in order to identify difficulties related to reading comprehension. Based on these studies, they developed a multi-lingual tool (*Open Book)* that aimed to identify and delete constructions that hinder reading comprehension for this audience. *Open Book* is a multilingual tool that processes texts in English, Spanish and Bulgarian, and also contains a personalisation service, which updates user-specific features based on the editing operations of the user. In an analysis of texts written for the target audience, a number of different simplification operations were identified, such as synonym replacement, sentence splitting, and the provision of definitions or explanations. Except for the lexical and syntactic simplification operations, *Open Book* also included other components that, for example, assigned images to difficult terms, explained idioms and other figurative expressions, and categorised texts based on topic.

Evans, Orasan, and Dornescu (2014) presented a set of rules for English syntactic simplification for individuals with ASD. The syntactic simplification rules, which focus on splitting compound sentences and deleting or modifying relative clauses, were evaluated with respect to accuracy against two gold standards which were automatically generated and validated by an expert. The accuracy was computed using the Levenshtein similarity between the generated sentence and the most similar gold standard sentence, and the readability of the generated sentence was computed.

Some text simplification efforts have targeted children. De Belder and Moens (2010) presented a text simplification method using both lexical and syntactic simplification. For the lexical simplification, a restricted form of Word Sense Disambiguation was used. They collected candidate words from WordNet, but instead of directly conducting the synonym replacement, this method was combined with a language model that takes both consecutive words and the contextual meaning into account. The words that were found in the intersection of WordNet and the language model were regarded as synonym candidates, and the word with the highest Kucera-Francis frequency (Kucera and Francis, 1967) was chosen for synonym replacement (as long as it had a higher frequency than the original word). For syntactic simplification, they focused on appositions, relative clauses, and splitting sentences. When an operation had caused the creation of new sentences, the operations were reapplied.

ERNESTA (Enhanced Readability through a Novel Event-based Simplification Tool) (Barlacchi and Tonelli, 2013) is a sentence simplification system for Italian targeting elementary school children with poor reading skills. The system worked in two steps. First, a text analysis was performed, where antecedents of pronouns and subjectless verbs were handled by a module for anaphora resolution. Secondly, the syntactic simplification was performed sentence-wise. The syntactic simplification method was event-based, and worked through a number of subtasks. First, they identified factual events ("actions that took place in the story"), discarding any event that was expressed in future and conditional tense, and then the arguments of each action were detected by the use of dependency parser tags, and necessary transformations were made using a set of rules. The simplification strategy of ERNESTA was built on psycholinguistic principles, and made reference to studies showing that individuals with poor text comprehension have difficulties with extracting the main events of a story and resolving anaphora. Moreover, they highlight that reading comprehension is largely dependent on working memory capacity, which suggests the use of operations that reduce the cognitive load of the reader.

Hmida, Billami, François, and Gala (2018) presented a method for lexical simplification for French children with reading difficulties. They presented two lexical simplification strategies that used a lexical resource of French synonyms with assigned reading difficulty rankings (Billami, François, and Gala, 2018), and showed that the use of such a resource improved performance when compared to a strategy based on word embeddings. The evaluation of the proposed synonyms was performed by two experts, who judged the candidates according to the following criteria: The substituting word should be simpler than the original complex word, and it should fit the context of the sentence.

Automatic Text Simplification for Swedish

Some research has focused on automatic text simplification for Swedish. Decker (2003) extracted simplification operations for Swedish. This was done through an analysis of a parallel corpus. The corpus contained news texts targeting second language learners of Swedish, and the news was written in one standard version and one simplified version. The syntactic structure of the parallel texts was compared, and the work resulted in 25 simplification rules. One subset of the rules was implemented in the CogFLUX system (Rybing, Smith, and Silvervarg, 2010), and another subset of the rules was implemented by Abrahamsson (2011), who also developed the system further by adding a synonym replacement module.

Keskisärkkä and Jönsson (2013) compared three different methods for lexical simplification through synonym replacement, and evaluated the strategies using readability measures (LIX, OVIX, average word length, and proportion of long words). The number of erroneous synonym exchanges and the number of exchanges in total were also measured. The strategies were synonym replacement based on word frequencies, word length, and level of synonymy. The results of this study indicate that synonym replacement strategies based on word frequencies and word length can improve readability, as measured by the given readability measures, but the risk of introducing errors is high.

Another study treating lexical simplification through synonym replacement replaced complex medical terms with simpler synonyms (Abrahamsson, Forni, Skeppstedt, and Kvist, 2014). The synonym replacement strategy was based on word frequencies, but they also took the frequencies of substrings of the words into account. Medical terms tend to occur rarely in general corpora, which complicates synonym replacement based on frequency measures of the words in their original form. However, the compound nature of the Swedish language suggests that it could be relevant to take the frequency of substrings into account. A rare compound word which comprises constituents that are used in standard language could be easier to understand, than a less frequent noncompound word. The evaluation included LIX and OVIX measures, and the results revealed that the resulting text was slightly more difficult according to LIX, but more readable according to OVIX.

A more recent attempt to create a complete text simplification tool for Swedish is SWEASY (Priscepov, 2020). The system includes syntactic and lexical simplification and a module for automatic text summarisation. The system was evaluated using the SVIT text complexity features (Heimann Mühlenbock, 2013) as well as a survey assessing grammaticality, meaning preservation and cohesion, with promising results.

4.2 Automatic Text Summarisation

Automatic text summarisation aims to reduce the length of a text while preserving the core content. It has been proposed as one text adaptation strategy which could increase readability.

Automatic text summarisation is typically divided into two types: *extractive* or *abstractive*. *Extractive* summarisation systems work by extracting the most important sentences from a text, and then concatenating them together in order to produce a summary of a text. *Abstractive* summarisation systems, on the other hand, generate new text, with the intention of capturing and paraphrasing the key content of the source text (Hahn and Mani, 2000).

There have been some attempts to compare abstractive and extractive summaries. Carenini and Cheung (2008) compared extractive and abstractive summaries focusing on the controversiality of the opinions they expressed. They found that the margin by which abstraction outperforms extraction is greater when controversiality is high. Souza, Meireles, and Almeida (2021) compared extractive and abstractive summarisation methods for facilitating labelling of subgroups in patent records.

Some efforts have been made on summarisation as simplification. Margarido, Pardo, Antonio, Fuentes, Aires, Aluísio, and Fortes (2008) tested three different extraction-based summarisation strategies on target readers, and found that all strategies improved the understanding of the text to some extent. They concluded that summarisation, in combination with other techniques, could be useful for simplifying texts, but that it is important to take the literacy level of the reader into account. Smith and Jönsson (2011a) showed that text complexity, given by several established text complexity measures, can be reduced by using extractive summarisation techniques. They propose summarisation as a first step to reduce the difficulty of a text, before applying other text adaptation strategies. More recently, hybrid approaches of text simplification and summarisation have been proposed. For example, Zaman, Shardlow, Hassan, Aljohani, and Nawaz (2020) adapted the Pointer generator model, a combination of abstractive and extractive summarisation models, to introduce a simplification factor to the loss function based on lexical complexity, and also used simplified summaries as training data.

4.3 Chapter Summary

This chapter has reviewed the research field of automatic text adaptation, by describing efforts made in automatic text simplification and automatic text adaptation for creating Easy Language texts.

Although the development within most fields of natural language processing is accelerating in conjunction with the rapid developments in computer science and the increase of available computational power, the task of automatic text simplification remains in part unsolved.

Typically, automatic text simplification has been approached at the sentence level. However, text comprehension does not depend solely on the decoding of individual sentences, and sentence-level simplification should be seen only as a partial solution for text simplification. One adaptation strategy that works at the document level is automatic text summarisation. It has been suggested as one text adaptation strategy which could increase readability, as the overall aim is to produce shorter text.

Despite the fact that it is common to validate research on automatic text simplification by focusing on the needs of the target audiences, the described methods and techniques are often rather general. Some approaches have taken on a target-centred perspective in their efforts to develop techniques to simplify text. This has been done by studying reading behaviour and assessing comprehension, using datasets specific to certain target audience, basing the simplification strategies on previous research on known psycholinguistic principles and difficulties among people with certain diagnoses, or using participants from the target audiences for evaluation of simplification.

This chapter also briefly introduced automatic text summarisation as an adaptation technique.

CHAPTER 5

Corpora for Automatic Text Adaptation

This chapter presents the role of corpora in text adaptation research, and describes the main resources available today.

5.1 Why Should We Study Corpora?

In Chapter 3, we noted that guidelines for writing Easy Language texts are usually a mixture of rather clear and concise advice, and more fuzzy text properties that are difficult to operationalise into rules that could be fed into an automatic system. However, text that has been written by professional writers following these guidelines could offer real-world examples of the applied guidelines, and could be used as a tool for learning how such text should be written. Moreover, adapting text to different target readers requires knowledge about the given audience. What are the specific difficulties that the readers experience, and how should the text be written and presented to suit the needs of various audiences? Expert writers have unique knowledge about writing for a particular audience, and that knowledge could be harvested by examining the texts they have written.

Thus, one of the answers to why we should study corpora is:

Answer 1: It provides a valuable resource for expert knowledge on how to write accessible text.

Studying corpora is an efficient way of gaining insight into the characteristics of how a language is used in a naturalistic setting. It allows for large-scale analyses, resulting in quantitative measures on how language is *actually* used. The simplest measures of such analyses are frequency counts. Frequency measures can give information about how commonly used certain linguistic constructions are in comparison with others. This could be useful when searching for a simpler synonym for a difficult word. Thus, a second answer would be:

Answer 2: Corpus studies allows for large-scale exploration and comparison of languages, genres and domains.

Modern natural language processing techniques require electronically available text resources. Thus, a third answer could be:

Answer 3: Large-scale and high-quality text corpora can be used to train models for a wide range of natural language processing purposes.

5.2 Corpora for Automatic Text Adaptation

In order to study how professionals adapt texts, or to use data-driven methods for automatic text adaptation, large-scale parallel or comparable corpora are required. A good corpus for this purpose would be a large collection of Easy Language texts where each text has one or more equivalent texts in standard language. However, there are not many such resources; instead, the creation of aligned *comparable* (rather than completely parallel) monolingual corpora has been proposed for various natural language processing tasks, such as paraphrasing (Barzilay and Elhadad, 2003; Dolan, Quirk, and Brockett, 2004), automatic text summarisation (Knight and Marcu, 2000; Jin, 2002), terminology extraction (Hazem and Morin, 2016), and automatic text simplification (Bott and Saggion, 2011a; Coster and Kauchak, 2011b; Klerke and Søgaard, 2012). Since most of the adaptation work described in this thesis treats simplification, this section mainly describes corpora for automatic text simplification.

Wikipedia has been regarded as a suitable resource for training text simplification systems, especially due to its size and availability. The English Wikipedia has a Simple English counterpart, which makes it easy to align articles on the same topic to create a large parallel or comparable resource. The Parallel Wikipedia Simplification (PWKP) corpus (Zhu, Bernhard, and Gurevych, 2010) is an aligned resource of sentences from Wikipedia and Simple English Wikipedia, that has been used to implement data-driven approaches to text simplification. The PWKP corpus was constructed by pairing around 65,000 articles from English Wikipedia and Simple English Wikipedia, and aligning the articles at the sentence level using a term frequency-inverse document frequency (tf-idf) approach. The corpus was used to train a tree-based translation model. Wikipedia and Simple English Wikipedia have also been aligned in other studies. In one article, a parallel resource from English Wikipedia and Simple English Wikipedia was presented (Coster and Kauchak, 2011c). In that case the tf-idf approach was used for initial paragraph alignment, and a dynamic programming approach was used to find the best alignments. Other approaches have aligned Wikipedia with Simple English Wikipedia by using various similarity cores. Hwang, Hajishirzi, Ostendorf, and Wu (2015) used a sentence-level similarity score that combined a semantic similarity measure (based on Wiktionary) and a measure which includes the dependency structure between the words of a sentence (Hwang, Hajishirzi, Ostendorf, and Wu, 2015). Kajiwara and Komachi (2016) used a method that combined the similarity of word vectors to create a sentence similarity method and aligned a resource for automatic text simplification purposes. *WikiLarge* (Zhang and Lapata, 2017) is an aggregation of aligned sentence pairs generated from Wikipedia data in earlier studies, comprising a total of 296,000 aligned instances.

According to Simple English Wikipedia, the articles target ”everyone”, which includes children and adults who are learning English. The writer's guidelines recommend using the 1,000 most common words in English, and using simple grammar and short sentences, but there are no strict rules. Suggestions given in the article *How to write Simple English Pages* [1] include the rewriting of passive to active voice, using simple verb forms, and using a simple sentence structure.

An article that reviewed Wikipedia as a resource for text simplification argued that it comes with some major problems (Xu, Callison-Burch, and Napoles, 2015). They claimed that half of the sentence pairs in the PWKP corpus are not examples of real simplifications, and that this might insert errors into text simplification systems that train on the dataset. They proposed that new resources should be sought, and that an example of one such resource could be the *Newsela corpus* (further described in Section 5.3). However, it has been shown that simplification models trained on the PWKP dataset have in fact learned simplification operations from the dataset, and it has been argued that the resource should not be completely dismissed for use in training of sentence simplification models (Alva-Manchego, Scarton, and Specia, 2020).

Wikipedia is a very commonly used resource, but many other resources are used for text simplification research, such as the *TurkCorpus* (Xu, Napoles, Pavlick, Chen, and Callison-Burch, 2016), which contains 2,350 sentences from PWKP, and multiple human reference simplifications collected through Amazon Mechanical Turk, or the *Simple PPDB* (Pavlick and Callison-Burch, 2016), a subset (4.5M paraphrases) of the Paraphrase Database (PPDB) which was released for text simplification purposes.

[1] https://simple.wikipedia.org/wiki/Wikipedia:How_to_write_Simple_English_pages

Although most simplification approaches have used simplified corpora to extract simplifications, it can also be fruitful to use general corpora. In one study, it was observed that words that are highly frequent in simplified texts also occur in general corpora, and it was hypothesised that it would be possible to use the general corpora to extract comprehensible synonyms (Glavaš and Štajner, 2015). The advantage of this method is that lexical simplification methods are made available to languages that lack resources such as corpora of simplified texts or lexicons.

There is one commonly used Easy Language corpus for Swedish. LäSBarT (Mühlenbock, 2008; Heimann Mühlenbock, 2013) is a specialised corpus of 1.3 million tokens consisting of Swedish simple texts of four genres: Easy Language news texts, fiction, community information, and children's fiction. The corpus is intended to reflect Easy Language use in various genres, and the intended reader audience is not one specified target group, but is defined as *persons that do not fully master everyday Swedish language* (Heimann Mühlenbock, 2013). Although quite limited in size, the LäSBarT corpus provides a good resource for automatic text simplification and adaptation since it is a compilation of professionally written texts carefully assembled to create a representative sample of Swedish Easy Language.

The DN-LC corpus (Monsen and Jönsson, 2021) was constructed for abstractive text summarisation of Swedish. It was compiled to imitate the English CNN/Daily Mail corpus (Nallapati, Zhou, Santos, Gulcehre, and Xiang, 2016; Hermann, Kocisky, Grefenstette, Espeholt, Kay, Suleyman, and Blunsom, 2015), and comprises $1,963,576$ article-summary pairs in total.

5.3 Resources Targeting Specific Audiences

Since there are few parallel corpora for automatic text simplification for different target audiences, the development and evaluation of automatic text simplification systems targeting specific audiences is hindered.

Dyslexia

Within the ALECTOR project, a parallel corpus of 183 texts in French was compiled (Gala, Tack, Javourey-Drevet, François, and Ziegler, 2020). The resource was primarily made for professionals working with individuals who struggle to read, but it could also be of use for researchers of text simplification. The corpus comprises 79 original texts in literary and scientific genres, and all texts were manually simplified by a group of researchers with respect to four linguistic dimensions: lexical, morphological, syntactic, and discursive. From the ALECTOR corpus, they also compiled a sub-corpus of 20 texts, which contained errors made by children with dyslexia, aligned with the correct word.

There are other resources that identify dyslectic writing errors. In one study, researchers created a resource for English dyslectic writing errors to aid in the development of a spell checker (Pedler, 2007). The resource comprises 3,134 words and 363 errors. A similar resource is DysList (Rello, Baeza-Yates, and Llisterri, 2014), a collection of Spanish writing errors extracted from texts written by individuals with dyslexia. DysList contains 887 misspelled words manually extracted from 83 texts. The words were richly annotated with information about, for instance, the intended target word, frequencies for the misspelled word and target word, number of syllables, type of error, visual information and phonetic information. This type of resource can be useful for assessing how comprehensible a text is, or for use in the development of different text adaptation techniques or other aids for persons with dyslexia.

Intellectual Disabilities

A corpus of news texts was compiled for research on features correlating with the readability of texts for adults with ID (Feng, Elhadad, and Huenerfauth, 2009). Most of the texts were originally written for children, due to the unavailability of texts written for individuals with ID. One part of the corpus (*LocalNews*) included 20 news articles that were manually simplified by a professional writer, targeting persons with ID. Both the original news texts and their simplified versions were evaluated by adults with mild intellectual disability. Using text comprehension tests, each text was given a readability score related to the ratio of correctly answered questions. The corpus was used for an analysis of the linguistic features of texts written for this audience, and a tool was developed to provide a readability assessment for the targeted users. However, due to the small size of the *LocalNews* corpus, this material was not used for training readability models.

Autism Spectrum Disorder

There are a few resources dedicated to individuals with Autism Spectrum Disorder (ASD). One article presented a corpus of texts with 1,034 annotated content words, tagged with both syntactic information and gaze fixation data gathered from eye-tracking sessions with both autistic and non-autistic adult readers (Yaneva, Temnikova, and Mitkov, 2016a). The participants read 9 texts and the reading skills were measured using a set of multiple-choice questions. This resource is useful for the exploration of specific linguistic issues that are experienced as problematic for individuals with ASD, and could serve as a resource for studies of text adaptation targeting this specific audience.

One study explored whether parallel corpora with texts written for children or second language learners could be useful in supporting automatic text simplification systems targeting individuals with ASD (Štajner, Evans, and Dornescu, 2014). This was done by conducting a corpus analysis of four text simplification corpora using a set of features that have proven to be relevant to individuals with hindered reading comprehension in other target audiences. The four corpora, and the corresponding target audiences, were *Weekly Reader* (second language learners), *Encyclopedia Britannica* (children), *Wikipedia* (general), and *FIRST* (people with ASD). Aside from measuring the statistical difference between the features in the corpora, classification experiments were also conducted that attempted to split the texts into simple and original classes using the features. The results of the corpus analysis (which the authors described as ”preliminary”) indicated that two of the corpora, *Encyclopedia Britannica* and *Weekly Reader*, could be suitable for training an automatic text simplification system targeting individuals with ASD.

Second Language Learners

The *OneStopEnglish* corpus (Vajjala and Lučić, 2018) is a collection of 189 English texts with simplified versions written at three different reading levels: *Elementary* level, *Intermediate* level and *Advanced* level. The corpus was compiled from a website for English learning resources where articles from the newspaper *The Guardian* are manually simplified on a weekly basis to these three reading levels by teachers. A version that is automatically aligned at the sentence level was also created, resulting in 1,674 sentence pairs for *ELE-INT*, 2,166 sentence pairs for *ELE-ADV*, and 3,154 sentence pairs for *INT-ADV*.

In one article, a corpus of Japanese news texts and their simplified versions, targeting non-native Japanese speakers, was presented (Goto, Tanaka, and Kumano, 2015). The simplifications were produced by teachers of Japanese with experience of teaching the language to non-native speakers. The resource was made from 490 parallel article pairs, and the resulting corpus is divided into a training set (10,651 automatically aligned instances), a development set (723 automatically aligned instances), and a test set (2,012 manually aligned instances).

In a PhD thesis focusing on lexical simplification for non-native English speakers (Paetzold, 2016), three user studies were conducted, which resulted in three different annotated datasets for non-native English speakers.

- The first user study addressed the issue of *Complex Word Identification (CWI)*, identifying words that were found to be difficult for the diverse group of non-native speakers of English. Sentences extracted from an aggregated resource of three different Wikipedia-based datasets were presented to a number of annotators for the given target audience. The annotators were instructed to mark the words that they could not understand. The resulting resource comprised close to 36,000 annotated words, of which 3,800 words were considered to be complex (as judged by at least one annotator).

- The second user study focused on ***Substitution Selection***, which describes the process of choosing the most suitable substitution for a complex word. From a number of words that were deemed complex, they generated candidates for substitution, and sentences containing these words were then extracted from Wikipedia. The sentences were annotated by fluent English speakers with respect to grammaticality and meaning preservation, resulting in a resource of approximately 25,000 sentences.

- The third user study concerned substitution ranking, which ensures that the selected substitution is simple. For this purpose, another resource was created by presenting non-native English-speaking annotators with 901 sentences extracted from the second user study. Gaps were introduced into the sentences, and the annotators were asked to choose the one of two presented candidate words that made the sentence easier to understand.

In the same PhD thesis (Paetzold, 2016), two new resources for lexical simplification were also presented: SubIMDB and NNSeval. SubIMDB is a subtitle corpus aiming to provide better frequency counts, and NNSeval is an evaluation dataset built on existing datasets, but filtered with the CWI dataset collected in the first user study. This results in an evaluation dataset better suited to the needs of non-native English speakers.

In relation to the other target audiences, the number of resources for text adaptation for second language learners is relatively high, especially for English.

Children

The *Newsela corpus*[2] contains 1,130 original news articles in English, manually simplified by professional writers into simplified versions that have up to five complexity levels. The simplified articles target children of varying reading levels, and the different complexity levels correspond to education grade levels. The Newsela corpus has been described as an alternative to the PWKP corpus, since the quality of the data can be considered higher, both due to a better correspondence between the original and simplified articles, and the fact that the Newsela writers are trained professionals (Xu, Callison-Burch, and Napoles, 2015). Thus, the Newsela corpus is a high-quality corpus that can be freely used for research purposes, but it comes with one big disadvantage: public release of model output (such as sentence alignments) based on this corpus is not allowed, which poses problems for the comparison of research results. The Newsela corpus has been used in various text simplification studies (Zhang and Lapata, 2017; Alva-Manchego, Bingel, Paetzold, Scarton, and Specia, 2017; Scarton, Paetzold, and Specia, 2018).

[2] https://newsela.com/data

In one article, an annotated resource in Italian consisting of two subcorpora was presented (Brunato, Dell'Orletta, Venturi, and Montemagni, 2015). One of the subcorpora comprised 32 short novels targeting children. The short novels were manually simplified and hand-aligned at the sentence level. The other subcorpus targeted mainly second language learners and comprised 24 text pairs where the simplified versions had been constructed by teachers. The corpora were used for a comparison of the different simplification strategies, and for the extraction of features related to text complexity.

Two corpora of web texts in Basque from the science and technology genre were compiled for the purpose of readability assessment (Gonzalez-Dios, Aranzabe, Díaz de Ilarraza, and Salaberri, 2014). The first corpus comprised 200 texts from a journal on science, and the second corpus comprised 200 texts from a science website targeting children. The corpora are not parallel.

Other

Some studies have compiled resources targeting specific reader audiences who do not fit into the categorisation used in this thesis. For instance, Bott and Saggion (2011b) used a corpus of 200 Spanish news articles and their simplified versions, targeting individuals with learning disabilities.

In the PorSimples project (Simplification of Portuguese Text for Digital Inclusion and Accessibility) (Aluísio and Gasperin, 2010), the target audience was low-literacy individuals, and people with other kinds of reading disabilities. Within the project, they created nine corpora of two different genres: news texts and popular science articles. The standard corpora were manually simplified at two different levels, and additional resources for the respective genre—targeting children—were also compiled. The corpora were collected to map the main linguistic components that make a text complex, and to propose simplification transformations.

The *EasyRead* corpus (Yaneva, Temnikova, and Mitkov, 2016b) targeted individuals with cognitive disabilities. The corpus comprised 353 English easy-to-read documents in three genres: news, medical documents, and general information. The documents were collected and assessed using readability formulas, where documents with a Flesch index (see Chapter 6, Section 6.2 for a description of the formula) above 65 were considered easy and were automatically included in the dataset, whereas documents with a Flesch index below 65 were manually examined and either included or discarded. As stated by the authors, this type of resource can be used as a gold standard material for the assessment of output from text simplification systems targeting persons with cognitive disabilities.

5.4 Chapter Summary

This chapter began by discussing the role of corpora in text adaptation research. First, corpora comprising texts written by trained professionals can give valuable information about how Easy Language texts should be written. Second, corpus studies allow large-scale exploration and comparison of languages, genres and domains, which could give us information about how such language is actually used, by real people, in the real world. Finally, large-scale and high-quality text corpora are needed for all data-driven natural language processing methods.

This chapter then described the main resources available today and was mostly concerned with corpora for automatic text simplification. Available resources for Swedish text adaptation were described, as well as corpora targeting different reader audiences.

CHAPTER 6

Text Complexity and Readability

This chapter introduces the notions of readability and text complexity, and provides a description of traditional readability formulas and features used to indicate text complexity. Finally, it describes earlier efforts on the visualisation of such measures.

6.1 Text Complexity or Readability?

Assessing the readability and complexity of a text means to get an objective measure on the ease of reading, understanding or comprehension of a text. Although appearing to be more or less synonymous, there is a significant difference between the terms *text complexity* and *readability*. According to Heimann Mühlenbock (2013), there are numerous definitions of *readability* and the largest difference seems to be the degree to which the reader is included in the equation. According to this perspective, readability covers the characteristics of text in relation to a reader (or a group of readers), whereas text complexity is more of a characteristic of the text itself.

A commonly referred definition of readability is the one provided by Dale and Chall (1949, p. 23):

> *The sum total (including all the interactions) of all those elements within a given piece of printed material that affect the success a group of readers have with it. The success is the extent to which they understand it, read it at optimal speed, and find it interesting.*

Here, readability is described not only as an inherent property of the text, but rather as something emerging from the combination of text and reader.

This thesis follows the terminology convention used by Falkenjack (2018b), which means that the preferred term used for the work conducted within our research projects is generally *text complexity*, as the contributions made so far mainly have concerned pure text characteristics. The features used do indeed say something about the measured text or corpora, but since the reader is not yet included in the equation, we do not know anything about the actual experienced *readability* of a text. However, when referring to the work of others, the terminology might differ depending on what terms the original authors used.

6.2 Traditional Readability Formulas

Early methods for readability assessment made use of basic features of the text to create readability indices.

The *Flesch Reading Ease Score* (Flesch, 1948) is one such example, given by the Equation 6.1, where $n(w)$ denotes the number of words, $n(s)$ the number of sentences, and $n(syll)$ the number of syllables. The score ranges from 0 to 100 and high scores indicate a text that is easy to read, whereas lower scores indicate a more difficult text.

$$\text{FRES} = 206.835 - 1.015\frac{n(w)}{n(s)} - 84.6\frac{n(syll)}{n(w)} \tag{6.1}$$

The Flesch Reading Ease Score was followed by the similar *Flesch-Kincaid Grade Level* (Kincaid, Fishburne, Rogers, and Chissom, 1975) where the resulting score instead correspond to a U.S. grade level, and the *Dale-Chall Readability Formula* (Chall and Dale, 1995) which, in addition to the number of words and sentences, also utilises a ratio of *difficult words*, defined as words not occurring in a given list of easy words.

For Swedish, LIX (Björnsson, 1968) is a commonly used readability metric. LIX is defined as the ratio of words longer than 6 characters added to the average sentence length in words. LIX is given by the Equation 6.2, where $n(w)$ denotes the number of words and $n(s)$ the number of sentences.

$$\text{LIX} = \frac{n(w)}{n(s)} + (\frac{n(words > 6\, chars)}{n(w)} \times 100) \tag{6.2}$$

Another readability formula often used for Swedish is OVIX, translated as *word variation index*. OVIX was developed by Hultman and Westman (1977) and is a measure similar to the type-token ratio, but using logarithms to compensate for problems with varying text length. OVIX is given by Equation 6.3 where $n(w)$ denotes the number of words and $n(uw)$ the number of unique words.

$$\text{OVIX} = \frac{log(n(w))}{log(2 - \frac{log(n(uw))}{log(n(w))})} \tag{6.3}$$

These are only a few of the traditionally used formulas for measuring readability (by the 1980s, 200 such formulas had been developed (DuBay, 2004)). Naturally, with the development of various tools for automatic analysis of texts, and the possibility of working with very large text resources, novel methods for assessing text complexity have emerged.

6.3 Other Complexity Measures

Features related to coherence and cohesion have received some attention.

Coh-Metrix (Graesser, McNamara, Louwerse, and Cai, 2004a) is a tool for assessing text cohesion as well as other measures of readability. Coh-Metrix goes beyond more superficial measures such as the length of linguistic units, and includes measures based on, for instance, Latent Semantic Analysis, which provides a representation of the semantic meaning of words and texts, as well as the density of connectives in a text, in order to grasp the more abstract notion of text cohesion. The extent to which such features succeed to capture true text cohesion is not clear. For instance, Todirascu, François, Bernhard, Gala, and Ligozat (2016) explored cohesive features for text complexity assessment for French. They found that such features did not seem to discriminate well in the task of complexity prediction.

There has also been some research on data-driven text complexity assessment for Swedish. Pilán, Vajjala, and Volodina (2016) classified texts for Swedish second language learning and showed that lexical features were the most predicative for document-level complexity assessment, whereas a combination of features was more successful for analysis at the sentence-level. Moreover, the type of lexical features provided matters. Pilán, Alfter, and Volodina (2016) combined receptive and productive text data, i.e. texts written *for* second language learners and texts written *by* second language learners in order to improve classification of texts into proficiency levels. They showed that using lexical features from receptive sources could enhance classification performance on essay texts written by second language learners.

Heimann Mühlenbock (2013) presented the SVIT model of readability: a combination of features on different linguistic levels that performed an accuracy of 84–100% at a binary classification task separating Easy Language texts from non-Easy Language texts. Falkenjack, Heimann Mühlenbock, and Jönsson (2013) presented the feature set that we use in most of our work: the SCREAM features.

SCREAM Text Complexity Features

The SCREAM features comprise approximately 120[1] features of text complexity divided into four categories. In this thesis, the categorisation used in Falkenjack (2018b) is adopted. This categorisation is based on the levels of *increasing linguistic involvement* and is mainly inspired by the work of Dell'Orletta, Montemagni, and Venturi (2011) and Heimann Mühlenbock (2013):

- **Shallow Features** include features that can be extracted without deeper text processing. Such features include mean word length and mean sentence length. The length of text units is a commonly used complexity feature, and can for that reason be found in, for instance, several traditional readability formulas, such as the LIX and FRES formulas. To use word length and sentence length as predictors of text complexity is coherent with common guidelines for writing Easy Language texts. For instance, the MTM guidelines (MTM, 2021) suggest using *short, simple* words, and many similar guidelines suggest expressing text ideas in a precise and concise way (PLAIN, 2011; Misako Nomura and Tronbacke, 2010), which in turn means that it should be as brief as possible and not use any unnecessary words (PLAIN, 2011).

- **Lexical Features** include features that are extracted by consulting a lexical resource of basic Swedish vocabulary (SweVoc (Heimann Mühlenbock and Johansson Kokkinakis, 2012b)). SweVoc is a lexical resource totalling about 7,600 word lemmas, divided into sub-categories. SweVoc is analogous to the list of simple words used in the Dale-Chall formula (Chall and Dale, 1995). The intuition is that a higher ratio of SweVoc words would imply a less complex text.

- **Morpho-syntactic Features** include features that require part-of-speech tagging, i.e. unigram probabilities for different part-of-speech tags and the ratio of content words.

- **Syntactic Features** includes features that can be extracted after dependency parsing. Such features incorporate, for instance, the average dependency distance of a text (where a longer distance could imply a more complex text (Liu, 2008))).

The complete list of SCREAM features can be found in Falkenjack (2018b) and Falkenjack, Heimann Mühlenbock, and Jönsson (2013).

[1] The exact number of feature varies slightly between the articles depending on implementation and, sometimes, selection of subsets.

6.4 Visualisation of Text Complexity

Various techniques have been suggested in previous studies on visualisation of readability and text complexity. Eika and Sandnes (2016) proposed a purely typographic and text-based visualisation, revealing specific issues in the texts. They focused specifically on surface-level indicators of readability, such as sentence and word length. For instance, long words are emphasised using uppercase characters, and long sentences and prepositional phrases are highlighted. This approach infers that the user, i.e. the author of a text, must make their own interpretation of the readability of a text, based on its formatting.

Kim, Park, and Seo (2014) used shaded consecutive dots to visualise readability, based on text length. The initially white dots darken as a sentence progresses, while punctuation marks (commas and periods) slightly lighten them. This provides an overview of document-level readability, as more difficult texts with long sentences and paragraphs appear darker. The visualisation was evaluated through a survey where participants rated text excerpts from books regarding readability (on a 5-point scale), and rated the perceived impression of the visualisations. The evaluation revealed a correlation between the predicted readability of the texts, and the corresponding visualisation.

Two suggestions for visualisation of readability at the paragraph level were presented by Karmakar and Zhu (2010a). First, the complexity of each paragraph was indicated by a coloured circle, where the colour represented a readability score. The various readability indices, represented by colour-coded abbreviations of the given readability index, could also be displayed on demand. Second, they used the shape and size of different features of a cartoon face to each represent a measure of overall complexity for a given paragraph. The idea was that the cognitive system that ensures fast processing of facial expressions would allow for a quick assessment of the readability. Karmakar and Zhu (2010b) also suggested visualisation by coloured bars. Each sentence was represented by a bar that either indicated word complexity (given by word length in syllables or characters, and a vocabulary-based measure using a list of simple words), or structural complexity (given by number of sub-clauses and parse tree depth), by coloured blocks within the bars.

Oelke, Spretke, Stoffel, and Keim (2012) used a continuous colour scale to indicate readability on three levels: document, paragraph and sentence. At the sentence level, beyond overall readability, the vocabulary, word length, nominal forms, sentence length and sentence structure were each indicated in their respective coloured box. The intention of the visualisation was to indicate complex parts of the text, but also to make the user understand why the specific parts were problematic. This was the reason for choosing a more fine-grained approach with three levels of analysis.

In a tool for Portuguese text classification support which classified texts according to reading proficiency level, Branco, Rodrigues, Costa, Silva, and Vaz

(2014) suggested to visualise certain features related to readability in radar charts. Out of 15 complexity features, four primary features with strong correlation to readability were presented in the chart, indicated by their accordance to a 5-point scale corresponding to language level.

Pilán, Alfter, and Volodina (2016) visualised lexical complexity by highlighting words in the text using a colour code representing different reading proficiency levels. The visualisation differed between words originating from sources of productive and receptive language respectively, using different shades of the same colour.

The evaluations of the above-mentioned visualisation have mostly consisted of case study demonstrations exemplifying the visualisations' usage (Eika and Sandnes, 2016; Karmakar and Zhu, 2010a; Karmakar and Zhu, 2010b), or evaluation of the text analysis performed by the tool (Oelke, Spretke, Stoffel, and Keim, 2012; Branco, Rodrigues, Costa, Silva, and Vaz, 2014). How the visualisations themselves are understood remains unexplored.

6.5 Chapter Summary

This chapter has provided an introduction to text complexity in general and how it relates to readability. Some of the most used traditional readability measures were presented, as well as modern data-driven text complexity features. Finally, we described efforts on text complexity visualisation.

Part II

Exploring Easy Language Texts

CHAPTER 7

Integrating Easy Language Guidelines into a Model of Comprehension

In Chapter 3 we described several sets of Easy Language guidelines. In this chapter, we introduce the CI model of text comprehension and use it as a framework to review Easy Language guidelines. According to the model, text is processed at three levels: Surface Model, Text-Base Model and Situation Model. This chapter presents a novel mapping of existing guidelines of writing Easy Language text to the different levels of the CI model and explore the applicability of such guidelines in an automatic system.

7.1 The CI Model

In an important theoretical model for reading understanding, Kintsch and Van Dijk (1978) presented a cognitive base for reading comprehension, the Construction-Integration model (CI model) (Kintsch, 1988; Kintsch and Van Dijk, 1978; Kintsch and Mangalath, 2011), see Figure 7.1. In the CI model, information is stored in the form of propositions which correspond to individual idea units. Comprehension of text occurs in two sequential stages. When a sentence is read, the following three sources of information are combined into a connected network: the current idea, a few important ideas from semantic long-term memory and a very limited number of ideas from the previously read text. This is called the *construction* phase. In the *integration* phase, associative connections between the propositions are calculated such that concepts that co-occur in working memory across cycles are strengthened. After the integration, only a few ideas are active in working memory, and the process starts over as another sentence is read.

Important features of the model include the different levels of processing and the emphasis on making inferences to form coherent representations at each level. In the model, text segments are processed at three levels simultaneously: the *Surface Model*, which is the initial grammatical representation of the sentence; the *Text-Base Model*, which corresponds to representation of the literal content of the text in propositional form; and the *Situation Model*, which is a representation of the text content related to the reader's background knowledge.

Establishing connections between propositions from the text requires that the reader understand how the ideas are related. The connections between the currently read proposition and preceding information establish local coherence. An example of such a connection is cause and effect, which may or may not be explicitly stated in the text. Local coherence of the text is established through the *microstructure*, a network of connections between propositions in the text, and the *macrosctructure*. The macrostructure of the text is the hierarchical discourse structure of the text, that is the global structure or organisation of the text. At this level the order of the content is key. To understand a text within the CI framework, the text must be coherent at the micro- and macrolevel, such that semantic relations are established. When

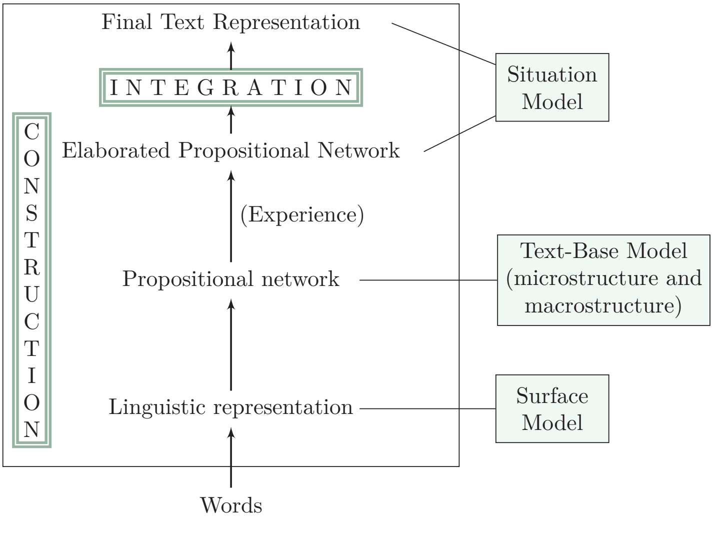


Figure 7.1: CI as a Cognitive Model of Text Comprehension. Adapted from Kintsch and Van Dijk (1978).

stated this way, a text's difficulty depends on how well the reader is able to repair breaks in text coherence by inference making and the effort required to fulfil that task (Arfé, Mason, and Fajardo, 2018). Text based inferences, establish local coherence between ideas in the text, and knowledge-based inferences form connections between the text ideas and world knowledge (Carlson, Broek, McMaster, Rapp, Bohn-Gettler, Kendeou, and White, 2014). Making knowledge-based inferences entails connecting the literal content, the Text-Base, and the reader's knowledge to form the Situation Model.

7.2 Surface Model Guidelines

The Surface Model of the CI model of text comprehension denotes processing at the level of the initial grammatical representation of the sentence. In this section, we have divided the Surface Model into the process of selecting words, and the straightforward expression of text ideas. We describe existing guidelines related to text comprehension at the Surface Model processing level, and investigate which of these guidelines are suitable for implementation in an automatic adaptation system.

Selection of Words

Building a mental representation of text at the surface level requires understanding the individual words. Plain English guidelines (PLAIN, 2011) suggest writers *use short, simple words*, which, they elaborate, requires writers pick *the familiar or frequently used word over the unusual or obscure*. Similarly, according to the MTM guidelines, Easy Language texts should consist of simple, short words (MTM, 2021). For example, the MTM guidelines suggests using *buses and trains* instead of *public transportation*.

The IFLA guidelines for easy-to-read-materials (Misako Nomura and Tronbacke, 2010) advise avoiding abstract and difficult words. However, they point out that the language used should still be *adult and dignified*, which means that, for example, unusual words could be used, but should then be explained. Jargon should be avoided if possible (PLAIN, 2011; Språkrådet, 2014), and if such words must be used, they should be explained at their first appearance, or even every time they are used (Inclusion Europe, 2020). One way to do that is to include an epithet connected to the word such as *the city Madrid* or *the virus Covid 19* (MTM, 2021).

Swedish words, especially compound words, can be very long, and it is tempting to divide long words when reaching the end of the line. However, this is not recommended (Inclusion Europe, 2020; Begriplig Text, 2019) as it could hinder the reading flow and make the word difficult to comprehend. The split word parts can also sometimes be interpreted as two different words.

The issue of how numbers should be written is not straightforward. According to the Inclusion Europe guidelines, numbers should be written in digits (Inclusion Europe, 2020). The instructions for writing plain language for Swedish public authorities (Språkrådet, 2014) have more extensive instructions depending on context. For example, letters are often used for low numbers (twelve or lower), as well as very high numbers (*one million*), whereas digits should be used when the numbers are in specific focus (tables, statistics, maths). The MTM guidelines (MTM, 2021) simply state that the reading of numbers should be simplified, for example by writing *half* rather than 50%.

The guidelines for writing clear text for public authorities (Språkrådet, 2014) contain much word-level advice. One such recommendation is to use clear and comprehensible words. For example, writers should avoid old-fashioned words and expressions and words that may have many different senses. There is a *black list* (Stadsrådsberedningen, n.d.) of such words to avoid when writing texts for public authorities. For instance, instead of writing *Såvitt anser (Eng: As far as considered)*, one should write *När det gäller (Eng: Regarding)*.

Other general guidelines covering word selection from the guidelines of writing clear text for public authorities (Språkrådet, 2014) include keeping words as short as possible, by not adding additional words to create compound words. For instance, it is advisable to use *learning* instead of *learning behaviour*. Technical terms, buzzwords and empty phrases should be avoided if possible. In Swedish legal texts, it is common to use *naked nouns*, i.e., non-conjugated singular nouns without any article, and this should generally be avoided.

Abbreviations should be avoided if possible (Begriplig Text, 2019; Språkrådet, 2014; PLAIN, 2011). Abbreviations might be difficult to read and comprehend, but they also have another disadvantage: they could be an obstacle for different kinds of reading tools, such as read-aloud applications. Some exceptions are noted in the guidelines for writing plain language for Swedish public authorities (Språkrådet, 2014), including very common abbreviations and units of measurements.

Automatic Adaptations for Selection of Words

There are several guidelines regarding the word selection for simple texts, but sometimes, the guidance is limited to different versions of *use simple words*. The question is then, what makes a word simple?

A common way of defining simple words is to say that common and colloquial words can be considered simple, and words that are less frequent are more difficult. The conventional way of measuring word frequencies is to use the relative frequencies in large text corpora. One way of doing this is to look up the frequencies in large text corpora. However, as pointed out by Wengelin (2015), the distribution of frequencies may vary depending on the corpora used. The genres of the texts in the corpora will affect the frequency distribution, as well as whether the corpora consist of spoken or written words, or target audiences of different demographics. Moreover, as noted by Alfter (2021), more frequent words are often more polysemous than less frequent words, which could indicate a higher level of complexity. Furthermore, for the individual reader, low-frequency words can be easy to read if the reader is familiar with the domain (Gernsbacher, 1984). For example, *snaffle bit* and *stirrup* may be words with low frequency in general corpora, thus considered difficult to read and comprehend, but can be simple for a reader who is familiar with the equestrian domain. Aside from frequency measures, other features that could indicate complex words include, for instance, word length, or psycholinguistic features such as age of acquisition (Morrison and Ellis, 1995) or concreteness given by resources such as the Bristol Norms (Stadthagen-Gonzalez and Davis, 2006) or the MRC Psycholinguistic Database (Wilson, 1988).

Regardless of how we define *simple words*, the substitution of words that are more complex for simpler words with the same meaning (a process commonly known as *lexical simplification*), is a rather well-studied area (see for instance Paetzold and Specia (2017) for an overview). Although it is a challenging task with many non-trivial subtasks (identifying complex words, disambiguating word senses, etc.), the guidelines of substituting complex words can be considered possible to automate. Avoiding jargon and technical terms can be supported with specialised term lists, such as the *black list* (Stadsrådsberedningen, n.d.) used by the Swedish public authorities. Not splitting words onto two lines and writing numerical expressions in an easy-to-comprehend way according to set standards are other guidelines that are relatively easy to automate.

Expression of Text Ideas

Information in a text is processed in the form of propositions (Kintsch, 1988). Easy Language guidelines offer many recommendations on how ideas in a text should be expressed to be as accessible as possible. Text ideas should be expressed in a precise and concise way (PLAIN, 2011; Misako Nomura and Tronbacke, 2010). A concise text should be as brief as possible and not use any unnecessary words (PLAIN, 2011). For example, the word *to* can be used instead of *in order to.* The sentences should also be short with preferably one proposition per sentence. However, to improve the reading experience, Myndigheternas skrivregler suggest mixing short and long sentences (Språkrådet, 2014). To make the text precise, writers of Easy Language texts should also avoid abstract language and be as concrete as possible. Metaphors are examples of abstract language that can be challenging to understand for many readers and should be avoided (Misako Nomura and Tronbacke, 2010; MTM, 2021).

The form and usage of a verb in a phrase is essential in Easy Language writing. In general, strong verbs with clear meaning should be used to express action in a sentence in a straightforward manner. This entails avoiding passive verbs and nominalisation (PLAIN, 2011; Språkrådet, 2014). For example, *Säpo investigates sabotage* is a better sentence than *Säpo is realising an investigation of sabotage*. The second version is more difficult to understand as *realising* carries less meaning compared to *investigate*, and *investigation* is a nominalisation of the verb. If possible, you should express how things are and avoid expressing how they are not. Negative statements are not as straightforward to understand as positive statements.

Regarding the length of the verb, if there are several different forms of a verb (*Swe: draga/dra, Eng: pull*), the shorter form is preferred. Another recommendation, applicable for Swedish, is to avoid merging phrasal verbs (such as *heat up/hetta upp* as one word *upphetta*). In these cases, the multi-word verb forms are considered easier to read compared to the merged verb forms (Språkrådet, 2014; MTM, 2021).

In certain types of texts, such as official texts, personal pronouns are rare, and sentences lack subjects. The general approach of Easy Language is to include subjects in the sentences. This could be done by addressing the reader directly rather than writing texts that have no apparent recipient (PLAIN, 2011; MTM, 2021; Inclusion Europe, 2020). Another suggestion, relevant for Swedish is to include the pronoun *man* (in English *one*) to refer to an unspecified person.

Regarding the order of the words in a sentence Plain language guidelines state *"The verb tells the audience what to do. Make sure they know who does what."* (PLAIN, 2011). A few related recommendations come from Myndigheternas skrivregler (Språkrådet, 2014). Sentence content is most efficiently expressed with the main verb early in the sentence. The subject of the sentence should precede the main verb. In general subject, verb and object should be kept close together in the sentences. Also, a modifier should be placed after the object which it modifies, for example *the red-haired girl* could be replaced by *the girl with red hair*.

Automatic Adaptations of Text Ideas

Automatic text adaptations can help keeping the text brief. For example, superfluous words and phrases can be recognised and deleted. Such simplification operations have been previously identified for Swedish Easy Language text (Decker, 2003), and while operations like these are relatively simple to implement from a technical point of view, there is a risk that relevant text information might be deleted in the process, which could cause confusion or impair the reading flow of the reader.

As previously mentioned, one guideline suggests limiting content to one proposition per sentence. Following this guideline is slightly more complicated than removing extraneous words, as it requires some semantic parsing. Event-based simplification (see for example Štajner and Glavaš (2017)), which identifies mentions of factual events and deletes sentences or parts of sentences that are irrelevant to these event mentions might offer a solution. Such simplification approaches could enhance text comprehension by deleting irrelevant information and highlighting the main information, but these adjustments will naturally result in some loss of information. It is clear that the deletion of words, phrases or information could result in a more readable text, but there is also a risk that that the resulting text is, in fact, less readable. This could be due to loss of core information, as described above, but it could also be due to typographic reasons, i.e., features of the adjusted text layout may make the text less appealing to read. Mixing long and short sentences, as suggested in another guideline, could be considered as a parameter when applying guidelines that aim to write as briefly as possible.

Avoiding abstract language requires a system that can identify markers of abstract language, such as metaphors, and either explain the abstract concept to the reader, or substitute replace the expression with a more concrete option. However, rewriting metaphors is a challenging task (Drndarević and Saggion, 2012), and has, to the best of the authors' knowledge, not yet been addressed in any adaptation system, although it has been identified as a suitable task for machine learning techniques (Wolska and Clausen, 2017).

Guidelines concerning specific verb choices (using strong verbs, nominalisations, phrasal verbs, etc.) border on lexical simplification and can, in many cases, be solved with similar methods. The rewriting of passive voice to active voice is another common simplification operation that has been solved by the use of hand-crafted simplification rules (Siddharthan and Mandya, 2014; Rennes and Jönsson, 2015).

It is possible to automate guidelines that change negative statements to positive statements, but a mechanism is required to identify such structures, along with a set of rewriting suggestions. For relatively simple cases, such as *no fewer than* → *at least* (PLAIN, 2011), the task is more or less analogous to lexical simplification, but for more complicated cases with, for example, double negations, the task is slightly more complex. Some work has been done on identifying and substituting negations within the medical domain (Burgers, Beukeboom, Sparks, and Diepeveen, 2015; Mukherjee, Leroy, Kauchak, Rajanarayanan, Diaz, Yuan, Pritchard, and Colina, 2017).

The use of personal pronouns is generally recommended, as is addressing the reader directly. Such linguistic adaptation strategies have previously been, at least partially, implemented in a rule-based simplification system (Rennes and Jönsson, 2015) (see also Chapter 10 of this thesis). In the same system, a rule for reordering sentences was implemented to keep a straightforward word order, with subject, verb and object kept close together.

7.3 Text-Base Model Guidelines

The Text-Base Model in the CI model represents the processing of a text's literal content in propositional form. The Text-Base Model is divided into microstructure and macrostructure. We have explored the Easy Language guidelines affecting comprehension at these processing levels and investigated the possibility of automatic implementation.

Microstructure

According to the CI model, the reader builds a representation at the microstructure level by making connections between different propositions in a text. The process of connecting statements is often referred to as making inferences, in this case text-based inferences, or connections between text ideas which make the text coherent. The task of Easy Language at this level is therefore to highlight the connections between statements. This could, for instance, be done by including if-then statements (PLAIN, 2011). Plain language guidelines even suggest using if-then tables, to highlight the connections.

Adding conjunctions such as *but*, *because* and *therefore* (MTM, 2021) between statements strengthens the connections between them. In order to facilitate connections between text ideas the writer should also avoid using synonyms when referring to the same object (PLAIN, 2011). The Easy Language text will therefore be more repetitious than varied in word choice. Another aspect of making connections at a more basic level is to make anaphoric references very clear. This could entail using the names of characters more often rather than referring to them as *he* or *she* (Inclusion Europe, 2020).

Making coherent connections within a text can be complicated if text ideas are split up. Therefore MTM (2021) suggest that inserted clauses should be avoided. For example, *The girl was hungry so she ate a hot dog* is preferred over *The girl, who was hungry, ate a hot dog.* Also, if you are presenting a main idea with conditions or exceptions, you should always present the main idea first. It is more difficult to make connections between the main idea and the exceptions when the exceptions are presented initially (PLAIN, 2011).

Automatic Adaptations for Microstructure

Automating guidelines about facilitating text connections is not straightforward, as it demands a deeper understanding of the semantic content in order to keep a text coherent.

The creation of if-then tables, as suggested in the Plain Language guideline (PLAIN, 2011), requires that the relevant text segments (for example conditions and consequences) are identified, and that relations between text segments are identified and lifted out of the text mass. The same goes with the guidelines to use conjunctions between statements, such as *but, because, therefore*. Such adaptation operations are relatively complex to automate as they require a semantic interpretation of the text.

The challenge of how to make anaphoric references clear is a known, and for Swedish, a partly unsolved, problem when dealing with different kinds of automatic text adaptation techniques. When information is deleted or re-ordered, the antecedents of anaphoric referencing expressions could be deleted or exchanged, which could affect text cohesion as a whole and even alter the meaning of the text. Thus, the guideline covering anaphoric references is not straightforward to implement in an automatic text adaptation system.

However, there are guidelines on this level that are easier to implement in an automatic adaptation system. To avoid using different words when referring to the same object is one such a guideline and deleting inserted clauses is another. The latter has been implemented in a rule-based simplification system for Swedish (Rennes and Jönsson, 2015).

Macrostructure

The macrostructure of the text describes, according to the CI model, the text structure and organisation. There are several Easy Language guidelines that cover aspects of macrostructure.

Some guidelines mention that texts should follow a *logical structure* (Misako Nomura and Tronbacke, 2010; Språkrådet, 2014), and that it should have a common thread (MTM, 2021; Språkrådet, 2014). The course of events should be described chronologically (MTM, 2021; Inclusion Europe, 2020; Misako Nomura and Tronbacke, 2010), and cross-references avoided (PLAIN, 2011). The Plain Language guidelines also state that a good organisation principle is to present general information in the beginning of the text, whereas more specialised information and exceptions should be given later in the text (PLAIN, 2011). It is also suggested that the most important information be presented in the beginning of the text (Begriplig Text, 2019). The text should not introduce too many new people to the reader (Misako Nomura and Tronbacke, 2010; MTM, 2021).

The guideline to make sure that *the main information is easy to find* (Inclusion Europe, 2020) is another, slightly more vague, guideline related to this. This can be done by presenting the most important information in the beginning, but some guidelines suggest other strategies to highlight the most important information. One such example is to emphasise key words and phrases in different ways, for example using bold face (PLAIN, 2011; Begriplig Text, 2019; Språkrådet, 2014). It is not recommended that uppercase letters be used for this purpose, as it makes the text more difficult to read (PLAIN, 2011). Other recommended strategies include using different sizes or weights of the font, or using other design elements such as shading and extra white spaces, but since these strategies border on being graphical elements, we will not go into detail on them.

The use of fact boxes (Språkrådet, 2014), bullet lists (Begriplig Text, 2019; Språkrådet, 2014; Inclusion Europe, 2020), tables (PLAIN, 2011), and clear and descriptive headlines (Begriplig Text, 2019; MTM, 2021; PLAIN, 2011; Språkrådet, 2014; Inclusion Europe, 2020) can also highlight and clarify text structures.

If a reader is provided with a summary of the text, they will have prior knowledge of the text's content when reading the full text. This could, in turn, facilitate reading of the full text (Begriplig Text, 2019; Språkrådet, 2014). A similar purpose is fulfilled by a summarising preamble in the beginning of the text (Begriplig Text, 2019).

The text should also be divided into paragraphs, each containing one theme (Begriplig Text, 2019; Inclusion Europe, 2020; Språkrådet, 2014; PLAIN, 2011). Short sections divide a text into smaller chunks, which gives a lighter appearance to the text, and facilitates organising the text, especially combined with boldface subheadings (PLAIN, 2011; Begriplig Text, 2019).

Automatic Adaptations for Macrostructure

Some of the guidelines concerned with enhancing comprehension at the macrostructure level could be easily implemented in an automatic text adaptation tool with existing techniques. For example, to *make the main information easy to find*, automatic extraction of keywords could be implemented, and those keywords could be presented in clear ways (boldface, headlines, bullet lists, etc.). Another example of a relatively straightforward implementation is to provide an automatic summary of the text. Keyword extraction and extractive summarisation (extracting the most important sentences of a text) are techniques that could be relatively easily implemented, whereas abstractive summarisation (rewriting the summary from scratch) requires more sophisticated methods and data for training. It might be possible to use the same keyword extraction and summarisation techniques for the guideline to let the general and the most important information be presented in the beginning.

Avoiding cross-references is a more complex problem as it requires some semantic processing of the text. Although such references could be relatively easy to identify in a text, resolving them is a more complex subject.

Several different techniques could be used to divide the text into paragraphs, where each paragraph treats one theme or topic, such as automatic text summarisation or event extraction systems. However, the main issue here is probably keeping the text-level cohesion, as text segments required to make sense of the text might be deleted in the process.

Other guidelines modifying a text's macrostructure are more complicated. To keep the *logical* structure of a text, and let the text keep a common thread are difficult to operationalise due to the abstract nature of the advice. Another common piece of advice is to keep the text in chronological order. Such discourse-level adaptation operations are, to be best of the authors' knowledge, still unsolved.

7.4 Situation Model Guidelines

The Situation Model in the CI model is a representation of the text content related to the reader's background knowledge. In this section, we describe the existing guidelines as they relate to the Situation Model, and discuss whether it is possible to implement these guidelines in an automatic system.

In order to make a coherent representation of a text, the reader must make knowledge-based inferences to establish global coherence. Therefore, Easy Language should aim to facilitate making such connections. In Plain Language guidelines this is stated using a couple of questions including: *Who is my audience? What do they know? What do they need to know? What questions will they have?* (PLAIN, 2011). Specifying the aim of the text is another aspect of target group adaptation (Språkrådet, 2014).

It is not an easy task to ensure that readers have the amount of knowledge required to understand a text. Concepts that are frequently used and obvious to one reader because of certain interests might not be familiar to another (Gernsbacher, 1984). For a second language learner, common cultural phenomena like traditions may be unfamiliar.

Another way of enabling readers to make inferences is to explain difficult concepts, preferably each time they are used (Inclusion Europe, 2020). If the concept is explained only once the reader must either learn the concept at the first encounter or go back and forth in the text to get the description. Both of these actions are demanding for poor readers.

Automatic Adaptations for the Situation Model

There are two main guidelines to be considered with regard to making model-based inferences. First, the more general guideline to *know the reader* means that the text should be written considering the background knowledge and references of the intended target audience. This is, for obvious reasons, a difficult guideline to automate as one single automated guideline. However, the more we know about the target audience, the more we can fine-tune the individual adaptations, thus, this general guideline could be seen as the sum of the individual parts. In this case, this is possible to implement in a system. However, it requires knowledge about the difficulties and challenges that differentiate the reader audiences.

The second guideline requires writers to explain difficult concepts to the reader, which could provide a partial solution to the more general guideline. To automatically identify words and concepts that could be difficult to understand, depending on the reader audience, and to provide definitions of such concepts is a task that can be automated, and could provide a partial solution to facilitate knowledge-based inferences.

7.5 Writer's Intuition

There are some guidelines that do not have a clear place in the CI model of text comprehension: guidelines that, in some way, refer to the common sense of the writer. Although it could be argued that such advice is outside the scope of this thesis, as the guidelines do not refer to anything purely linguistic *per se,* they are worth mentioning since they often have something to do with the degree or extent of the linguistic adaptations. For instance, the guidelines of Begriplig Text (Begriplig Text, 2019) state: *Is the text as long as it should be for you to say what you want to say?* and similarity, the Plain Language guidelines (PLAIN, 2011) state: *Use language your audience knows and feels comfortable with.* Such guidelines are present at all text levels and require writers to make decisions based on a feeling for language, intuition or expertise, in order to adapt a text so that it is as comprehensible as possible, but without affecting the reading flow that is experienced in a negative way.

It is difficult to identify specific adaptations that can be automated and implemented in a system to address these kinds of intuitive assessments. Aside from the specific adaptations we have suggested, it is possible that these guidelines are examples of advice that simply cannot be implemented in a clear way in automatic adaptation systems.

7.6 Connection to Reader Audiences

Some features of the Surface Model related to guidelines regarding the selection of words, as well as guidelines intended to ensure straightforward expression of text ideas. For example, the text should contain common words and straightforward word order (PLAIN, 2011). Ideally, lexical adaptations should support comprehension in readers with limited grammatical understanding and vocabulary, although studies have shown that it may not be that simple (Fajardo, Avila, Ferrer, Tavares, Gomez, and Hernandez, 2014). Another aspect of the possible effects of Surface Model adaptations is to limit working memory load in readers where decoding ability is not automated, for example in individuals with dyslexia. Connected to the CI model, adaptations are meant to increase construction of comprehension through simplification of the linguistic content and load on the cognitive system.

The Easy Language guidelines related to the microstructure level are advice that facilitate making inferences to make the text coherent. An example is to include connectives (MTM, 2021). Difficulties in inference making are an important deficit for many poor readers for instance in children with automatic decoding but severely depressed reading comprehension, i.e. poor comprehenders (Cain, Oakhill, and Bryant, 2000; Hulme and Snowling, 2011). In a study on poor comprehenders, Hua and Keenan (2014) found that the students could make the inferences given that they could recall the aspects to connect, which was frequently not the case. Inference making is highly dependent on paying attention to the most important aspects of a text (Kendeou, Van Den Broek, Helder, and Karlsson, 2014). Clearly stating connections between statements likely help in building a more coherent representation in microstructure. Adding connectives to enhance comprehension has been successful in children (Beck, McKeown, Sinatra, and Loxterman, 1991). But efficiency of such a strategy may depend on aspects such as type of connectives and difficulty of the text (Kleijn, Pander Maat, and Sanders, 2019).

Guidelines related to the macrostructure level concern advice regarding the (re-)organisation and structure of the text content. For example, this could be done using headlines or highlighting important words (PLAIN, 2011). Essentially these adaptations aim to give processing instructions to the readers, to emphasise important aspects of the text which the readers may otherwise miss. This approach appears to work in typical readers (Sanchez, Lorch, and Lorch Jr, 2001).

The guidelines related to the Situation Model concerned text that encourage making knowledge-based inferences. An example is to explain rare terms repeatedly in the text (Inclusion Europe, 2014). Here, other factors than purely language-related guidelines are covered, such as matching the text to the reader's knowledge. In reading comprehension instruction for school age children this is often operationalised as activating prior knowledge before reading the text (Palincsar and Brown, 1984).

7.7 Chapter Summary

This chapter reviewed Easy Language text adaptations using an established framework of reading comprehension; this perspective differs from the purely text-based perspective that is commonly used within the field. This position holds that it is important to take into account the strengths and weaknesses of individuals in the different groups of poor readers when evaluating their response to adapted text.

Most guidelines can be mapped to the different levels of the CI model, and guidelines can be identified to work at all levels of the model. However, some guidelines were more difficult to address in the CI model, in particular, the advice referring to the common sense of the writer. Such intuition-based guidelines are present at all levels, but instead of taking the form of concrete guidelines, such advice is connected to the degree or extent of adaptations, the gained expertise of the writer, and some knowledge about the intended reader.

We explored the possible paths for implementing the guidelines in a system for automatic adaptation. For instance, substituting words and highlighting information are tasks well suited for automatic adaptations, whereas adaptations within sentences and clarification of connections between sentences are more difficult to automate. The intuition-based guidelines, which require drawing on the writer's common sense, experience and expertise, are perhaps also examples of advice that are more difficult to implement.

CHAPTER 8

Text Complexity and Visualisation

This chapter treats the research field that is concerned with the assessment of text complexity, a field closely related to automatic text adaptation. My own contributions in this field mainly focus on visualisation of text complexity, but text complexity features have played an important role in several parts of the thesis work. For instance, we have implemented features that indicate text complexity in the tools and services that will be described in Chapter 12, and the features have been used to assess the collected corpora described in Chapter 9.

We present the results of two studies. In the first study, we explored a component-based text complexity analysis method, and showed that the extracted components can be used to classify texts in genres in a corpus of texts in standard Swedish. In the second study, we focused on visualisation of text complexity facets extracted by a factor analysis. We saw that linguistically similar texts had similar complexity shapes. The shapes can aid the interpretation of text characteristics, and can be used by readers, writers, or teachers.

8.1 Text Complexity Features

In the main part of our work, we have used the SCREAM features (Falkenjack, 2018b). This large set of features are developed for Swedish, and include both calculations of traditional readability formulas as well as features describing text at the lexical, syntactic and morpho-syntactic levels. Although the features are briefly described in Chapter 6, the interested reader should delve into the complete description of the work by Falkenjack (2018b).

8.2 Making Sense of the Measures

Understanding the complexity of a given text could be useful for various reasons. It could allow individuals with reading impairments to get a quick overview of the complexity level of a given text. It could also, for example, support teachers in finding texts that are at an appropriate reading proficiency level for a given student or student group or assist text producers when writing Easy Language text.

Even though it is clear that the SCREAM features (and other similar measures) are predicative in the task of distinguishing simple language text from standard text, the features themselves are rather opaque and difficult to interpret for non-experts. Thus, in order to make them useful for an end user, whether that user is a reader or a writer, something must be done to make sense of them. We have approached this issue from two directions. Firstly, we have attempted to cluster the features into smaller groups of measures. Secondly, we have visualised the measures in diagrams. The next sections will describe these two tracks of work.

Clustering Features

This section describes the application of a Principal Component Analysis (PCA) to the SCREAM features. The PCA was performed to reduce the number of dimensions and divide the features into a set of components. Our aim was to investigate whether it was possible to distinguish different genres in a corpus of standard text using text complexity measures factorised into components. The ways in which text complexity measures can be combined and clustered to be more comprehensive has been studied before, c.f. Falkenjack, Santini, and Jönsson (2016).

Corpus

We used texts from the Stockholm Umeå Corpus, SUC (Ejerhed, Källgren, and Brodda, 2006), which is annotated by genres. *Genre* is defined as a text variety based on conventionalised textual patterns[1]. Applying this definition of genre to the nine top genres of SUC, we ended up with six genres, see Table 8.1.

[1] *Genre* should be distinguished from the term *domain*, which refers to the shared topic of a group of texts. For a more detailed description, see Falkenjack, Santini, and Jönsson (2016).

Table 8.1: The six proper SUC genres used.

Genre	Size
Press Reportage (**A**)	269
Press Editorial (**B**)	70
Press Review (**C**)	127
Biographies/Essays (**G**)	27
Learning/Scientific Writing (**J**)	86
Imaginative Prose (**K**)	130

Procedure

A Principal Component Analysis was conducted in order to group the SCREAM features. We first excluded features that did not have any values (either a prediction of 0 or no result at all), that were already represented by other features by having one-to-one correlations, or that did not have a predictability higher than 0.65 (.503–.646). After the exclusion of these features, a total of 93 features remained.

An analysis using the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO=.595) and Bartlett's Test of Sphericity indicated the validity of PCA to interpret the data set ($p < .05$).

A PCA, using oblique Promax rotation, was conducted on the remaining 93 features, and 28 components were extracted. Using the components, we evaluated the performance on genre classification using an 18×15 *softmax* neural network with a linear activation function. As the number of texts belonging to each genre in SUC is uneven, we sample the data as a tensor, Batches × Samples × Components (where a batch is a 10×6 matrix of sampled measures of SUC texts). Genre **G** was excluded due to a low number of data points.

We obtained components by quantitatively analysing correlation between features and removing features such that we obtain maximal classification. The correlation cut-off was $|0.8|$ where we found local optimum of classification rate 84.0%. Examples of components are given in Table 8.2.

The variables chosen for each component had a magnitude over 0.3 and under -0.3. The 28 components explained 60.5% of the total variance explained by the 28 components. The first component explained 8% on its own.

Results on Genre Classification

The genre classification task aimed to explore whether the components extracted in the PCA had any discriminating power. The results of the classification are presented in Table 8.3.

There are some noteworthy results. The genre **B**, Press Editorial, had a significantly lower F1 score than the other genres. This could be due to the fact that we included other genres related to *Press* (**A** and **C**), and it

Table 8.2: Example of extracted components.

Comp.	Feature	Weight	Explanation
1	pos_PN	.816	Pronouns
	pos_NN	-.808	Nouns
	nrValue	-.807	Nominal ratio
	avgNoSyllables	-.730	Average number of syllables
	dep_PA	-.729	Complement of preposition
	dep_ET	-.714	Other nominal post-modifier
	dep_MS	.612	Macrosyntagm
	ratioSweVocC	.607	SweVoc lemmas fundamental for communication
	dep_IO	.573	Indirect object
	pos_AB	.572	Adverb
	dep_SS	.525	Other subject
	dep_DT	-.524	Determiner
	avgPrepComp	-.522	Average number of prepositional complements per sentence in the document
	pos_PS	.487	Possessive pronoun
	dep_NA	.473	Negation adverbial
	dep_MA	.446	Attitude adverbial
	dep_I	.425	Question mark
	pos_RG	-.407	Cardinal number
	dep_AA	.400	Other adverbial
	dep_.F	.388	Coordination at main clause level
	dep_PL	.382	Verb particle
	dep_OO	.365	Direct object
	pos_HA	.322	WH-adverb
	dep_AT	-.302	Nominal (adjectival) pre-modifier
	ratioSweVocTotal	.301	Unique, per lemma, SweVoc words in the sentence.
2	pos_PM	-.858	Proper noun
	dep_HD	-.788	Head
	lexicalDensity	.710	Lexical density
	ratioSweVocTotal	.706	Unique, per lemma, SweVoc words in the sentence.
	ratioSweVocH	.573	SweVoc other highly frequent lemmas (category H)
	ratioSweVocC	.544	SweVoc lemmas fundamental for communication
	dep_SS	.429	Other subject
	dep_AN	-.393	Apposition
	ratioSweVocD	.356	SweVoc lemmas for everyday use (category D)
	ratioVerbalRoots	.347	The ratio of sentences with a verbal root
	pos_NN	.332	Noun

Table 8.3: F1 Scores for the components

Genre	F1
Press Reportage (**A**)	0.814
Press Editorial (**B**)	0.793
Press Review (**C**)	0.831
Learning/Scientific Writing (**J**)	0.826
Imaginative Prose (**K**)	0.9324

is conceivable that these genres are relatively similar textually. Genre **K**, Imaginative Prose, had a significantly higher F1 score compared to the other genres, which had consistent scores. This indicates that this genre differs linguistically from the other genres, which makes it easier to discriminate in a text classification task.

Table 8.4: Confusion matrix for the classified genres. Each genre has been classified 150 times.

	A	B	C	J	K
A	120	6	9	8	7
B	11	111	8	15	5
C	8	4	125	9	4
J	4	8	7	128	3
K	2	1	2	0	145

The confusion matrix, presented in Table 8.4, reveals that genres **A, B, C, J** had many False Positives (FP) and many False Negatives (FN). Genre **K**, however, only has many FN. This indicates that the other genres are falsely classified as genre **K**, but that genre **K** is seldom misclassified as any other genre.

Interpretation of Results

In this study, we stated that it is possible to group text complexity features into meaningful components. The components were evaluated in a text classification task, and the results suggested that they can be used to correctly classify genres in a corpus of Swedish standard texts. The aim of this study was not to interpret the feature clusters linguistically, but rather to explore the discriminative power of the components.

The PCA components are difficult to interpret functionally, as the main purpose is to reduce the number of features. A previous study (Santini and Jönsson, 2020) conducted a factor analysis, as applied in Biber's Multi-Dimensional Analysis (MDA) framework (Biber, 1988), investigating whether it was realisable to isolate *facets* (as opposed to components) of text complexity across the different registers[2] of SUC. In MDA, a factor can be described as a textual dimension informative of co-occurrences of certain linguistic features in text and a text complexity facet, as defined by Santini and Jönsson (2020), is one side of that factor, either positive (+) or negative (-). Each side comprises groups of features, and since the features in the respective side are normally mutually exclusive, Santini and Jönsson (2020) interpreted each side as a facet which could distinguish a text complexity facet. Both MDA and PCA are techniques for data reduction, but the MDA allows us to interpret the extracted factors functionally in a way that PCA does not.

Visualising Text Complexity Features

The results of the study described in Santini and Jönsson (2020) revealed that it is possible to extract facets of text complexity using MDA. The resulting facets were normalised and visualised in a bar chart. However, despite being rather informative, interpreting the bar chart was challenging, even for experts, and we were interested in whether we could visualise the various facets of text complexity that had been extracted in Santini and Jönsson (2020) in a more intuitive manner, using radar charts. Visualising information or data in charts or graphs enhances the interpretation and understanding of trends and patterns in data. A clear visualisation of a text's characteristics could be useful for poor readers or for teachers and writers of Easy Language texts. The visualisation can be used as a guide to choose the best-suiting type of text to read for a given target audience.

A radar chart visualises multivariate data, and as each variable is dedicated to an axis (using the same scale), the radar chart presents the data in a polygonal shape. Radar charts have been used previously for visualising linguistic data (see Branco, Rodrigues, Costa, Silva, and Vaz (2014) and Egbert and Biber (2018)).

In this experiment, we interpreted the text complexity facets by visualising them, and hypothesised that the shapes of the visualisation would be informative of text characteristics. Our intuition was that different types of texts would exhibit different types of patterns, so that it is possible to get a visual understanding of the complexity of a certain register.

We used radar charts to profile the registers of the SUC corpus with five text complexity facets and with readability levels. The idea was that the radar chart would provide an overview of how the text complexity features of a text complexity facet were distributed in the register.

We noted that the shapes of the reviews (Figure 8.1), scientific writing (Figure 8.2) and reportage (Figure 8.3) were very similar, having a strong nominal facet associated with a pronounced appositional facet. The

[2] In this chapter, we use the terms *genre* and *register* interchangeably.

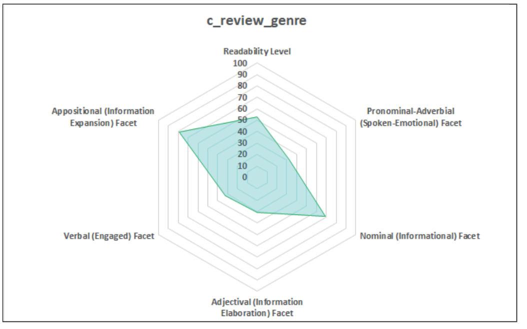


Figure 8.1: Review

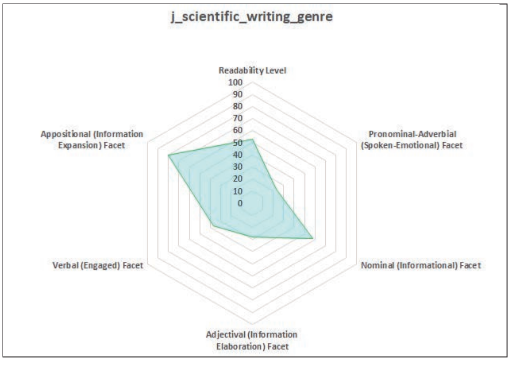


Figure 8.2: Scientific writing

pronominal-adverbial facet is very flat, and the verbal and adjectival facets are weak.

The shapes of bio-essay and imaginative prose are similar (see Figures 8.4 and 8.5). The bio-essay and imaginative prose registers are characterised by strong pronominal-adverbial, adjectival and appositional facets.

The hobby and miscellaneous registers are illustrated in Figures 8.6 and 8.7. These registers are strong on the nominal-appositional facet, but also characterised by some prominence of the verbal facet, while the pronominal-adverbial facet and the adjectival facet are rather flat.

The editorial and popular lore registers are depicted in Figures 8.8 and 8.9). The shapes are not similar to other registers in the SUC corpus. In the editorial register, we note a strong nominal facet, but a rather weak appositional facet. The texts in the editorial register exhibit

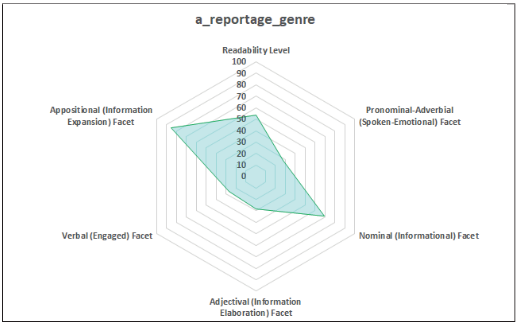


Figure 8.3: Reportage

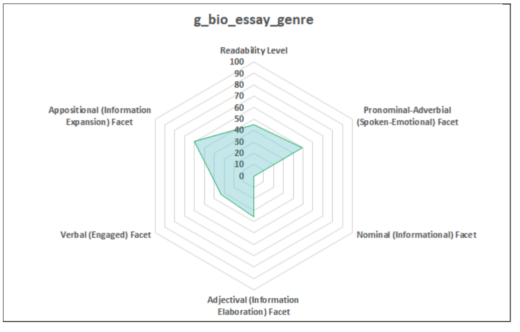


Figure 8.4: Bio-Essay

a pronounced verbal facet that could imply more complex syntax. The adjectival facet is weak, as is the pronominal-adverbial facet.

In summary, the visualisation of text complexity with radar charts highlights the linguistic similarities and dissimilarities of different registers given by the combination of various text complexity facets. Visualising text complexity in this way could serve several purposes. For instance, it can be used to guide readers of different audiences in choosing the type of text that best suits them. The visualisation could also be useful for producers of Easy Language texts writing for a specific reader audience, or as an aid for teachers searching for texts material at an appropriate complexity level for their students. For instance, individuals with aphasia may struggle with long sequences of adjectives. This group of readers may then use the visualisation to decide to read editorials (with a low frequency of adjectives) rather than imaginative prose (with a high frequency of adjectives).

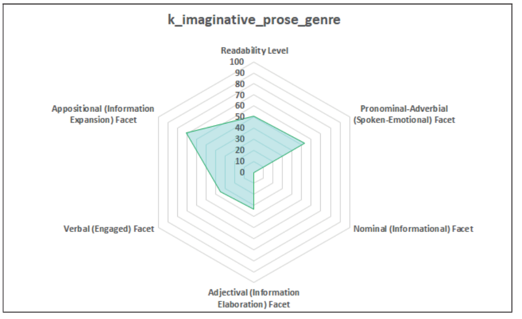


Figure 8.5: Imaginative prose

A Note on the Design and Usability of Complexity Visualisation

Producers of Easy Language texts could benefit from visualisation of text complexity. Such visualisation could provide instant feedback on the texts, and makes it possible to tailor the text to certain text characteristics. In order to create visualisation techniques that are useful for producers of Easy Language texts, we must know how such visualisations are interpreted by them.

In a series of design workshops, we invited web editors to give feedback on the design and usability for the purpose of communicating text complexity features. In the workshops, we presented two different design concepts (radar charts and bar charts), and discussed issues on text complexity in itself, as well as visualisation of text complexity.

The workshops resulted in input in how the web editors would like text complexity to be presented and visualised. For instance, using numbers to indicate complexity should be avoided. Instead, it might be better to display complexity levels relative to a gold standard, indicating levels both above and below it. Another opinion was that more detailed features should be presented in a separate and optional view, and that the overall focus should be on grouped complexity categories, as the presentation of specific complexity features causes confusion. It was also considered important to include genre and reader audience when designing the diagrams.

Regarding the visualisation, it was lifted that features could be better emphasised by using colours for indicating complexity level. On the issue on bar charts versus radar charts, the participants considered the first version of the radar charts to be detailed and informative, but more difficult to interpret. The simplicity of the bar charts was generally appreciated, but since they contain less information (when compared to a radar chart), they were also considered rather blunt and unspecific.

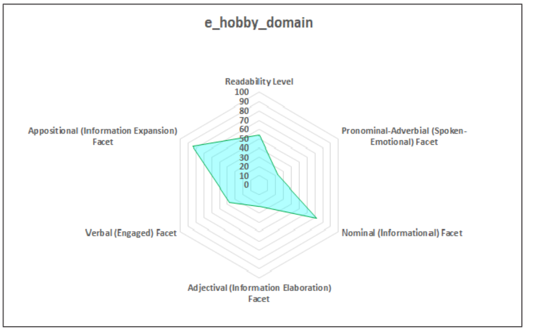


Figure 8.6: Hobby

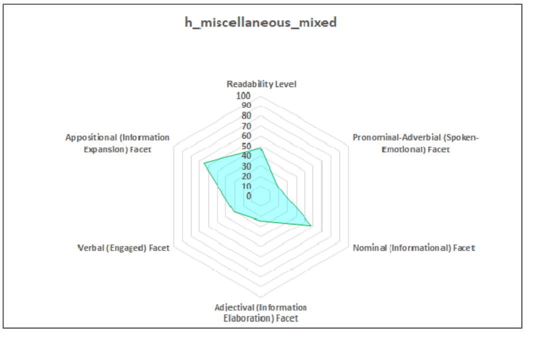


Figure 8.7: Miscellaneous

Based on the workshops, new design suggestions were developed, that combined the two visualisations and used the best qualities from both. The bar chart seems to have been preferred partly because of the use of colours and less explicit features, while the radar charts are compact and informative. Thus, we suggested to present specific features in an optional detailed view. The final visualisation combines a bar chart for overall complexity with a radar diagram giving a more refined picture on demand.

8.3 Chapter Summary

The purpose of the experiments presented in this chapter was to find methods that could help the interpretation of a large number of text complexity features. We showed that clustering and visualisation techniques are useful aids to make sense of this type of features.

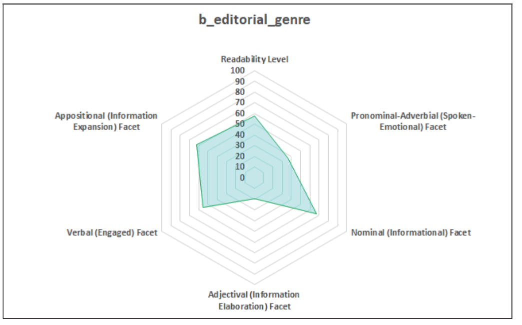


Figure 8.8: Editorial

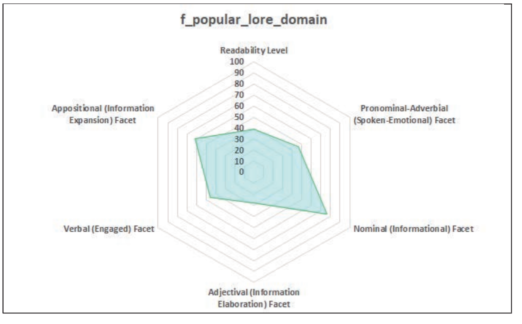


Figure 8.9: Popular lore

We investigated a component-based text complexity analysis method, and showed that it can be used to classify texts in genres. Since genres differ in terms of text characteristics, the resulting components may also be indicative of text complexity. As the complexity features in themselves, and the factors resulting from the PCA and MDA analyses, are difficult to interpret, we turned to visualisation of text complexity by the use of radar charts.

Visualisation through the use of radar charts creates shapes which could be used as an aid for quick interpretation of a text's overall complexity, and we saw that linguistically similar texts had similar complexity shapes. Visualising text complexity in this way could be useful to guide the readers of different audiences, but also for producers of Easy Language texts, or as an aid for teachers searching for texts material at the appropriate complexity level.

Although the work described in this chapter is based solely on Swedish, it is transferable to other languages. Some of the complexity features are probably language-dependent, but given a language-specific set of features, we believe that the methods for clustering and visualisation can be applied just as well to analyses in other languages.

CHAPTER 9

Corpora for Swedish Automatic Text Adaptation

In this chapter, our work on creating new resources for Swedish text is outlined through a description of the corpus collection procedure and the work on aligning the corpora at the sentence level. The insights of our work are summarised and discussed.

9.1 Creating a Comparable Swedish Corpora

Even if it is possible to learn from a corpus of Easy Language texts written by professionals, data-driven adaptation techniques require aligned resources. The automatic alignment of corpora refers to the process of linking text fragments. This is typically done for translation corpora of different languages, but many other modern natural language processing techniques require large corpora of aligned monolingual material, including automatic text simplification and adaptation. Alignment can be done on different levels. Typically, alignment is done at the sentence-level, but it is also possible to align smaller units (such as words) and larger units (such as paragraphs or whole texts).

From parallel material, it is possible to automatically extract patterns or rules in an automatic manner, or use the corpus as training data for language models. Since there was no such aligned resource for Swedish, we wanted to create such a resource from existing corpora.

To create an aligned resource for automatic text adaptation, we used two corpora: the Stockholm-Umeå Corpus (SUC) (Ejerhed, Källgren, and Brodda,

2006) and LäSBarT (Mühlenbock, 2008; Heimann Mühlenbock, 2013). SUC is a balanced corpus comprising one million words of Swedish texts written in the 1990's. While SUC is intended to represent the full spectra of Swedish texts, LäSBarT comprises Easy Language texts of different genres and domains. However, as the corpora were not parallel or comparable in content, the alignment task was challenging. The alignment method we chose was originally developed for test reuse detection (Sanchez-Perez, Sidorov, and Gelbukh, 2014).

Algorithm

For the development of alignment algorithm, we followed the procedure described in Sanchez-Perez, Sidorov, and Gelbukh (2014). The algorithm used a modified version if the tf-idf vector space model. The similarity between text fragments[1] was calculated with a slight modification of the original implementation: each sentence was considered a "document", and the full collection of sentences in the original document was considered the "document collection". Thus, rather than an inverse document frequency measure, we calculated an inverse *sentence* frequency. The aim of this procedure was to construct a monolingual corpus consisting of reference fragments (hereafter: **RS**) aligned with fragments of Easy Language (hereafter: **ES**).

The similarity scores used for aligning text fragments were cosine and the Dice coefficient. The cosine similarity measure is given by Equation 9.1 and calculates the cosine angle between two non-zero n-dimensional vectors, as the dot product of two vectors normalised by the vector lengths. The cosine similarity measure ranges between -1 and 1, where a value closer to 1 indicates a high similarity.

$$cos(A,B) = \frac{A \cdot B}{||A|| \cdot ||B||} \tag{9.1}$$

The **Dice coefficient**, given by Equation 9.2, is defined as the product of 2 times the number of features (i.e. lemmas or part-of-speech tags) in common by the sum of the length of the reference fragment and the length of the Easy Language fragment. The Dice coefficient ranges from 0 to 1 and represents the similarity of two text fragments, where a value closer to 1 means a higher similarity between the fragments.

$$Dice(A,B) = \frac{2\,|A \cap B|}{|A| + |B|} \tag{9.2}$$

Both similarity scores were calculated for each pair of text fragments, and text fragments were aligned if the score exceeded a certain threshold, originally 0.33 (following (Sanchez-Perez, Sidorov, and Gelbukh, 2014)).

[1] We use the term *text fragment* instead of *sentence* since we did not limit the algorithm to only sentences

We created and tested two conditions. The first method (hereafter: **M1**) was a replication of the algorithm proposed by Sanchez-Perez, Sidorov, and Gelbukh (2014). The vectors in **M1** were based on lemmatised words. The second method (hereafter: **M2**) worked in a similar fashion, but also included part-of-speech tags, in addition to word lemmas. We hypothesised that the inclusion of part-of-speech tags would improve the precision for word disambiguation and give a better result.

Evaluation

We assessed the performance in two ways. First, we calculated a number of text complexity metrics on the collected pairs of text fragments. This was done in order to get an idea of how well the algorithm performed in extracting more readable text fragments. Second, we performed a human evaluation through crowdsourcing.

For the evaluation, we selected RS that were paired with at least one ES at each cosine value (0.40, 0.50, 0.60, 0.70, 0.80) rounded to 2 decimals.

Text Complexity Measures

The text complexity measures used were:

- **LIX**, *readability index* (Björnsson, 1968), further described in Chapter 6, Section 6.2. LIX was calculated for the collection of Easy Language text fragments and reference text fragments that were produced by the alignment process, as well as for the original corpora.
- **OVIX**, *word variation index* (Hultman and Westman, 1977), further described in Chapter 6, Section 6.2. OVIX was computed by treating the collection of aligned fragments that originated from SUC and the aligned text fragments that originated from LäSBarT as separate corpora.
- **Length of words and fragments**:
  - **Average word length** defined as the average number of characters per word
  - **Average number of long words per fragment** defined as words longer than six characters
  - **Average number of words per text fragment**
- **N-gram overlap** describes the shared *n*-grams between pairs of text fragments. We treated each paired fragment (one RS paired to one ES) as a collection of *n*-grams, with *n* ranging from 1 to 4. We computed overlapping *n*-grams in ES and RS and divided the result by the total number of *n*-grams in RS. The resulting value ranges from 0 to 1, where

1 indicates that the ES is either an exact copy of, or is contained within, the RS.

Crowdsourcing

Participants were recruited via public postings on social media and e-mails to undergraduate students at Linköping University. Similarity judgements were made on randomly ordered aligned pairs, following the annotation scale proposed in the Cross-Level Semantic Similarity Task of SemEval 2014 (Nakov and Zesch, 2014), translated into Swedish:

4. *The two items have very similar meanings and the most important ideas, concepts, or actions in the larger text are represented in the smaller text.*

3. *The two items share many of the same important ideas, concepts, or actions, but those expressed in the smaller text are similar but not identical to the most important in the larger text.*

2. *The two items have dissimilar meaning, but the shared concepts, ideas, and actions in the smaller text are related (but not similar) to those of the large text.*

1. *The two items describe dissimilar concepts, ideas and actions, but might be likely to be found together in a longer document on the same topic.*

0. *The two items do not mean the same thing and are not on the same topic.*

There was no time restriction for the individual tasks, nor for the evaluation as a whole. The participants could continue the annotation task until all aligned items had been given a ranking, or until the participant left the annotation web page.

Results

A comparison of the overall results is presented in Table 9.1. M2 resulted in a higher arithmetic mean for both cosine and Dice scores, and produced $18,115$ more aligned clusters than M1. M2 contained about eight million more aligned text fragments than M1. Regarding the *n*-gram overlap, M1 had a higher unigram overlap (0.31) than M2 (0.18). This indicates that for M1, the easy fragments are more similar to the reference fragments than they are for M2.

Table 9.2 presents shallow features of the original corpora, and the corresponding values of the aligned total (RS_{total} and ES_{total}), and the subset later evaluated by crowdsourcing (RS_{subset} and ES_{subset}).

RS_{total} and ES_{total} contained more tokens and sentences compared to the original corpora. This is due to the fact that the same source text fragment

Table 9.1: Comparison of M1 and M2 regarding descriptive features.

Descriptive feature	**M1**	**M2**
Arithmetic mean Dice	0.47	0.50
Arithmetic mean cosine	0.49	0.50
Std. deviation, Dice	0.13	0.13
Std. deviation, cosine	0.09	0.10
Total number of aligned clusters	113,993	132,108
Total number of aligned text fragments	9,294,015	17,422,338
Unigram overlap	0.31	0.18
Bigram overlap	0.05	0.04
Trigram overlap	0.01	0.01
Quadrigram overlap	0.00	0.00

could be aligned several times. In both methods, the average sentence length and the average word length exhibited a tendency toward convergence for RS and ES. This suggests that the extracted fragments might be the simplest fragments of the original corpora. For both methods, we saw that the LIX and OVIX scores were lower for the ES and RS when compared to the corresponding original corpora. A tendency toward convergence could be observed also for the readability formulas.

Table 9.2: Feature values of corpora, alignments and subsets for M1 and M2.

M1	SUC	LäSBarT	RS_{total}	ES_{total}	RS_{subset}	ES_{subset}
Tot. no of tokens	1,048,657	1,142,666	5,378,071	344,951,473	181	1,308
% of long words	26.55	18.03	0.09	0.14	0.05	0.02
No. of unique tokens	106,853	49,776	17,362	26,889	30	179
LIX	41.97	27.46	18.4	24.22	-*	-*
OVIX	90.9	68.83	49.22	50.04	31.93	29.33
No. of sentences	68,038	121,212	617,167	33,247,883	19	95
Avg. sentence length	15.41	9.43	10.37	8.71	6.20	6.77
Avg. word length	5.21	4.58	4.10	3.93	3.9	3.33
M2	SUC	LäSBarT	RS_{total}	ES_{total}	RS_{subset}	ES_{subset}
Tot. no. of tokens	1,048,657	1,142,666	2,455,941	419,768,883	252	1,377
% of long words	26.55	18.03	0.09	0.15	0.007	0.05
No. of unique tokens	106,853	49,776	15,369	21,871	50	221
LIX	41.97	27.46	24.93	16.67	-*	-*
OVIX	90.9	68.83	49.66	49.23	21.55	32.03
No. of sentences	68,038	121,212	314,399	41,658,621	25	125
Avg. sentence length	15.40	9.43	7.80	10.07	4.30	6.25
Avg. word length	5.21	4.58	3.50	3.88	3.30	3.80

* LIX is applicable only to documents rather than unique sentences

Evaluation Results

220 aligned items were selected for the crowdsourcing evaluation. Each item contained one RS and one ES. 95 of the aligned items were from M1 and the remaining 125 were from M2.

Table 9.3 presents some examples of selected items from each alignment method with corresponding Dice and cosine measures. The examples are also presented with ranks that are typical of how they were rated by annotators.

Table 9.3: Example sentences per rank category for sentences aligned by M1 and M2.

M1 Rank	RS	ES	Cosine	Dice
0	Jag vet inte, sa jag. *I don't know, I said*	Jo, jag vet. *Yes, I know.*	0.50	0.57
1	Ja, det är du, säger Oscar. *Yes, you are, says Oscar.*	– Ja, det säger alla! *– Yes, everybody says that!*	0.60	0.60
2	Vad är det med dig? *What's up with you?*	– Vad är det som har hänt? *– What has happened?*	0.40	0.43
3	Vad är det? *What is it?*	–Vad skulle det vara? *– What would that be?*	0.80	0.86
4	Jag vet inte, sa jag. *I don't know, I said.*	Jag vet inte. *I don't know.*	0.70	0.55
M2				
0	Majsan rycker på axlarna. *Majsan shrugs.*	Tanne ryckte tag i henne. *Tanne grabbed her.*	0.40	0.36
1	Vad menar du? *What do you mean?*	Vad hade han gjort dig? *What did he do to you?*	0.60	0.60
2	undrar jag. *I wonder*	undrade hon *she wondered*	0.80	0.67
3	Varför då? *Why?*	Jaså, varför det? *Oh, really? Why?*	0.50	0.44
4	Det visste jag. *I knew that.*	Jag visste det, fan jag visste det. *I knew it, damn, i knew it.*	0.80	0.80

The comparison of cosine similarity and participant ranking is presented in Figure 9.1, with cosine values divided into intervals. There was a tendency for M2 to score consistently higher than M1.

The comparison of Dice similarity and participant ranking is presented in Figure 9.2, with Dice values divided into intervals. The somewhat peculiar look of the first and last intervals is explained by skewed data (not enough data points on these intervals), so these intervals can be considered outliers in this context. From interval 0.31–0.40 to interval 0.71–0.80, the Dice score seemed to stabilise for values over 0.5. These results suggest that M2 provides a more stable relationship between the human ranked categories and the similarity measures.

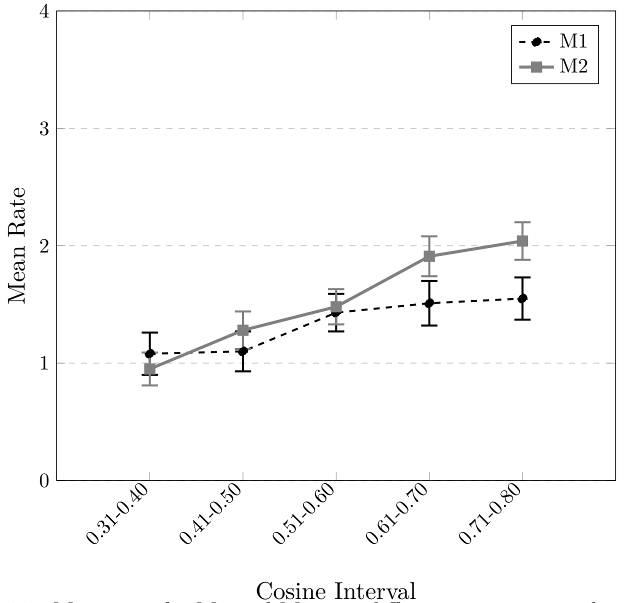


Figure 9.1: Mean rate for M1 and M2 over different cosine intervals

Summary of Results

In this work, we hypothesised that we could extend the work of Sanchez-Perez, Sidorov, and Gelbukh (2014) to create an aligned resource to be used for text simplification and adaptation purposes, by expanding the method to include part-of-speech tags.

We found that M2 seemed to relate better to human ranking than M1. The relationships between the annotated rankings and the cosine and Dice scores showed trends which we believed can be used to accept a cosine higher than 0.33 as the threshold for deciding which fragments to treat as candidate paraphrases.

As can be seen in the examples in Table 9.3, the Easy Language fragments seemed to be both syntactically and lexically similar to the reference fragments. Ideally, the resulting resource should contain paraphrases that are lexically different, but semantically equivalent.

9.2 Collecting a Web Corpus of Easy Language Text

Our work on aligning LäSBarT and SUC resulted in a large resource, but we saw that the aligned items were very similar and did not fully suit our

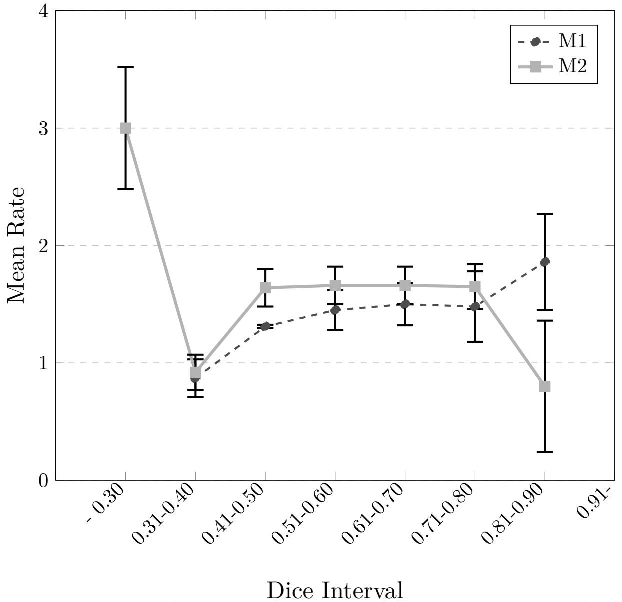


Figure 9.2: Mean rate for M1 and M2 over different Dice intervals

purposes. In order to create a more useful resource, we wanted to use texts which were comparable in content, treating the same topic, but written in both Easy Language and standard language.

In Sweden, the websites of public authorities and municipalities have Easy Language versions and the texts are often based on guidelines for writing Easy Language, such as the guidelines given by the Swedish Agency for Accessible Media (MTM, 2021). A more thorough description of such guidelines can be found in Chapter 7.

Thus, scraping these websites for Easy Language texts and the standard language counterparts could be a way to effectively construct a Swedish resource for automatic text simplification and adaptation.

Scraping the Web

In order to create a resource from the web texts of Swedish public authorities and municipalities, we first collected all the URLs of the websites. Websites that did not contain pages that were explicitly described as being written in easy Swedish were removed from the list. Other filtering operations included removing duplicate or erroneous URLs, as well as URLs that were automatically identified as spam.

To collect the relevant texts, a web crawler was built in Python 2.7.0 with Scrapy[2], an open-source web scraping framework. The crawler was given two lists: one list containing all the start URLs of the easy Swedish versions of the websites, and one list containing the start URLs of the standard versions of the website. The *simple* pages were collected in the first run, and in the second run, all other pages on the same domain were collected. The crawler was instructed to respect the robots.txt exclusion directives, and to limit the request frequency to 1 request per 4 seconds.

The final number of documents downloaded was 138,130, of which 1,629 consisted of pages with text written in easy Swedish.

Alignment

To align the corpus at the sentence level, three different algorithms were tested, broadly following (Kajiwara and Komachi, 2016). The alignment algorithms, originally proposed by (Song and Roth, 2015) were Average Alignment (AA), Maximum Alignment (MA), and Hungarian Alignment (HA). All algorithms use similarity of word embeddings to create a sentence similarity score that is used to find sentences that are semantically similar.

- The AA algorithm uses the average of the pairwise word similarity of all words of a pair of sentences to create a sentence similarity score.
- The MA algorithm only use the word pairs that maximise the word similarity score of all words of the given pair of sentences, and the sentence similarity is then calculated by summing the word similarity scores.
- The HA algorithm calculates the lowest cost (thus, the highest cosine value) for every possible word pair, and normalises the resulting sum by the length of the shortest sentence in the sentence pair, resulting in a sentence similarity score.

In order to make the alignment methods work for Swedish, a few adjustments were made to the original implementations. Firstly, since we were now working on Swedish, new word embeddings were needed. For this implementation, we used Swectors (Fallgren, Segeblad, and Kuhlmann, 2016). However, when switching to Swedish word vectors, another problem emerged. Since the original implementation used word embeddings trained on a very large corpus, there was no real need for handling Out-Of-Vocabulary (OOV) words, so they were simply ignored. Since the size of the vocabulary of the Swedish vector set was much smaller, the OOV words needed to be addressed.

To handle OOV words, Mimick (Pinter, Guthrie, and Eisenstein, 2017) was used to train a recurrent neural network at the character level, in order

[2]https://scrapy.org/

to predict OOV word vectors based on a word's spelling. The idea is that such word embeddings could identify characteristics of a word based on its spelling, which would result in more precise vector estimations when compared to other OOV handling strategies, such as using a completely randomised word vector.

There were two thresholds that could be altered for each algorithm: 1) The *word similarity threshold*, denoting the threshold for when a pair of words is considered to be similar enough, and 2) the *sentence similarity threshold*, denoting the threshold for when a pair of sentences is similar enough and thus, aligned.

Finding the Thresholds: Human Evaluation

In order to tune the sentence similarity threshold, we performed a web survey wherein aligned sentences were rated according to similarity. For this evaluation, the word threshold value was set to 0.49 following (Kajiwara and Komachi, 2016).

All three algorithms were used to generate aligned sentences from the collected corpus of web texts from Swedish public authorities and municipalities. In the evaluation, we randomly selected three sentence pairs per similarity interval (0.51–0.60, 0.61–0.70, 0.71–0.80, 0.81–0.90, 0.91–1.0), where a lower similarity interval indicated a low level of similarity, and vice versa. However, the AA algorithm did not produce sentence pairs represented at all similarity intervals, and due to this we chose to remove AA from the evaluation. For the other two algorithms, MA and HA, a total of 30 sentence pairs were extracted.

The participants rated the aligned sentence pairs from MA and HA on a four-point similarity scale. The grading was inspired by the categories previously used to create a manually annotated data set (Hwang, Hajishirzi, Ostendorf, and Wu, 2015), but with slight modifications: the categories were translated into Swedish and were slightly reformulated to suit non-expert participants. The categories were as follows:

1. **Meningarna handlar om helt olika saker**
   The sentences treat completely different things

2. **Meningarna handlar om olika saker men delar en kortare fras**
   The sentences treat different things, but share a shorter phrase

3. **En menings innehåll täcks helt av den andra meningen, men innehåller även ytterligare information**
   The content of a sentence is completely covered by the second sentence, but also contains additional information

4. **Meningarnas innehåll matchar helt, möjligtvis med små undantag (t. ex. pronomen, datum eller nummer)**
   The content of the sentences matches completely, possibly with minor exceptions (such as pronouns, dates or numbers)

61 participants responded to the survey, and the results of the human evaluation are presented in Table 9.4.

Table 9.4: Results of the human evaluation of MA and HA. Good=3, Good Partial=2, Partial=1 and Bad=0.

MA	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1.0
Mean	0.363	1.282	2.451	1.989	2.522
Std.Dev.	0.646	0.918	0.774	0.796	0.652
HA					
Mean	0.344	0.300	1.464	0.645	1.539
Std.Dev.	0.624	0.504	0.848	0.874	1.314

The human evaluation revealed that the MA algorithm produced sentence pairs that were generally considered more similar than the HA algorithm. More specifically, when looking at the MA algorithm, a sentence threshold over 0.71 seemed to produce sentence pairs that were considered similar. The HA algorithm, however, did not reach an average value above 2.

It should be noted, though, that the standard deviation was rather high through all intervals, which indicates that the results should be interpreted with some caution.

Finding the Thresholds: Gold Standard Evaluation

We supplemented the web survey with an additional method of finding the best parameter settings regarding both word and sentence thresholds for all three alignment algorithms (AA, MA, HA). The intended outcome was to find the best combination of parameters that maximised the F1 score of the algorithms.

First, we needed to identify a gold standard. To do this, we aligned sentences from automatically aligned document pairs from the corpus of Swedish public authorities and municipalities web texts, and let three annotators rate each sentence pair, broadly following the procedure in (Hwang, Hajishirzi, Ostendorf, and Wu, 2015). The annotators (one graduate student and two payed undergraduate students) were asked to rate each sentence pair according to the descriptions of the scale.

Since the document pairs were randomly distributed, the distribution of rated sentences was uneven, but for this evaluation, only sentences with exactly three annotations were included. This resulted in a set of 4,548 sentence pairs. Of these pairs, 4,457 were rated as *Bad*, 37 were rated as *Bad Partial*, 24 were rated as *Good Partial*, and 30 were rated as *Good*. The inter-annotator agreement was calculated using the Intra-class Correlation Coefficient (ICC), and revealed excellent agreement, $ICC(2,3) = 0.964$.

The results are given in two conditions: **GGPO** (the sentences rated as *Good* and *Good Partial* were considered correct alignments) and **GO** (only the sentences rated as *Good* were considered correct alignments).

Since the AA algorithm, again, produced a very low number of aligned sentences, it was discarded from analysis.

For the **GGPO** condition the results are presented in Table 9.5. The AA algorithm achieved a maximum score of $F1 = 0.034$, aligning 3 sentences (no difference was observed when changing parameters or vector conditions). The MA algorithm achieved a maximum score of $F1 = 0.758$, aligning 39 sentences (Mimick vectors, word similarity threshold of 0.39, sentence similarity threshold of 0.7). The HA algorithm achieved a maximum score of $F1 = 0.762$, aligning 49 sentences (Mimick vectors, word similarity threshold of 0.79, sentence similarity threshold of 0.7).

Table 9.5: The best-performing algorithm conditions in the GGPO setting.

	Max F1	No. sentences
AA	0.034	3
MA	0.758	39
HA	0.762	49

For the **GO** condition the results are presented in Table 9.6.

The AA algorithm achieved a maximum score of $F1 = 0.060$, aligning 2 sentences (Mimick vectors, word similarity threshold of ≥ 0.29 and sentence similarity threshold of ≥ 0.4). The MA algorithm achieved a maximum score of $F1 = 0.892$, aligning 33 sentences (Mimick vectors, word similarity threshold of ≥ 0.39 and sentence similarity threshold of 0.8). The HA algorithm achieved a maximum score of $F1 = 0.800$, aligning 38 sentences (Mimick vectors, word similarity threshold of ≥ 0.59 and sentence similarity threshold of 0.9).

Table 9.6: The best-performing algorithm conditions in the GO setting.

	Max F1	No. sentences
AA	0.060	2
MA	0.892	33
HA	0.800	38

From this evaluation, we have established that the conditions using Mimick for generating vectors for out-of-vocabulary words performed better in terms of precision, recall and number of aligned sentences. The best-performing algorithms were the MA in the GO setting, and HA in the GGPO setting.

The Sentence-aligned Corpus

Based on the two evaluations, we aligned the corpus of web texts from Swedish public authorities and municipalities using the MA algorithm, with a word similarity threshold of 0.39 and a sentence similarity threshold of 0.7. The aligned resource consisted of 45,671 sentence pairs. After removing duplicates, 15,433 sentence pairs remained. We were now interested in whether the sentence pairs in the aligned resource in fact differed in complexity, i.e. if we really had a corpus of standard and easy Swedish.

In a corpus-level analysis, we looked at frequency and ratio measures to get a general overview of the corpus. This included measures such as the total number of words, unique words, and ratio of long and extra-long words. Such corpus-level measures have been used for analysing corpora of texts in simple and standard Swedish (Heimann Mühlenbock, 2013). A few measures were excluded from analysis, since they are not applicable at the sentence level. The measures we excluded from the analysis were LIX (Björnsson, 1968), type-token ratio and OVIX (Hultman and Westman, 1977). Thus, the measures used for the corpus-level analysis were:

- **Total number of words**, calculated as the number of all alphanumeric word tokens in the subcorpus.
- **Number of unique words**, calculated as the number of all unique alphanumeric word tokens in the subcorpus.
- **Ratio of long words**, defined as the ratio of words longer than 6 characters to the total number of words in the subcorpus.
- **Ratio of extra-long words**, defined as the ratio of words longer than 13 characters to the total number of words in the subcorpus.

In a sentence-level surface analysis, we looked at complexity measures of each part of the corpus, i.e. all simple sentences and all standard sentences separately. Significance testing was performed using a two-tailed *t*-test. The measures considered for the sentence-level surface analysis were:

- **Word length (chars)**, calculated as the mean word length in number of characters. This value was calculated for each sentence, and then averaged over the entire subcorpus.
- **Word length (syll)**, calculated as the mean word length in number of syllables. For simplicity, we let the number of vowels correspond to the number of syllables. This value was calculated for each sentence, and then averaged over the entire subcorpus.
- **Sentence length (words)**, calculated as the number of tokens of a sentence. This value was calculated for each sentence, and then averaged over the entire subcorpus.
- **Number of long words**, defined as the number of words longer than 6 characters. This value was calculated for each sentence, and then averaged over the entire subcorpus.

- **Number of extra-long words**, defined as the number of words longer than 13 characters. This value was calculated for each sentence, and then averaged over the entire subcorpus.

In a third analysis, we calculated the measures of *SCREAM-sent*, a subset of a feature set used for text complexity classification (Falkenjack, Heimann Mühlenbock, and Jönsson, 2013). For an overview of text complexity measures in general and the SCREAM features in particular, see Chapter 6 in this thesis. The selected subset of measures was chosen specifically to be suitable for sentence-level analysis, and the selection was done according to (Falkenjack, 2018a).

To calculate the linguistic measures used for the *SCREAM-sent* analysis, we used a new version of SAPIS (Fahlborg and Rennes, 2016), an API service for text analysis and simplification. The version used for this analysis now uses EFSELAB[3] (Östling, 2018) for part-of-speech tagging.

The selected features were:

- **avg_dep_distance_dependent**, calculated as the average dependency distance in the document.
- **avg_n_syllables**, calculated as the average number of syllables per word in the document.
- **avg_prep_comp**, calculated as the average number of prepositional complements in the document.
- **avg_sentence_depth**, calculated as the average sentence depth.
- **avg_word_length**, calculated as the average word length in a document.
- **n_content_words**, calculated as the number of content words (nouns, verbs, adjectives and adverbs).
- **n_dependencies**, calculated as the number of dependencies.
- **n_lix_long_words**, calculated as the number of long words as defined by the LIX formula; words with more than 6 characters.
- **n_nominal_postmodifiers**, calculated as the number of nominal post-modifiers.
- **n_nominal_premodifiers**, calculated as the number of nominal premodifiers.
- **n_right_dependencies**, calculated as the number of right dependencies.

[3]https://github.com/robertostling/efselab

- **n_sub_clauses**, calculated as the number of sub-clauses.
- **Lemma frequencies**, derived from the basic Swedish vocabulary SweVoc (Heimann Mühlenbock and Johansson Kokkinakis, 2012a):
  - **n_swevoc_c**, calculated as the number of words that belong to the SweVoc C word list. SweVoc C contains lemmas that are fundamental for communication.
  - **n_swevoc_d**, calculated as the number of words that belong to the SweVoc D word list. SweVoc D contains lemmas for everyday use.
  - **n_swevoc_h**, calculated as the number of words that belong to the SweVoc H word list. SweVoc H contains other highly frequent lemmas.
  - **n_swevoc_s**, calculated as the number of words that belong to the SweVoc S word list. SweVoc S contains supplementary words from Swedish Base Vocabulary Pool.
  - **n_swevoc_total**, calculated as the number of words that belong to the total SweVoc word list. SweVoc Total contains SweVoc words of all categories.
- **n_syllables**, calculated as the number of syllables in the document.
- **n_tokens**, calculated as the number of tokens in the document.
- **n_unique_tokens**, calculated as the number of unique tokens in the document.
- **n_verbal_roots**, calculated as the number of sentences where the root is a verb.
- **n_verbs**, calculated as the number of verbs.
- **right_dependency_ratio**, calculated as the ratio of the number of right dependencies to the number of total dependencies.
- **sub_clause_ratio**, calculated as the ratio of sub-clauses to the total number of sub-clauses.
- **total_token_length**, calculated as the length of all tokens of a document.

The *SCREAM-sent* measures were calculated at the sentence level, all measures indicating an average should be regarded as absolute for a given sentence.

Thus, three sets of analyses were performed: one corpus-level surface analysis, and two sentence-level analyses. The significance testing was performed using two-tailed *t*-tests, assuming non-equal variances.

Results

The results of the corpus-level surface analysis are presented in Table 9.7. This analysis provides an overview of the sentence-aligned corpus, and shows that the subcorpus of simple sentences has a lower total number of words, whereas the subcorpus of standard sentences has a larger word variation (number of unique word tokens), and a slightly higher ratio of long and extra-long word tokens.

The results of the sentence-level surface analysis are presented in Table 9.8, and here we note a tendency towards shorter word lengths (in number of characters as well as number of syllables) in the subcorpus of simple sentences. We can also see that the subcorpus of simple sentences has a shorter sentence length and a lower number of long and extra-long words.

The results of the sentence-level analysis using the *SCREAM-sent* measures are presented in Table 9.9. Measures related to the length of the sentence, such as the number of syllables and the number of tokens, are higher in the *standard* sentences. There is also a significant difference in sentence depth and number of right dependencies, which could indicate higher complexity in the *standard* sentences. The *simple* sentences generally exhibit shorter token length, and fewer long words (>6 characters). We observed no difference regarding the SweVoc measures from category C (core vocabulary), D (words referring to everyday objects and actions, and H (highly frequent words). Statistically significant differences were observed for the SweVoc category S (supplementary words from the Swedish Base Vocabulary Pool), and SweVoc Total.

Table 9.7: Overview of the characteristics of the sentences in the simple part of the corpus (*simple*) and the standard part of the corpus (*standard*).

Measure	*simple*	*standard*
Total number of words	177,011	181,111
Number of unique words	10,373	11,593
Ratio of long words	22.55%	22.97%
Ratio of extra-long words	3.28%	3.44%

9.3 NyLLex: a Graded Easy Language Lexicon

It is well-established that one key element of Easy Language texts is to use *simple* words, and avoid *difficult* words. One way of gaining further knowledge about word complexity is to compile lexical resources of texts targeting poor readers. Such resources can provide an important source of information about the receptive vocabulary of the target audience, which by extension could

Table 9.8: Sentence-level surface analysis.

Measure	$\overline{X}_{simple}$	$\overline{X}_{standard}$	t	p
Word length (chars)	5.36	5.40	-3.03	*
Word length (syll)	1.93	1.95	-3.67	*
Sentence length (words)	11.47	11.74	-3.96	**
Number of long words	2.96	3.10	-5.66	**
Number of extra-long words	0.38	0.40	-3.47	**

* $p<0.05$, ** $p<0.001$

be used, for instance, in automatic text simplification applications, or for measuring text complexity. For instance, the classical Dale-Chall readability formula (Chall and Dale, 1995) utilises a list of simple words in its calculation, where the ratio of difficult words (i.e. words not occurring in the list) is used for estimating the readability of a text.

For Swedish, there are some lexical resources which could be used for this purpose. SweVoc (Heimann Mühlenbock and Johansson Kokkinakis, 2012b) is a Swedish base vocabulary consisting of about 7,600 Swedish lemmas, divided into sub-categories. SVALex (François, Volodina, Pilán, and Tack, 2016) is a lexicon of approximately 16,000 words originating from the COCTAILL corpus (Volodina, Pilán, Eide, and Heidarsson, 2014), a corpus of reading comprehension texts collected from coursebooks targeting second language learners of Swedish. Whereas SVALex contains the receptive vocabulary of second language learners, the SweLLex (Volodina, Pilán, Llozhi, Degryse, and François, 2016) lexicon is focusing on productive vocabulary. SweLLex is extracted from the SweLL corpus (Volodina, Pilán, Enström, Llozhi, Lundkvist, Sundberg, and Sandell, 2016), containing essays written by second language learners of Swedish. Both SVALex and SweLLex are annotated with the CEFR level classification indicating reading proficiency level.

We aimed to create an Easy Language lexical resource which complements the available resources for Swedish with regards to source material and target audience, and to validate the resource by examining its unique aspects and overlaps with similar resources.

Material

The source material consists of 247 books from *Nypon och Vilja förlag*, Sweden's largest publishing company for texts in Easy Language. Each book is classified according to the readability gradation presented in Table 9.10.

The books varied in genre, and included both fiction and non-fiction works. Some of them belong to the same series of books, for example, the *Gå till …* [Go to...] series, describing various everyday tasks (such as going to the library or the dentist), aimed mainly at second language learners; and the

Table 9.9: Results from the t-test comparing the sentences in the simple subcorpus (*simple*) with the sentences in the standard subcorpus (*standard*). The *n_lix_long_words* differs from the *Number of long words* in Table 9.8, since the former uses the lemma form in its calculation.

Measure	$\overline{X}_{simple}$	$\overline{X}_{standard}$	t	p
avg_dep_distance_dependent	2.44	2.46	-3.81	**
avg_n_syllables	1.80	1.81	-3.24	**
avg_prep_comp	1.46	1.51	-3.77	**
avg_sentence_depth	5.95	6.01	-2.63	*
avg_word_length	5.07	5.11	-2.91	*
n_content_words	6.64	6.77	-3.51	**
n_dependencies	13.26	13.58	-4.46	**
n_lix_long_words	2.41	2.56	-6.64	**
n_nominal_postmodifiers	0.85	0.90	-4.06	**
n_nominal_premodifiers	0.28	0.30	-3.48	**
n_right_dependencies	9.18	9.39	-4.19	**
n_sub_clauses	0.26	0.26	-0.78	
n_swevoc_c	5.38	5.46	-1.89	
n_swevoc_d	0.26	0.26	-0.02	
n_swevoc_h	0.79	0.80	-0.73	
n_swevoc_s	0.61	0.63	-2.28	*
n_swevoc_total	6.32	6.44	-2.40	*
n_syllables	21.02	21.73	-5.96	**
n_tokens	13.26	13.58	-4.46	**
n_unique_tokens	12.45	12.73	-4.64	**
n_verbal_roots	0.81	0.80	3.32	**
n_verbs	2.45	2.46	-0.45	
right_dependency_ratio	0.70	0.70	0.63	
sub_clause_ratio	0.25	0.26	-0.89	
total_token_length	62.80	65.00	-6.18	**

* $p<0.05$, ** $p<0.001$

Ett liv [A life] series of easy-to-read biographies adapted for different levels of education. Furthermore, there are also easy-to-read versions of classic novels like *Kallocain* by Karin Boye and *1984* by George Orwell, amongst others. All the books were provided in digital format as PDF files.

In Sweden, the LIX value is commonly used by writers of Easy Language texts. As seen in Figure 9.3, the LIX value does indeed increase in conjunction with the reading proficiency levels.

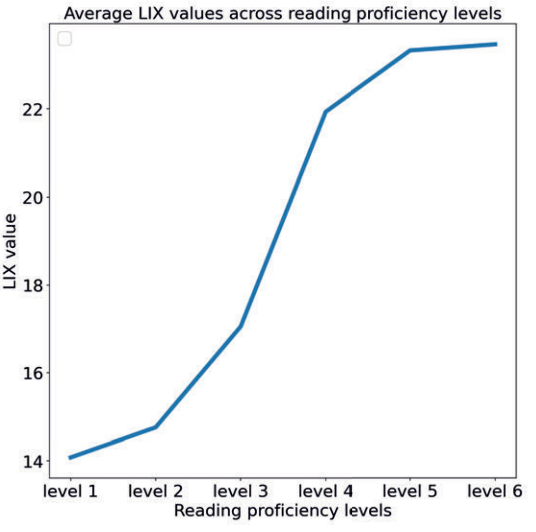


Figure 9.3: Average LIX values of the books in the different reading proficiency levels.

Pre-processing

Each book was parsed from PDF to plain text with a Python implementation of Tika Parser[4].

The plain texts were subsequently tokenised, part-of-speech tagged, and lemmatised in the EFSELAB pipeline[5]. Each word was kept in its lemma form together with its part-of-speech tag, meaning that a word that appears with two different part-of-speech tags has two separate entries in the resource. We excluded all entries that were tagged as punctuation (the SUC tags MAD, PAD, and MID) as well as all personal names (listed as the SUC tag PM). Entries tagged as an ordinal number (SUC tag RO) in the form similar to *2:a* [2nd], *3:e* [3rd] were manually merged to their base lemmas; *andra* [second] and *tredje* [third].

In line with similar resources, NyLLex includes multi-word expressions (MWEs). These were identified by matching *n*-grams in the source texts to MWE entries in the SALDO lexicon (Borin, Forsberg, and Lönngren, 2013).

[4] https://github.com/chrismattmann/tika-python
[5] https://github.com/robertostling/efselab

Table 9.10: The Nypon reading proficiency levels, interpreted and loosely translated by the authors. The complete descriptions of each level can be found at https://www.nyponochviljaforlag.se/om-oss/om-lattlast/lattlastnivaer/lattlastnivaer-nypon/
.

Level	Description
Level 1	Each page contains very little text. Simple words and sentences. Many illustrations that support the story.
Level 2	Every-day language. The text is divided into short paragraphs with short line lengths. The content depicts relatable situations and focuses on sequences of events. In the books targeting a younger audience, there are illustrations for each spread.
Level 3	Well-known words and expressions. The story is chronologically presented and the connection between cause and effect is clear. There is a sequence of events and descriptions of characters and environment. A number of illustrations. The graphical form includes larger font size, line spacing and margins.
Level 4	Chapter books with few or no illustrations. Slightly more difficult names, words and expressions and longer sentences. The graphical form is spacious with large font size.
Level 5	Adopts an easy-to-read focus regarding language, content and graphical form, but present a larger challenge to the reader.
Level 6	Books produced with special care regarding language, content and graphical form. Books at this level are supposed to be a gateway to traditional books.

Table 9.11: Number of books and tokens in the Nypon dataset. Before and after filtering.

	Books	Tokens	
		Before	After
Level 1	57	23,301	22,942
Level 2	47	83,990	82,998
Level 3	60	212,000	208,723
Level 4	64	362,289	352,595
Level 5	17	110,476	106,966
Level 6	2	22,573	22,007
Total	**247**	**814,629**	**796,231**

Filtering

The vocabulary of the full Nypon dataset consists of a total of $16,841$ entries spread across the six readability levels. $10,167$ (61.7%) entries were only found in one of the readability levels. Of these level-specific entries, $6,781$ were unique and only appeared once in the whole dataset, while 3386 could be found multiple times in the level. Many of the level-specific entries were a consequence of errors during the reading of the PDFs, even after the manual corrections of the raw texts produced by the PDF reader. There was also a substantial number of entries that represented a correct word, but where most of the usage stems from very specific contexts[6]. In order to get a more representative vocabulary, with fewer outliers and highly specific entries, we applied a filter to remove these kinds of entries. We tried filters that only kept entries that had a specified raw frequency count of N, but found that while this approach was fairly effective for pruning rare words, the entries that were faulty readings from the PDFs were not easily caught with a relatively low N. As mentioned before, the faulty readings mainly stemmed from graphical elements in the books. Since these elements tended to reoccur several times in the same book, the same kind of error was often repeated, resulting in multiple instances of the same faulty entry. The issue then became to find an N high enough to remove even repeated faulty entries, but low enough to not remove too many correct entries. However, since the exploration of the best value of N would require a manual evaluation of the entire resource for a wide range of Ns, we used the approach applied in Forsbom (2006), where a filter based on genre[7] contribution was applied. Since the goal was to keep as many lemmas as possible, we also set the threshold as low as possible, and filtered out entries that were not present in at least two of the six reading proficiency levels. This approach was found to strike a good balance in the pruning of both rare and faulty entries.

As seen in Table 9.11, even though the filter removed 62% of the unique entries, our resource still covers 98% of the total number of tokens present in the data set.

Manual Editing

Table 9.12: Overview of the final resource.

	Entries	MWEs	Avg. entry length	Rare entries
Level 1	1,881	72 (3.8%)	5.26	45 (2.4%)
Level 2	3,347	206 (6.2%)	5.53	151 (4.5%)
Level 3	5,120	315 (6.2%)	5.95	413 (8.1%)
Level 4	6,041	382 (6.3%)	6.14	556 (9.2%)
Level 5	4,381	250 (5.7%)	5.89	304 (6.9%)
Level 6	1,897	108 (5.6%)	5.61	73 (3.8%)
Total	**6,674**	**443 (6.6%)**	**6.20**	**771 (11.5%)**

Similarly to SVALex, the manual editing of NyLLex was carried out in a circular fashion that allowed for the frequency estimations to be recalculated

[6] For example; *kafferast* [coffee break], *debattera* [to debate], and *kakdeg* [cookie dough]

[7] In this work each reading proficiency level is seen as a genre

when a correction was made to an entry. We also took advantage of the fact that SVALex had undergone this process of manual editing before. As seen in Table 9.13, the total overlap of the NyLLex and SVALex is 4,544 entries, all of which could be seen as correct. Initially that left us with 2,214 entries present in NyLLex, but not in SVALex, all of which became the subject of our manual editing. In total, 83 of the 2214 entries were flagged as erroneous, and subsequently manually assigned to their correct lemma and word forms. The remaining 2,130 no-matches with SVALex were thus correct, but novel, NyLLex entries.

Frequency Estimation

Each vocabulary item of the resource is associated with a per-level frequency estimation. For the calculation of frequencies across each of the six readability levels, we followed the procedure described in SVALex (François, Volodina, Pilán, and Tack, 2016) using dispersed frequency. For calculating frequencies on the dataset as a whole, we used adjusted frequency, which has previously been used for ranking the words of the Swedish Base Vocabulary Pool (SBVP) (Forsbom, 2006).

Final Resource Description

The final resource totals 6,674 items (of which 443 are MWEs) distributed over six reading proficiency levels as defined by *Nypon and Vilja förlag*. Each item is not restricted to a single level, and since the filter method we applied (each entry must be present in at least two levels) removed all hapaxes, the rarest items in our resource are the items present only once in two levels respectively. Our resource includes 771 of these rare items, of which 50.2% are nouns (e.g. *kärlekshistora* [love story] and *ledtråd* [clue]). Their distribution across different reading proficiency levels is listed in the column *Rare entries* in Table 9.12. Conversely, 922 entries are present in all of the six levels. It comes as no surprise that the most frequent entries in this category are highly used words like *jag* [I], *vara* [be], and *och* [and].

As the well-established LIX metric indicates (see Figure 9.3), the complexity of the books increases in conjunction with the reading proficiency level. Another (simple, but fairly effective) metric for text complexity is the average word length of the text (see for example Falkenjack, Heimann Mühlenbock, and Jönsson (2013)), where longer average word lengths indicate a more complex text. This increase is however not clearly visible in the average entry lengths in NyLLex. Whereas level 1 is observed to be slightly lower, the rest of the levels follow no obvious pattern. The same trend can be seen with the number of MWEs present across different levels. While Volodina, Pilán, Llozhi, Degryse, and François (2016) found that the number of MWEs increased steadily with a higher CEFR level, the clearest pattern is that level 1 is the odd one out with a lower proportion of MWEs compared to the other levels. On the other hand, levels 2 through 6 can be seen to have a fairly similar MWE ratio. Why this happens is an interesting topic for further study, and we plan to investigate additional text complexity metrics on the same dataset.

We also compared NyLLex with the aforementioned similar resources for Swedish; SVALex, SweLLex, and SweVoc and the results of this comparison are presented in Table 9.13. It should be noted that NyLLex had a varying degree of overlap with the other resources. SVALex, which is the most extensive of the other resources, was found to have the highest overlap, with 4,544 identical entries. For the two comparatively smaller resources, SweLLex and SweVoc, the percent of overlaps was also lower. For SweVoc, 3,505 entries overlapped, while 3,169 entries were novel to NyLLex. For SweLLex, 2,733 entries overlapped, while 3,941 entries were novel. The resource size probably accounts for most of the difference in overlap percentages between SVALex and SweVoc and SweLLex. Additionally, the fact that SweLLex is a productive vocabulary sets it apart from the other resources, and can possibly explain some of the proportionally large overlap difference between SweLLex and SweVoc, even though they are similar in size.

Table 9.13: Comparison of the final resource to other resources.

	Total entries	Entry overlap	New entries
NyLLex	6,674	-	-
SVALex	15,686	4,544 (68.09%)	2,130 (31.91%)
SweLLex	6,967	2,733 (40.95%)	3,941 (59.05%)
SweVoc	7,408	3,505 (52.52%)	3,169 (47.48%)

Since SVALex was the most similar resource to NyLLex in terms of overlapping entries, and has similar structure with words annotated at different reading proficiency levels (CEFR), we did a more extensive comparison of the overlap of our resource and SVALex. This comparison is presented in Table 9.14. Level 1–6 correspond to the six Nypon reading proficiency levels, and A1–C1 correspond to the CEFR levels.

Each row displays the proportion of Level 1–6 entries also present in each of the CEFR levels of SVALex. Even though level 6 has slightly more entries represented than level 1, it only consists of two books (see Table 9.11) compared to the 57 books of level 1. It is therefore hard to draw any general conclusions about level 6. For the rest of the levels, although the degree of overlap between the resources is quite modest, there seems to be a slight trend toward a higher Nypon reading proficiency level also having a higher degree of overlap with a higher CEFR level.

Table 9.14: Overlap with CEFR levels in SVALex.

	A1	A2	B1	B2	C1
Level 1	20.5%	21.2%	15.5%	13.1%	13.0%
Level 2	16.2%	21.9%	19.3%	17.4%	16.7%
Level 3	13.3%	21.9%	22.6%	21.4%	20.0%
Level 4	12.0%	20.8%	23.8%	22.8%	21.4%
Level 5	14.0%	20.9%	22.0%	20.5%	20.0%
Level 6	16.2%	18.5%	15.8%	13.8%	14.2%

9.4 Chapter Summary

The decision to collect a new corpus of Swedish simple and standard texts and align it at the sentence-level was motivated by the lack of available parallel data for data-driven text simplification.

First, our intention was to collect a corpus of paraphrased fragments from corpora that were not comparable in content by using an algorithm previously used for plagiarism detection. We showed that such an algorithm can be used for creating parallel corpora. However, the resulting corpus was not ideal for the purpose of Easy Language Adaptation since the collected corpus contained fragments that were lexically very similar, which is not desirable for the task of extracting paraphrases. If the aligned fragments are too similar, it is less likely that useful information can be derived from them. Moreover, even though the corpora used were representing both Easy Language and the full spectra of Swedish texts, there is no guarantee that all *sentences* of an Easy Language text are simple, and vice versa. Thus, it is possible that the algorithms find the simplest sentences of the general corpus, since they probably are more similar to sentences of the Easy Language corpus. This is indicated by the fact that the resulting subcorpora are similar to each other according to OVIX (and to some extent also LIX), while scoring consistently lower than the original corpora.

Part of the solution to overcome such issues could be to use a dataset which we know contains texts treating the same topic, but using texts of different complexity levels. Since there was no such resource available at the time, we turned to the Easy Language versions of the websites for Swedish public authorities and municipalities.

Although there is no clearly defined target audience for the simple texts of Swedish public authorities and municipalities, they are typically written by professional writers with knowledge of simple writing, following the guidelines of MTM (MTM, 2021). Therefore, it seemed like a suitable resource that could be used to study how simple text is actually written. The sentences in the sentence-aligned corpus were extracted from this corpus of standard and simple documents, and therefore we assumed that the corpus comprise good representatives of standard and simple sentences.

The corpus of web texts from Swedish public authorities and municipalities was the first corpus of simple and standard sentences for Swedish. Unfortunately, the resulting corpus turned out to be limited in size and heavily unbalanced, and is not ideal as input for data-driven text simplification or adaptation. However, the corpus is useful for other purposes. For example, different versions of this corpus have been used in other projects. In Santini, Danielsson, and Jönsson (2019), a subset of the sentence-aligned corpus was one of the corpora used when comparing text categorisation feature representations, and Holmer and Jönsson (2020) used the same subset of sentence pairs as well as a version containing complete texts for training BERT models for classification.

Finally, we presented NyLLex, a novel lexical resource of words annotated with six reading proficiency levels. We believe that NyLLex could work as a complementary lexical resource, which could be used for further work in, for instance, applications for text complexity assessment or lexical simplification.

Part III

Automatic Text Adaptation for Target Audiences

CHAPTER 10

Automatic Text Simplification

The natural language processing field that is most directly concerned with adapting text to suit the needs of poor readers is Automatic Text Simplification (ATS). This chapter includes our work on automatic text simplification: developing a platform for rule-based syntactic simplification, and two methods for extracting comprehensible synonyms from text corpora.

10.1 Syntactic Simplification

For the implementation of syntactic simplification operations, we developed a simplification framework, StilLett. The framework was developed for experiments on various automatic text simplification techniques. This chapter accounts for a description of the original implementation of StilLett. The simplification module has been further developed, and is now built on a different platform.

Platform

The pre-processing module of the original StilLett comprised Stagger (Östling, 2013) for part-of-speech tagging, and two different versions of MaltParser (1.2, 1.7.2) (Nivre, Hall, Nilsson, Chanev, Eryigit, Kübler, Marinov, and Marsi, 2007), enabling use of both phrase-structure parsing and dependency parsing. Note that the rules and node operations described in this chapter modify phrase-structure trees.

The syntactic simplification module was partly built on a previous tool for automatic text simplification, CogFLUX (Rybing, Smith, and Silvervarg, 2010). Each operation type in CogFLUX is based on a target phrase (an expression of the linguistic structures that should be covered) and a modification phrase (an expression of what the targeted phrase should be replaced with). CogFLUX included two node operation types: REPL and DEL. In the REPL node operation, the modification phrase consisted of a replacement phrase structure, while the DEL operation removed the text matched by the target phrase.

Node Operation // [Target Phrase] → [Modification Phrase]

StilLett introduced new functionality which allowed for more extensive modification of the texts. We added new node operation types, new simplification rules, and extended the syntax for rule notation.

Node Operation Types Extension

The functionality of CogFLUX was further developed by the addition of new node operation types.

- SHIFT
  node operation for rearrangement of the sentence structure, which is supposed to be used for changing word order without losing grammaticality. The role of the target phrase is the same as for the REPL and DEL operations, but the modification phrase instead indicates which part of the targeted structure that should change position. Example:

  SHIFT//S-AVP VB NP → AVP NP &P(#)[1]

  This expression shifts the nodes AVP and NP resulting in the structure S-NP VB AVP.

- SPLIT
  node operation for splitting a sentence into two new sentences when a matching target phrase is found. This node operation splits the tree at the node given by the modification phrase, and the remaining two parts become two separate trees. Example:

  SPLIT//S-S KN S → KN &P(#)

  This expression splits the phrase at the conjunction, resulting in two separate phrases.

[1] &P denotes an additional demand on the target phrase. For instance, &P(S) demands that the parent node of the target phrase should be S. &P(#) means that there is no demand on the parent node.

- DROP
  node operation for removing nodes from the matched node. The node that is suggested for removal can be specified with part-of-speech tags, dependency tags, or a child of such a node. It is possible to delete one or several nodes below the matched node.

- ADD
  node operation for duplicating a specific node and adding it to a new phrase. The main usage is in the splitting operations, for instance for duplicating the subject of a phrase.

- COND
  node operation for adding restrictions to the child nodes of the matched nodes. If the COND restriction is not fulfilled, the rule is ignored.

- NOT
  node operation that excludes matches containing the matching nodes. For instance, NOT(?SN) suggests that there should not be any occurrence of the SN-tag within the ? node.

Syntax Extension

The simplification rules notation of CogFLUX followed the *X-rules* syntax, but in order to enhance functionality, we extended the syntax in three ways.

First, we included the possibility of using dependency tags in addition to the part-of-speech tags. The dependency tags were given within brackets to indicate the dependency relation a specific part-of-speech tag must have in a target phrase. For instance, NP(SS) would mean that the noun phrase must have a subject function.

Second, we added the *?* tag, which indicates one, many or no tags of any sort. The question mark is also used by the program and is handled as a node with all available node operations. Thus, the target phrase S-? NP VP would match a phrase such as S-AVP NP VP.

Third, we added the capability to provide a node index, which could serve as a limitation on the ? functionality. For instance, S-? NP_1 VP would limit the phrase to only match structures where the noun phrase is placed at index 1, meaning that only one node is allowed before the noun phrase.

Finally, we added functionality for adding child demands on nodes. The demands are given within curly brackets. For instance, NP{NN PP} states that the child nodes of the noun phrase must be nouns followed by a preposition phrase.

Rules

In the original version, the simplification rules proposed by Decker (2003) had been implemented. In StilLett, we extended the tool further by adding four new simplification rules.

1. Changing from passive to active voice

   To transform a sentence from passive to active voice in Swedish, relatively large transformations must be made. The subject of the passive sentence must become the object of the active sentence. The verb must be conjugated in the correct form, and if the sentence contains an agent, it must become the subject. In order to perform this, a sequence of operations was applied when a sentence in passive voice was detected. In addition to this, the agent must drop the preposition *av* (*Eng: by*).

   SHIFT//S-NP(SS) VB/VP ? PP(AG) → S-PP NP &P(#)

2. Quotation Inversion

   The quotation inversion changes the place of the speaker in a quotation, from *[quotation], said X* to *X said: [quotation]*) (Bott, Saggion, and Mille, 2012).

   The quotation inversion operation is triggered by quotation marks followed by specific words from a lexicon that might indicate a quotation (such as *said, exclaimed, whispered*, etc. and the quotation (specified by the quotation marks) switches place with the verb phrase and the noun phrase, such as:

   Example input: *"Go to bed!" said Kalle.*
   Example output: *Kalle said: "Go to bed!"*

3. Rearranging to straightforward word order

   This rule shifts the word order. In the example, a clause initiated with an adverb phrase or adjective phrase is captured and rearranged.

   SHIFT//S-AVP/AP VB/VP NP(SS) ? → S-NP(SS) AVP &P(#) (1)

   → S-AVP ? &P(#) (2)

   The application of this simplification operation might result in the following example:

   Example input: *Yesterday bought Kalle a new car.*[2]
   Example output: *Kalle bought a new car yesterday.*

[2] This word order is correct in Swedish.

4. Sentence split

   The SPLIT operation splits a sentence in two. For instance, a main clause comprising two main clauses joined by a conjunction is split at the conjunction to form two separate sentences. We also included functionality to split at appositions and split at relative clauses.

$$\text{SPLIT//S-S KN S} \rightarrow \text{KN \&P(\#)}$$

The performance of the initially implemented rule set was evaluated in Rennes (2015) and the results for each rule are given by Table 10.1. The sentence split rule was divided into three subtypes: split at conjunction (split -k), split at relative clause (split -r), and split at apposition (split -a).

Table 10.1: Performance of the first implemented rule set.

	Prox.	P2A	QI	SWO	SPLIT -k	SPLIT -r	SPLIT -a
Precision	0.979	0.890	1	0.848	0.537	0.732	0.111
Recall	0.960	0.564	0.650	0.683	0.254	0.449	0.026

The rule set was refined by Johansson (2017), and the evaluation results are presented in Table 10.2.

Table 10.2: Performance of the refined rule set.

	Prox.	P2A	QI	SWO	SPLIT -k	SPLIT -r	SPLIT -a
Precision	1	0.978	0.842	1	0.981	0.957	0.0286
Recall	1	1	0.842	0.859	0.922	0.658	0.857

A Note on the Syntactic Simplification

This work was conducted outside the scope of this thesis work (in the author's master's thesis (Rennes, 2015)), but the syntactic simplification module constitutes an important part of the complete adaptation systems to this date. StilLett was developed further within our projects and has now been completely reworked. The current version of StilLett is built on Python3 and makes use of a new pre-processor (see Chapter 12), which no longer includes any phrase-structure parser.

One benefit of using hand-crafted simplification rules is that it is possible to create relatively accurate simplifications, and any errors that get introduced are relatively transparent and easy to debug. Rule-based approaches can also serve as a good complement to data-driven methods, or as a complete solution when there is not enough data available. However, there are some drawbacks and challenges related to rule-based methods. One of the most prominent problems, is that it is very time-consuming to write rules manually. Moreover, the rules must be permissive enough to capture many complex constructions, but restrictive enough not to make many erroneous modifications.

While working with the development of the original implementation of StilLett, we found that the syntax complexity increased rapidly as more rules were developed. We discovered that the development of new rule types demanded far-reaching extensions, both to the script syntax as well as the node operations, and concluded that rule-based syntactic simplification systems are only profitable if there is a well-developed and dynamic platform.

10.2 Lexical Simplification

In addition to the syntactic simplification, StilLett includes the possibility of replacing words with simpler synonyms. The initial version of the synonym replacement module was based on the work of (Abrahamsson, 2011), using the Synlex lexicon (Kann, 2004) with an included frequency list. The Synlex lexicon includes 82,000 word pairs that are annotated with a *level of synonymity*. This score was calculated using ratings made by voluntary Internet users, who graded the synonym pairs based on how synonymous they were.

In addition to these strategies, we developed and evaluated other methods for finding more comprehensible synonyms. The first method, described in Section 10.2, was based on a corpus of texts in Swedish Easy Language, and the other method, presented in Section 10.2, was based on theories from the field of cognitive linguistics about basic-level words.

Extracting Synonyms from a Corpus of Simple Texts

We explored two novel methods to automatically extract Swedish synonyms from a corpus of texts in Swedish Easy Language.

This work was based on three assumptions. The first assumption was that synonyms can be extracted by finding semantically related words based on the distributional hypothesis. This means that words appearing in similar contexts often have similar meanings (Harris, 1970), and this can be used for finding synonymous words to a given input word. The second assumption was that words with overlapping translations are likely to be synonyms. This was in part inspired by Lin, Zhao, Qin, and Zhou (2003), who noted that translations of a word are often synonymous. The third assumption was that a corpus of Easy Language texts would contain more comprehensible words than corpora of standard texts.

The aim was to explore whether we could extract synonyms that were also considered more comprehensible.

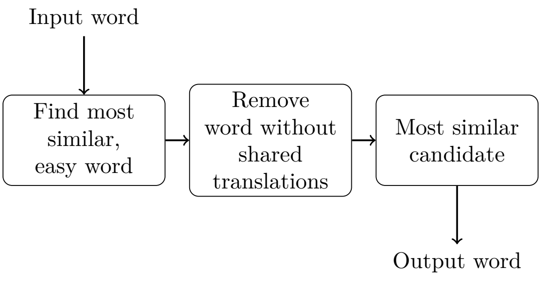


Figure 10.1: A simplified flowchart of Method 1.

Resources

The corpora used were LäSBarT (Heimann Mühlenbock, 2013), the Stockholm-Umeå-Corpus (SUC) (Ejerhed, Källgren, and Brodda, 2006), and the Swedish Wikipedia Corpus (Denoyer and Gallinari, 2006). As the LäSBarT corpus comprises Easy Language texts, we considered all words present in that corpus to be easy to comprehend. For translation of words into English, the online dictionary *bab.la*[3] was used.

Both methods used word2vec word embeddings to calculate the semantic similarity between words. Only the CBOW approach of Mikolov, Chen, Corrado, and Dean (2013) was used.

Two Synonym Replacement Methods

Two versions of synonym replacement methods were developed. **Method 1**, depicted in Figure 10.1, was inspired by Lin, Zhao, Qin, and Zhou (2003), and aimed to find the most semantically similar word to a given input word, which also shared at least one English translation with the input word.

For Method 1, the 40 most semantically similar words to the input word were extracted. The words were only extracted if they also occurred in the LäSBarT corpus. The words were translated into English, and only words that shared a translation with the input word were kept. The resulting words were compared to the input word, and the word that was most semantically similar out of the remaining words was selected as a synonym candidate.

Method 2, depicted in Figure 10.2 was in part inspired by the methods used to find synonyms described by Kann and Rosell (2005). In Method 2, the input word was translated to English, resulting in various suggested English words. Each English word was then translated back to Swedish, and words that did not occur in the LäSBarT corpus were discarded. The resulting words were then compared to the input word, and the word that was

[3] http://bab.la/

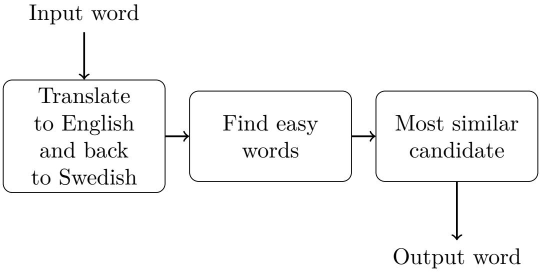


Figure 10.2: A simplified flowchart of Method 2.

most semantically similar to the remaining words was selected as a synonym candidate.

Survey

The two methods were evaluated through an online survey. In the survey, participants were presented with 45 word pairs of one original word paired with one synonym suggestion. Only input words for which both methods suggested a synonym candidate were evaluated. 30 of the suggested synonyms were extracted by Method 1, 30 of the suggested synonyms were extracted by Method 2, and 15 of the suggested synonyms were extracted by both Method 1 and Method 2 (hereafter referred to as the *intersection*).

The survey, inspired by Kann and Rosell (2005), posed the question *Is the word X a synonym to the word Y?*. The alternatives were *Disagree*, *Doubtful*, *Sometimes*, *Totally agree*, and *I do not understand the word/s*.

Results

99 participants responded to the survey. Ratings of *I do not understand the word/words* made up a very small part of the ratings (0.61–2.42%) and were, thus, excluded from data analysis.

On average, Method 1 (M = 2.58, SE = .04) performed better than Method 2 (M = 2.41, SE = .04), $t(98) = 10.90, p < .001, r = .90$).

In general, participants chose the alternatives *Disagree* and *Doubtful* more frequently for Method 2 synonyms than for Method 1 synonyms, and *Sometimes* and *Totally agree* were more frequently chosen answers for Method 1 synonyms. The response *Sometimes* was the most common for both methods.

We performed separate analyses, treating the *intersection* as its own method. When doing this comparison (disregarding the *intersection* synonyms from Method 1 and Method 2), we saw that the most common response for Method 1 and *intersection* was *Sometimes*, but for Method 2 the most frequent response was *Disagree*.

A repeated measures ANOVA, presented in Table 10.3, revealed that the methods differed significantly in performance, $F(2,196) = 197.41, p < .001$. Post hoc tests using the Bonferroni correction further revealed that Method 1 performed significantly better than Method 2, and that the intersection between Method 1 and Method 2 performed significantly better than both Method 1 and Method 2.

Table 10.3: Pairwise comparisons from the Bonferroni post-hoc test.

Pairwise	Comparisons	*M.* Diff.	Sig.
Method 1	Method 2	.166	.000
	Method 1 ∩ Method 2	-.136	.000
Method 2	Method 1	-.166	.000
	Method 1 ∩ Method 2	-.302	.000
Method 1 ∩ Method 2	Method 1	.136	.000
	Method 2	.302	.000

Discussion of Results

This study aimed to find comprehensible synonyms using a bilingual dictionary and a corpus of Easy Language texts.

We found that *Sometimes* was the most frequent answer for both Method 1 and Method 2. This is not surprising since there are not many words that are synonymous in the strict sense, but rather tend to be synonymous only within certain contexts. We also found that Method 1 generated more synonymous words than Method 2, and when the *intersection* was employed as a third method, the *intersection* proved to be significantly better. This indicates that the combination of methods could be a way forward.

It should be noted that the methods differed in terms of the number of generated synonym candidates. Method 1 proved to be stricter in filtering candidates, and sometimes failed to return a suggestion. Method 2, on the other hand, was generally more successful in finding synonym suggestions. Since only input words to which both methods suggested a synonym candidate were selected for the survey evaluation, this had a negative impact on the performance of Method 2 as some Method 2 synonyms were discarded when Method 1 failed to return a synonym suggestion.

Synonyms Based on Extraction of Basic-level Words

Another approach to finding more comprehensible synonyms was inspired by the field of cognitive linguistics. The aim of this study was to use word taxonomies in order to extract superordinate words, under the assumption that such words are more basic and, thus, can enhance comprehension.

Basic-level Words

In human categorisation some concepts are more representative than others (Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976). For example, *furniture* can be regarded as higher up in the taxonomy than *chair* or *table*, whereas *kitchen chair* or *dining table* can be found at a lower level with higher specificity.

There are concepts which can be considered more *basic* than others. In the taxonomy above, we could claim that *table* is the most basic word. Such concepts seem to be the most important to human categorisation (Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976).

Words describing basic concepts have some specific characteristics. For example, basic-level words seem to be more frequently used in language, and they emerge early in a child's language acquisition. Another feature of basic-level words is that they also often comprise one single lexeme, while subordinate terms more often consist of multiple lexemes (Evans, 2019).

Hyponomy

Hyponomy is a type of semantic relation between words, where a hyponym denotes a subordinate word and hypernym denotes a superordinate word. Hyponomy could be interesting in the context of lexical simplification. Drndarević and Saggion (2012) analysed a parallel corpus in standard and simple news texts in Spanish, and found that the exchanged words sometimes could be hypernyms, hyponyms and meronyms. The number of hypernym levels can also serve as an indication of concreteness: words with many hypernym levels are often more concrete, and words with fewer hypernym levels are often more abstract (Graesser, McNamara, Louwerse, and Cai, 2004b).

We hypothesised that a corpus of Easy Language text is characterised by a higher proportion of basic-level nouns than a standard corpus. We further hypothesised that hypernyms with characteristics of basic-level words could be useful for the task of synonym replacement.

Corpus Analysis

For this reason, we analysed and compared two pairs of corpora.

The first pair of corpora were the *Stockholm-Umeå Corpus, v.3 (SUC3)* (Ejerhed, Källgren, and Brodda, 2006) and the *LäSBarT* corpus (Mühlenbock, 2008). The hypothesis was that the *SUC3* corpus would exhibit a higher average number of steps to the top-level noun than the *LäSBarT* corpus.

The second pair of corpora were a corpus of the Swedish newspaper *8 Sidor*, which comprises news articles in Swedish Easy Language, and a corpus of news articles (*GP2D*). The corpora were of the same genre, but not parallel. The reason for this comparison was to investigate whether the genre did play a role. The hypothesis was that the *GP2D* corpus would exhibit an even higher average number of steps to the top-level noun than the *8 Sidor* corpus.

Procedure

First, all nouns were extracted from the resources, together with their most probable sense according to SALDO (*Svenskt Associationslexikon*) version 2 (Borin, Forsberg, and Lönngren, 2008). SALDO is a descriptive lexical resource that includes a semantic lexicon in the form of a lexical-semantic network. SALDO was also used for extracting lexical relations. For each extracted noun, we recursively collected all *primary parents*. The *primary* descriptor describes an entry which fulfils two requirements better than any other entry: 1) it is a semantic neighbour of the entry to be described, which means that there is a direct semantic relationship, such as synonymy, hyponymy, and meronymy, between words; and 2) it is more central than the given entry. However, there is no requirement that the *primary* descriptor is of the same part of speech as the entry itself.

We counted the number of steps taken to reach the top-level noun in the taxonomy, and the procedure ended when no more parents tagged as a noun were present. The method was inspired by the collection of synonym/near-synonym/hypernym relations in Borin and Forsberg (2014).

The frequency counts of the nouns occurring in the corpora were collected, as well as their superordinate nouns and an indication of compositionality. The frequency measures used were relative frequencies gathered from the *WIKIPEDIA-SV* corpus, accessed through Språkbanken[4].

Corpus Analysis Results

The following numbers of instances were extracted: 206,609 (*SUC3*), 177,390 (*LäSBarT*), 180,012 (*GP2D*), and 543,699 (*8 Sidor*). The distribution of the number of words per superordinate level is presented in Figure 10.3.

First, we compared the *SUC3* corpus with the *LäSBarT* corpus. To compare the medians, a Mann-Whitney *U* test was performed. On average, the words of the *SUC3* corpus had a slightly lower number of steps to the top-level noun ($M = 0.93, Md = 1.0$) than the words of the *LäSBarT* corpus ($M = 1.02, Md = 1.0$). This difference was significant ($U = 17489728875.50, n1 = 206,609, n2 = 177,390, p < 0.001, cles = 0.32$).

Secondly, we compared corpora of the same genre (news texts): *GP2D* and *8 Sidor*. To compare the medians, a Mann-Whitney *U* test was performed. On average, the words of the *GP2D* corpus had a slightly higher number of steps to the top-level noun ($M = 1.03, Md = 1.0$) than the words of the *8 Sidor* corpus ($M = 0.93, Md = 1.0$). This difference was significant ($U = 46166030968.50, n1 = 180,012, n2 = 543,699, p < 0.001, cles = 0.37$).

[4] https://spraakbanken.gu.se/verktyg/korp/korpusstatistik

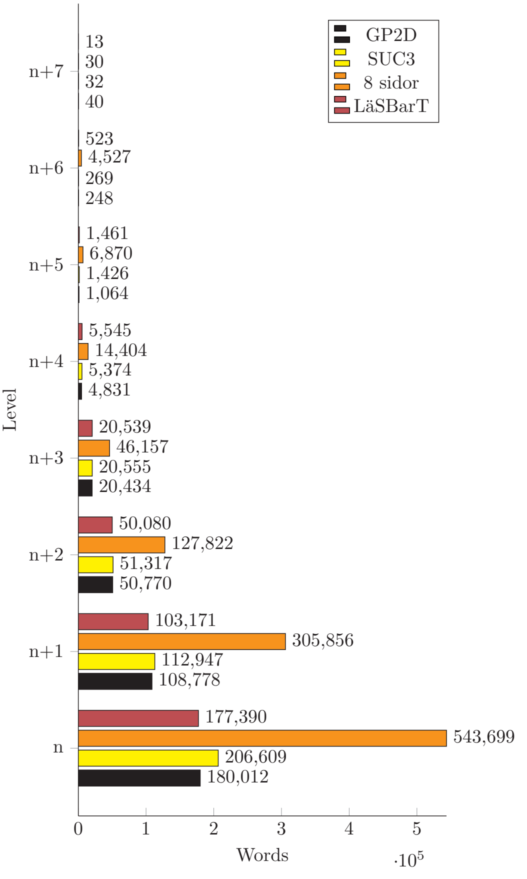


Figure 10.3: Number of words in the corpora at the various levels. Words at level n are the words in the corpora.

Table 10.4: Average relative frequencies at each level of the words of the corpora. Highest level frequencies in boldface.

	SUC3	LäSBarT	GP2D	8 Sidor
Level n	140.66	**190.02**	155.44	274.54
Level n+1	219.69	176.59	195.59	212.82
Level n+2	199.97	165.67	163.39	203.38
Level n+3	**280.56**	126.48	310.78	**317.01**
Level n+4	84.60	48.68	113.92	88.22
Level n+5	74.25	38.10	93.04	34.64
Level n+6	51.04	30.88	79.37	33.24
Level n+7	83.47	36.03	**401.41**	53.76

The analyses of the relative frequencies of the corpora are presented in Table 10.4. The words at level n are the words that appear in the corpora[5], and each n+i step refers to the superordinate words. Three of the corpora (*LäSBarT*, *GP2D* and *8 Sidor*) had words represented at the level n+8, but since these words were very few (1, 4 and 1 words respectively), they were excluded from the analysis.

The *SUC3* corpus had the highest relative frequencies at level n+3. The *LäSBarT* corpus had the highest relative frequencies at level n. The *GP2D* corpus had the highest relative frequencies at level n+7. The *8 Sidor* corpus had the highest relative frequencies at level n+3.

All corpora, except for the *LäSBarT* corpus exhibited a tendency to peak at level n+3 (see Table 10.4 and Figure 10.4). For the news corpora, we observed that the *8 Sidor* corpus had the highest relative frequency at level n, and the standard news *GP2D* had the highest relative frequency at level n+4.

Synonym Replacement Algorithm

We know that basic-level words occur more frequently in language, and that they tend to consist of single lexemes. Thus, a synonym replacement algorithm should reward monolexemic synonym candidates with high relative frequency. In order to include this in the algorithm, we consulted the frequency corpus for information about whether or not the word could be interpreted as a compound.

From the corpus analyses, we learnt that in the two standard corpora, there seems to be a frequency peak at level n+3. One explanation for this

[5] We use the notation *level n* to describe the words of the corpora instead of, for example, level 0 words, as we do not know the level of inclusiveness at which they actually appear. The words at level n are the words as they appear in the corpora, thus, they could be anywhere on the vertical axis of inclusiveness of the category. The only thing we know is the number of superordinate words, and therefore we chose to use the notation n for the corpus-level and n+i for each superordinate level.

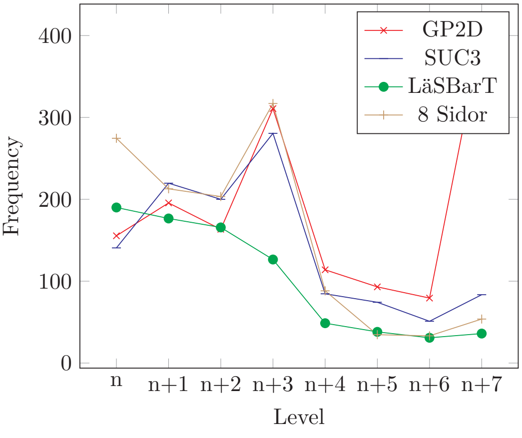


Figure 10.4: Relative frequencies at each level of the word hierarchy in the corpora.

could, however, be that when climbing higher up in the word taxonomy, more general words are found. When searching for synonyms, we hypothesise that the more general words are not necessarily good synonym candidates. For instance, whereas *horse* could be a good-enough synonym candidate for the word *Shetland pony*, the word *animal* might be too general. We conducted experiments with varying levels and chose to restrict our synonym-seeking algorithm to not go beyond level n+2.

Based on the above reasoning, we developed an algorithm for choosing the best synonym from the extracted nouns and their superordinate words.

The resulting algorithm (hereafter: *FM*) is presented in Algorithm 1. It selects, from words at most two levels up in the hierarchy, the most frequent monolexemic word. If no such word exists, it selects the most frequent word.

```
Data: candidates: a word chain containing the word of the corpus
      and the superordinate words collected from Saldo.
Result: best synonym from candidates
candidates.sort(key=frequency);
bestSynonym = candidates[0];
for word in candidates[:3] do
    if word is monolexemic then
        bestSynonym = word;
        break;
    end
end
```

Algorithm 1: The FM algorithm for choosing synonym.

Assessment of Synonym Replacement Algorithm

The performance of the *FM* algorithm was assessed by comparison with two baseline algorithms. The *OneLevel* baseline algorithm selects the first hypernym of a given word (if such a word existed), and the *Freq* baseline algorithm selected synonyms based on the overall highest relative frequency. The *Freq* baseline algorithm did not take any information about monolexemity into account.

We ran all algorithms on the nouns extracted from the standard corpora: *SUC3* and *GP2D*.

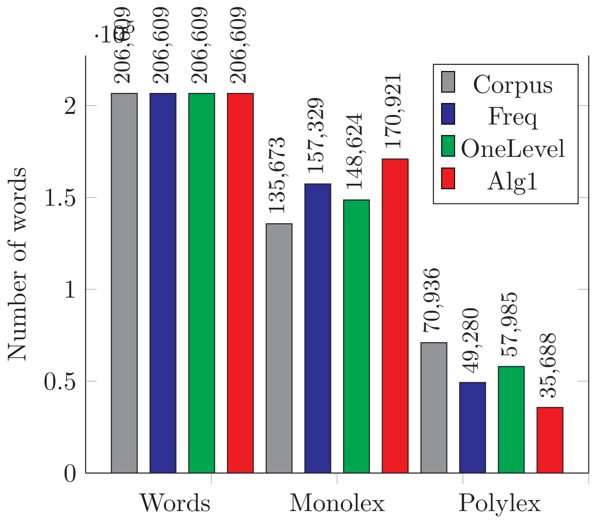


Figure 10.5: Number of total words, monolexemic words, and polylexemic words in the SUC3 corpus after applying the algorithms. *Corpus* denotes the original values of the specific corpus.

Table 10.5: Example synonyms chosen by the different algorithms

Example word chain	FM	OneLevel	Freq
procent - hundradel - bråkdel - del *percent - centesimal - fraction - part*	procent	hundradel	del
universitet - högskola - skola *university - college - school*	universitet	högskola	universitet
rubel - myntenhet - mynt - pengar *ruble - currency unit - coin - money*	mynt	mynthenhet	mynt

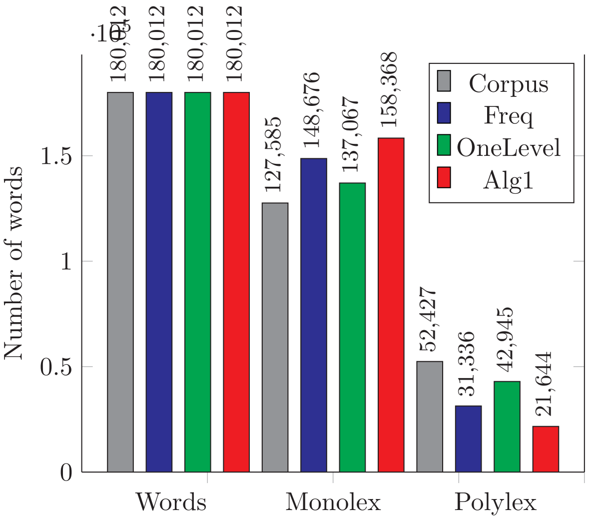


Figure 10.6: Number of total words, monolexemic words, and polylexemic words in the GP2D corpus after applying the algorithms. *Corpus* denotes the original values of the specific corpus.

The number of monolexemic and polylexemic words of both corpora are presented in Figure 10.5 and Figure 10.6. The relative frequencies after running the algorithms are illustrated in Figure 10.7.

For the *SUC3* corpus, all algorithms increased the number of monolexemic words. The largest increase was observed for the FM algorithm (+35, 248), followed by Freq (+21, 656), and OneLevel (+12, 951). All algorithms increased the average relative frequency of the substituted words. The largest increase was observed for Freq (+153.68), followed by FM (+120.92), and OneLevel (+34.68).

For the *GP2D* corpus, all algorithms increased the number of monolexemic words. The largest increase was observed for the FM algorithm (+30, 783), followed by the Freq algorithm (+21, 091), and OneLevel (+9, 482). All algorithms resulted in a higher average relative frequency, and the largest increase was observed for the Freq algorithm (+149.54), followed by the FM algorithm (+110.58), and OneLevel (+7.2).

Table 10.5 shows examples of the synonyms chosen by the algorithms.

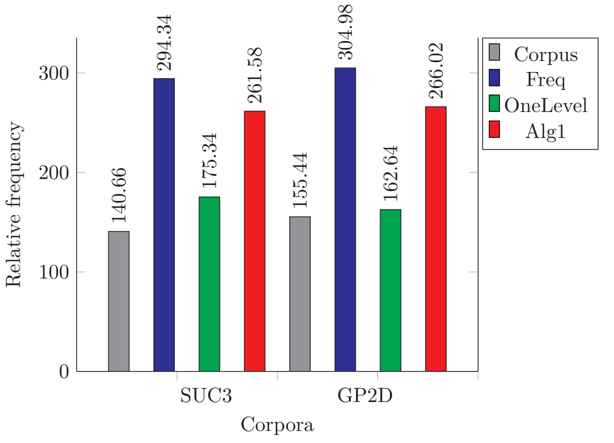


Figure 10.7: Relative frequencies for each corpus after applying the algorithms. *Corpus* denotes the original values of the specific corpus.

Discussion of Results

Our intuition was that simple texts would comprise more basic-level words than texts written in standard Swedish. We found no clear support for this claim. In the statistical analysis we compared very large samples, and the presence of statistical significance is not surprising. When comparing the means and medians of the datasets, it is clear that the differences are small, and the results should be interpreted with caution.

The *SUC3* corpus had a significantly lower average number of steps to the top-level noun, compared to the *LäSBarT* corpus. These results contradicted our expectations. The simple news corpus *8 Sidor* had a significantly lower number of steps to the top-level noun than the standard news corpus *GP2D*. The relative frequency analysis showed that the *8 Sidor* corpus has relatively high average relative frequency at the base level (level n), although it peaks at level n+3. The *GP2D* corpus, on the other hand, generally had lower average frequencies at level n and the highest frequencies at level n+7.

We expected the standard corpora to have lower relative frequencies at the base level than the simple text corpora. This difference can be observed in the *LäSBarT* corpus, but is less prominent in the *8 Sidor* corpus. However, the *8 Sidor* corpus shows consistently high frequencies, even at the lower levels. The level n score is the second highest frequency score for this corpus, and much higher when compared to the level n score of *GP2D*.

The *GP2D* corpus had the highest average frequency at level n+7. This indicates that the words used in this corpus are more specific than in the other corpora. However, this score is based on a relatively low number of words (40), and is also due to the fact that this corpus exhibits a frequency peak at level n+3.

For *SUC3* and *8 Sidor*, the most frequent words were found at level n+3. This indicates that more basic-level words could be found if we selected the words three levels up in the word taxonomy. However, this could also indicate that the words at this level are higher up at Rosch's vertical axis, which means that they are more inclusive than the basic-level words, and therefore more frequent (compare: *Shetland pony*, *horse*, *animal*).

In conclusion, we presented results from a corpus study indicating the proportions of basic-level words in corpora of simple and standard Swedish. The studies revealed that the corpus of simple news text did include more basic-level nouns than the corpus of standard news. This indicates that lexical simplification may benefit from traversing a word taxonomy upwards. This strategy could serve as a complement to other synonym replacement methods. We then developed an algorithm which aimed to reward high relative frequencies and monolexemity, while not climbing too high on the word hierarchy. The algorithm seemed to perform well with respect to these criteria.

10.3 Chapter Summary

This chapter described the automatic text simplification efforts made in our research projects.

First, we presented the original implementation of a rule-based syntactic simplification system. The system has been developed further, and as of today it includes simplification operations for rewriting from passive to active tense, quotation inversion, rearrangement to straightforward word order, and splitting a long and complex sentence into multiple sentences. We briefly discussed the potential benefits and drawbacks of using hand-crafted simplification rules for syntactic simplification.

This chapter also presented two different approaches to lexical simplification. We developed, implemented and tested different strategies for synonym replacement. The first methods used a bilingual dictionary for extracting more comprehensible synonyms from an Easy Language corpus. Synonyms generated by a combination of the two methods were perceived as more synonymous. Another approach to lexical simplification was to derive word taxonomies from corpora and traverse the word hierarchy upwards in order to find near-synonymous words with basic-level characteristics. We argued that this could serve as a complement to the often-used replacement methods that rely on word length and word frequency measures.

From the various guidelines for writing Easy Language (elaborated in Chapter 7), we know that writers are advised to use words that are *short*, *simple*, *familiar*, and *concrete*. But knowing what this means in reality, and what it means for a specific target audience, is more difficult. There are also other characteristics that could indicate complexity in a word, such as word length, age of acquisition or concreteness.

We also know that the various target audiences may have differing needs regarding the characteristics of a word. This motivated us to explore several ways of automatically extracting more comprehensible synonyms for a given input word. The first two methods intended to capture the comprehensiveness by only admitting words occurring in an Easy Language corpus, and the third method was based on theories from cognitive linguistics.

CHAPTER 11

Summarisation as a Simplification Tool

The adaptation tools developed within our projects include one system for abstractive summarisation and one system for extractive summarisation.

Some of our work concerned exploring automatic text summarisation as an aid for making texts easier to read. First, we explored cohesion errors made by an extractive summarisation system and the effect they have on reading such texts. This work is presented in Section 11.1. Then, we compared the perceived readability and text quality in an extractive and an abstractive summarisation system, and this work is presented in Section 11.2.

11.1 Cohesion Errors in Extractive Summaries

Since extractive summaries are created by the extraction of complete sentences, there is a risk that relations between sentences are lost in the summarisation process, resulting in cohesion errors. A lack of cohesion may result in an erroneous interpretation of a text (Otterbacher, Radev, and Luo, 2002). Anaphoric references in particular are known to cause problems in automatic text summarisation (Hassel, 2000; Mani, Bloedorn, and Gates, 1998), and the higher the level of summary is, the more errors are found (Kaspersson, Smith, Danielsson, and Jönsson, 2012).

We conducted an eye-tracking study, investigating how such cohesion errors affect the reading of extractive summaries.

Eye-tracking provides a link to the on-going cognitive processes during the execution of a certain task. Eye movement is a result of both goal driven and stimulus driven processes (Duchowski, 2007), and depends strongly on the type of cognitive task that is being performed. In this study, we used the following measures:

- *Number of fixations*. A fixation denotes the period of time where the eye is relatively still (about 200-300 ms).
- *Fixation duration*.

  The fixation duration indicates the effort needed for the cognitive processing, but the average fixation duration varies depending on the task and stimuli. The more complicated a text is, the longer the average fixation duration, and factors like stress might result in shorter fixations (Holmqvist, 2011).

  All words of a text are not fixated during reading. Long words are more likely to be fixated than short ones (Just and Carpenter, 1980), but other aspects such as frequency and predictability from context are also proven to be a reason for shorter fixations or word skipping (Reichle, Pollatsek, Fisher, and Rayner, 1998).
- *Pupil size*, which increases during problem solving and correlates to the difficulty of the task, which in turn implies that this could be used as a measure of cognitive activity (Hess and Polt, 1964).

Error Types

The categorisation of errors was derived from Kaspersson, Smith, Danielsson, and Jönsson (2012):

1. Erroneous anaphoric reference
   a) Noun-phrases
   b) Proper names
   c) Pronouns
2. Absent cohesion or context
3. Broken anaphoric reference
   a) Noun-phrases
   b) Proper names
   c) Pronouns

Erroneous anaphoric references describe errors that occur when an anaphoric reference refers to an incorrect antecedent. This is often the case when the summary has not included the correct antecedent and at the same time there is another antecedent in the text that fits. There are three sub-types of erroneous anaphoric references: noun-phrases, proper names and pronouns.

Absent cohesion or context describes the case when the extracted sentences lack cohesion or context, which affects comprehension of the summary.

Broken anaphoric references are errors which occur when the summariser does not extract the antecedent that is referred to in an anaphoric reference. There are three sub-types of broken anaphoric references: noun-phrases, proper names and pronouns.

Areas of Interest

To analyse the data recorded by the eye tracking equipment, areas of interest (AOIs) were defined. There were four different AOIs corresponding to the four error types. Due to the nature of the errors, the AOIs varied in size. Therefore, the number of fixations was normalised by the size of the AOIs, in order to get comparable scores.

Procedure

The study included 23 participants (13 male, 10 female). They were all native Swedish speakers without any writing or reading disability and with normal or corrected-to-normal vision. The average age was 23.2 ($SD = 2.76$).

The eye tracking equipment used for this study was an SMI iView RED II 50 Hz Pupil/Corneal reflex camera mounted underneath a 19" computer monitor. The software applications used for recording and analysing the eye tracking data were iView X, Experiment Center 3.0 and BeGaze 2.

The experimental procedure consisted of four parts: answering a questionnaire, text reading, error marking and text rating.

Texts and Errors

The original texts used in the tests were four texts from the Swedish popular science magazine *Forskning och Framsteg*. The summaries were produced by the extractive text summariser CogSum (Smith and Jönsson, 2011b) in a previous project, where they also had been tagged for errors (Kaspersson, Smith, Danielsson, and Jönsson, 2012). The summary length was set to 33%, which means that a third of the sentences of the original text were extracted. The level of summarisation was selected to get as many errors as possible in a text, while keeping it at a reasonable length that was still readable (Kaspersson, Smith, Danielsson, and Jönsson, 2012). The four texts varied in length from 11 to 14 sentences and the number of tagged errors varied from 6 to 12 per text. There were 34 errors in total.

Not all error types were present in the texts of the experiment. The error types and number of errors for each type that were present in the texts were:

- 1(c) Erroneous anaphoric reference - Pronouns, a total of 4 errors
- 2. Absent cohesion or context, a total of 16 errors
- 3(a) Broken anaphoric reference - Noun-phrases, a total of 4 errors
- 3(c) Broken anaphoric reference - Pronouns, a total of 10 errors

The number of errors and number of sentences for each text is presented in Table 11.1. Text 2 was the shortest text, and also contained the fewest errors, which resulted in a relatively low percentage of errors per sentence. Text 3 and 4 were equally long but text 3 had a higher percentage of errors per sentence.

Table 11.1: Description of the texts used in the test. The row labelled *Error rate* represents the ratio of the number of errors and the number of sentences.

	Text 1	Text 2	Text 3	Text 4
No. of errors	7	6	12	9
No. of sentences	12	11	14	14
Error rate	58.33%	54.55%	85.71%	64.29%

Questionnaire

The questionnaire aimed to capture the participants' reading strategies and prior attitudes toward reading. The participants were presented with five assertions, which they were asked to rate according to a 5-point unipolar Likert scale. The assertions were:

- I usually understand what I read
- I am a slow reader
- I find it easy to read
- I find it exhausting to read
- I am often pleased to get a rough idea of a text's content

In the questionnaire, the participants also reported age, gender, profession or current education, and whether glasses or contact lenses were used during the experiment.

Experimental Procedure

In the test, the participants read all four summaries, while their eye movements were recorded. The participants were instructed to read the text for as long as they wanted, and they were told that they were going to perform a task after having finished the reading part. However, they were not told what the task consisted of. The participants were not aware that the texts that they read were summarised, and the order of the texts were not the same for all participants.

After reading the four texts, the participants were given printed copies of the texts and were instructed to mark the parts of each text that they considered to be most problematic to read. There was no limit in how many parts or areas they could mark. The participants were then asked to rate each marking according to how difficult they were on a scale of 1–3, where 1 corresponded to the least difficult and 3 the most difficult. The participants were then asked to explain their markings.

In a final task, the participants rated the texts according to how difficult, boring, interesting and exhausting they were, again using a 5-point Likert scale. After having finished the test, the participants were asked whether they felt that the presence of the eye movement camera had any impact of their performance. It was also revealed to them that the texts were automatically produced summaries. They were then asked if they thought that their attitude towards the texts would be different if they knew in advance that the summaries were automatically produced.

Results

This section presents the results from all parts of the experiment conducted in this study. First, we present the results of the questionnaire and text rating questions. Then we present the results of the manual error marking and the subjective rating of the marked errors. Last, we present the results of the eye tracking experiment.

Attitude to Reading

The questionnaire that evaluated the participants' prior attitudes toward reading produced the results presented in Table 11.2. The participants generally considered themselves to be good readers.

Text Rating

The texts used were evaluated regarding three different criteria: whether they were easy to understand, boring or exhausting to read. The results are shown in Table 11.3.

Table 11.2: Mean and standard deviation of participants' self-rated reading abilities and attitudes towards reading.

Assertion	Mean	Std.Dev.
I usually understand what I read	4.61	.58
I am a slow reader	2.52	.90
I find it easy to read	4.70	.70
I find it exhausting to read	1.65	.88
I am often pleased to get a rough idea of a text's content	3.70	1.02

The texts differed slightly. According to the means, text 2 was considered the easiest, least boring and least exhausting text while text 1 was the most boring text and text 3 was the most exhausting text to read.

Table 11.3: Mean and standard deviation (within parentheses) of the text ratings.

Assertion	Text 1	Text 2	Text 3	Text 4
Easy	3.43 (.90)	3.96 (1.33)	2.91 (1.20)	3.78 (1.17)
Boring	3.52 (1.08)	1.78 (.85)	3.00 (1.08)	2.17 (.98)
Exhausting	2.96 (1.14)	2.00 (1.09)	3.48 (.99)	2.39 (1.1)

A repeated measures ANOVA revealed that for the criterion *easy* the texts differed significantly $F(3, 66) = 4.02, p < .05$. A Bonferroni post-hoc test showed that there was a significant difference between text 2 and text 3 ($p < .05$), indicating that text 2 was easier than text 3.

We also found significant differences for the criterion *boring*, $F(3, 66) = 15.28, p < .001$. A Bonferroni post-hoc test revealed that there are significant differences between text 1 and text 2 ($p < .001$), text 1 and text 4 ($p < .05$), and between text 2 and text 3 ($p < .05$). The results indicate that text 1 and text 3 were significantly more boring than text 2.

Regarding the criterion *exhausting*, we found significant differences $F(3, 66) = 9.37, p < .001$. A Bonferroni post-hoc test revealed that there were significant differences between text 1 and text 2 ($p < .05$), text 2 and text 3 ($p < .001$), and between text 3 and text 4 ($p < .05$). Text 1 and text 3 were, thus, more exhausting to read than text 2, and text 3 was more exhausting to read than text 4.

Error Marking and Subjective Rating

From the analysis of the error markings made by the participants, we found that the participants marked problems other than the previously identified error types. We performed a categorisation of the markings based on the comments that the participants made. The participants markings are presented in Table 11.4.

The predefined errors (*Cohesion error*) made up 38.3% of the total number of markings. The second most frequent marking (17.55%) was different types of language related problems, e.g. long sentences or a complicated word order. Difficult words accounted for 11.7% of the total number of markings. 9.04% of the markings treated general problems with understanding the context. Summariser errors and numbers made up 7.45% and 4.79% respectively of the total number of marked areas. The category *Other* accounts for 11.17% of the markings and contains markings that the participants were not able to explain.

Table 11.4: Distribution of cohesion errors and other categories that were marked by the participants.

Category	Percentage
Cohesion error	38.3%
Language	17.55%
Difficult words	11.7%
Context	9.04%
Summariser errors	7.45%
Numbers	4.79%
Other	11.17%

The mean of the subjective rating (ranging from 1 (least difficult) to 3 (most difficult)) of the marked errors is presented in Table 11.5.

Table 11.5: Mean and standard deviation of the subjective rating for each error type.

Error type	Mean	Std.Dev.
Error 1c	1.82	0.77
Error 2	1.85	0.81
Error 3a	1.70	0.86
Error 3c	1.88	0.78

All error types scored relatively similarly, with scores ranging from 1.70 to 1.88. No statistical significance was found between the subjective ratings of each error type. All participants reported that their attitudes would have been more lenient if they knew in advance that the texts used in the test were summaries.

Eye Tracking Results

The results of the eye tracking experiment are presented in Table 11.6. The row labelled *None* refers to an area which has not been marked as an AOI, i.e. the rest of the text.

Table 11.6: Mean and standard deviation of the number of fixations, fixation duration and pupil size for each error type. The values corrected for the size of the AOI are within parentheses.

Error	Number of fixations	
	Mean	Std.Dev.
1c	13.61 (2.28)	6.22 (1.04)
2	210.30 (8.69)	51.91 (2.15)
3a	12.70 (1.75)	4.30 (.59)
3c	22.61 (6.14)	5.08 (1.38)
None	841.44 (1.25)	193.77 (.29)
	Fixation Duration	
	Mean	Std.Dev.
1c	291.88	76.14
2	280.25	41.63
3a	269.17	52.83
3c	279.20	59.19
None	273.33	41.13
	Pupil Size	
	Mean	Std.Dev.
1c	10.82	1.20
2	10.76	1.21
3a	10.73	1.20
3c	10.79	1.22
None	10.82	1.21

A repeated measures ANOVA was used to test for differences between the four error types and the rest of the text. For fixation duration and pupil size, we found no differences ($p > .05$). For the corrected number of fixations there was a significant difference $F(2.160, 47.522) = 251.86, p < .001$, Greenhouse-Geisser corrected.

Bonferroni post-hoc tests revealed significant differences, presented in Table 11.7. Statistically significant differences are marked in bold.

Table 11.7: Pairwise comparisons from the Bonferroni post-hoc test. Significant differences are marked in bold.

Pairwise	Comparisons	*M.* Diff.	Sig.
Error 1c	Error 2	**-6.41**	.000
	Error 3a	.53	.065
	Error 3c	**-3.86**	.000
	None	**1.03**	.000
Error 2	Error 1c	**6.41**	.000
	Error 3a	**6.94**	.000
	Error 3c	**2.55**	.000
	None	**7.44**	.000
Error 3a	Error 1c	-.53	.065
	Error 2	**-6.94**	.000
	Error 3c	**-4.39**	.000
	None	**.50**	.002
Error 3c	Error 1c	**3.86**	.000
	Error 2	**-2.55**	.000
	Error 3a	**4.39**	.000
	None	**4.90**	.000
None	Error 1c	**-1.03**	.000
	Error 2	**-7.44**	.000
	Error 3a	**-.50**	.002
	Error 3c	**-4.90**	.000

All error types were fixated significantly more than the rest of the text (p <.05).

Error type 2, absent cohesion or context, had significantly more fixations than all other error types (p <.001). Error type 3c, broken anaphoric reference (pronouns), had significantly more fixations than error types 1c and 3a (p <.001).

Significant differences were found between all error types except for 1c, erroneous anaphoric reference (pronouns), and 3a, broken anaphoric reference (noun-phrases) (p = .065). The marginal significance level suggests a tendency toward slightly more fixations on error type 1c than on error type 3a.

None of the participants reported that the presence of the eye movement camera had any significant impact on their reading performance.

Discussion of Results

This section discusses the results of all parts of the experiment.

Text Ratings

We observed several differences between the texts. Text 2 was clearly the easiest, least boring and least exhausting text. It was also the shortest text, and had the lowest percentage of errors present per sentence.

Text 3 was considered the most exhausting text, and it also had the highest number of errors per sentence and was one of the longest texts. When compared to text 4, which was equally long but with fewer errors, we observed that text 4 was considered less exhausting than text 3, but we noted no difference in difficulty between the two texts. This indicates that the experience of the text depends more on the number of cohesion errors than the text length.

Text 1 was rated as the most boring text, even though it was similar to text 2 in length and number of errors. The reason that text 1 was considered more boring might be that the topic was considered boring. Text 1 treated the Nobel Prize while text 2 treated polar bears, and it is possible that the second topic was more attractive to the reader.

Interestingly, all participants claimed that they would have been more lenient with the texts if they knew that they were automatically summarised. This could indicate that the experienced errors could be considered less severe in a real-word application.

Error Marking and Subjective Rating

The majority of areas marked by the participants (38.3%) corresponded to the previously identified cohesion errors. Regarding the rest of the error markings, we observed that some of them were not necessarily due to the text being automatically summarised. For instance, 11.7% of the markings represented difficult words in the text. 17.55% of the markings corresponded to other linguistic factors, such as long sentences or phrases with difficult word orders.

The nature of the extractive summariser could explain some of the issues. As the extractive method extracts the most important sentences of a text, it is inevitable that information is lost, which could cause difficulties in understanding the general context or could leave behind sentences with strange word orders. For instance, 9.04% of the markings were areas that the participants claimed were out of context. These error markings did not correspond to any of the predefined error types. This can be explained by the fact that the error type is rather vague, and probably affects the text as a whole, rather than one specific sentence. 11.17% of the markings belonged to the *Other* category. Since the participants failed to explain this error category, we can only speculate about the reasons behind it, but it is possible that at least some of them could be due to missing context. Regarding the subjective rating of the errors, we found no significant difference, which suggests that none of the error types was considered more problematic.

Eye Tracking Results

The statistical analysis of the eye tracking data indicates that error type 2 and error type 3c were the most problematic, according to the number of fixations. However, we observed no difference for the fixation duration or pupil size.

Long fixations indicate cognitive processing, and the absence of such fixations would imply that the errors did not cause any substantial cognitive effort. However, according to Ehmke and Wilson (2007), many short fixations might indicate confusion when expected information is missing. Although this claim is made within the field of usability research and is applied to a web stimulus, it could be seen as a possible interpretation to the pattern of many short fixations within the areas of error type 2 and 3c.

Error type 1c, erroneous anaphoric reference, had significantly more fixations than the rest of the text, but fewer fixations compared to the other error types (except for error 3a where no statistical significance was found). The reason that this error type is fixated less might be because it is difficult to identify. The anaphoric expression of this error type does refer to an existing (but erroneous) antecedent. Since there is an antecedent present, the reader might not discover that the antecedent is erroneous, which could cause an erroneous interpretation of the text.

We observed no statistically significant change in pupil size. This suggests that the participants did not find any of the cohesion errors more cognitively demanding. Pupil size is also a sensitive metric, and since our experiment did not control for factors like fatigue or light variation, this could be another possible explanation for the minimal change in pupil size.

Summary of Results

From this study, we concluded:

- Cohesion errors affect the experience of reading a summary negatively. This is suggested by the higher number of fixations for error types 2 and 3c.
- More cohesion errors could make a text more exhausting to read. However, we saw that a text with a high number of cohesion errors per sentence was not more difficult. This suggests that the errors indeed cause problems during reading, but the impact may be restricted to the effort required to read, rather than the comprehension.
- Other factors may constitute a source of disturbance in extractive summaries. The majority of areas marked by the participants were the previously identified cohesion errors, but the participants pointed out other aspects as problematic. Such factors included linguistic factors (17.55%) and difficult words (11.7%). This indicates that there are other factors affecting the experience of reading automatically produced summaries.

- Cohesion errors affect reading. However, the disturbances need not be severe. This is suggested by the number of fixations. We found no difference in average fixation duration and pupil size, which could indicate that participants did not find the cohesion errors more cognitively taxing than the rest of the text.

The participants in the study were not aware of the fact that the texts they read were summarised. Moreover, all participants claimed that they would have had a more lenient attitude towards the texts if they knew that they were automatically produced. Thus, it is possible that we would have gotten a different result if the participants were aware of this fact before starting the test. An interesting topic for future investigation could be whether summaries are preferred over the original texts, despite their weaknesses.

11.2 Perceived Readability and Text Quality in Extractive and Abstractive Summaries

In a second study, we aimed to compare aspects of perceived readability and text quality in extractive and abstractive summaries of Swedish news texts. In this study, we trained two automatic text summarisers (one extractive, one abstractive) on the same data set, and evaluated the resulting summaries via an online survey, exploring the notions of fluency, adequacy and simplicity (Wubben, Bosch, and Krahmer, 2012). *Fluency* is defined as the extent to which a summary contains proper grammatical sentences. *Adequacy* is defined as the extent to which a summary conveys the same meaning as the source document. Finally, *simplicity* is defined as the extent to which a summary is easy to understand.

Data

The corpus used for training the summarisation models was the DN-LC dataset (Monsen and Jönsson, 2021). The corpus comprises news articles published in *Dagens Nyheter (DN)* during the years 2000–2020. The preamble of each article was used as the summary. For training the models, we used 349,935 article-summary pairs. 9,000 of them were used for testing and 1,000 for validation. The average article length in the training set was 476 words (30.3 sentences), and the average summary length was 33 words (2.5 sentences).

Summarisation Models

The abstractive model used in this study was trained based on the methodology proposed by Rothe, Narayan, and Severyn (2020), utilising a pre-trained Swedish BERT model (Malmsten, Börjeson, and Haffenden, 2020) to warm start an encoder-decoder model. Both the encoder and the decoder were warm started with the Swedish BERT model weights, which were also shared between the two components. The warm-started model was fine-tuned in Google Colab on a Tesla V100-SXM2-16GB GPU for $300,000$ steps with a batch size of 10. The model achieved a ROUGE-1 score of 33.73 and a ROUGE-2 score of 13.31 on the test set.

The extractive model was trained using the same pre-trained Swedish BERT model as a base. The TransformerSum[1] framework was used for fine-tuning the model for extractive summarisation. To use the DN dataset, which is intrinsically abstractive, it was transformed into an extractive dataset with the convert_to_extractive.py script. However, the script was modified and adapted for Swedish text by substituting the inherent spaCy tokeniser with a Swedish NLTK tokeniser. This script reformatted the abstractive dataset by determining the best extractive summary for each article-summary pair that maximised ROUGE scores.

The model was fine-tuned by running the main.py script. We used a batch size of 16 and trained the model for three epochs in Google Colab on the Tesla V100-SXM2-16GB GPU. This extractive model achieved a ROUGE-1 score of 30.83 and a ROUGE-2 score of 10.40 when extracting the top three candidate sentences.

Survey

15 articles with one abstractive and one extractive summary were evaluated in an online survey. The articles were randomly sampled from the test set, and we restricted the sampling to articles between 300 and 350 words.

The length of the extractive summaries was restricted to a length corresponding to a third of the original article. The abstractive summaries were adjusted to be of a similar length.

We calculated ROUGE scores for the 15 articles and their respective summaries. The abstractive summaries had a ROUGE-1 score of 27.03, and a ROUGE-2 score of 10.77 and the extractive summaries had a ROUGE-1 score of 24.65 and a ROUGE-2 score of 7.79. The lower scores (compared to the test set) are explained by the fact that the survey summaries generated were longer than the test set summaries Appendix A shows examples of summaries produced by respective models.

The 15 articles were divided into 5 different survey versions with 3 articles in each. The survey was distributed and shared on social media from the researchers' personal accounts, resulting in a convenience sample of 37 participants. The submitted responses were approximately evenly distributed between the different articles. The least attended article received four re-

[1]https://github.com/HHousen/TransformerSum

sponses, while the most attended article received nine. On average, each article received 7.4 responses.

The participants were presented with the original news article, as well as one extractive and one abstractive summary of the given article. They were then asked to answer questions related to the fluency, adequacy and simplicity (Wubben, Bosch, and Krahmer, 2012) of each summary.

Fluency is defined as the extent to which a summary contains proper grammatical sentences. *Adequacy* is defined as the extent to which a summary conveys the same meaning as the source document. Finally, *simplicity* is defined as the extent to which a summary is easy to understand.

The questions were posed as follows:

(a) *The summary contains grammatically correct sentences*

(b) *The meaning of the summary conforms to the meaning of the original text*

(c) *All the important information of the original text is contained in the summary*

(d) *The summary contains superfluous information*

(e) *The summary contains words that do not fit the context*

(f) *The summary is easy to understand*

Question (a) was intended to account for *fluency*, (b)–(e) for *adequacy*, and (f) for *simplicity*. The answers for questions (d) and (e) were reversed on the 5-point Likert scale to facilitate the analysis and so that 5 had a positive connotation like in the other questions. Each question was assessed using a 5-point Likert scale ranging from *Strongly Disagree* (1) to *Strongly Agree* (5).

After reading and assessing each of the three summaries, the participants were asked which of the summaries they found to be the best (*Summary 1/Summary 2/No Difference*), and they were also asked to give an explanation of their reply in a free-text field.

Analysis

A statistical analysis was done on the data collected from the survey. A Wilcoxon signed-rank test was used on each category to compare the differences between extractive and abstractive summaries regarding fluency, adequacy, and simplicity.

The free-text answers motivating the choices of the best summary were analysed by reading them through and searching for overarching themes.

Results

In Table 11.8 mean values and standard deviations are presented for all questions and in Table 11.9 test statistics for all questions are presented.

Table 11.8: Mean values (M) and standard deviations (within parenthesis) for extractive (Ext) and abstractive (Abs) summaries on all questions. The adequacy questions are presented separately and combined.

Question	Ext M	Abs M
(a)	4.27 (0.953)	3.98 (1.070)
(b)	3.72 (0.962)	2.51 (1.242)
(c)	3.15 (1.169)	2.40 (1.154)
(d)	3.98 (1.144)	3.46 (1.242)
(e)	4.26 (1.059)	3.98 (1.191)
(b)–(e)	3.78 (0.817)	3.09 (0.862)
(f)	3.55 (1.241)	3.41 (1.232)

Table 11.9: Test statistics for all questions between extractive and abstractive summaries. Statistics for the adequacy questions are presented both separately and combined.

Question	W(111)	Cohen's d
(a)	577*	0.271
(b)	465**	0.751
(c)	841**	0.506
(d)	582**	0.359
(e)	523*	0.209
(b)–(e)	900**	0.645
(f)	1037***	0.094

** p<0.05, ** p<0.001 ***ns*

There was a statistically significant difference between the perceived fluency in extractive (M = 4.27, SD = 0.953) and abstractive (M = 3.98, SD = 1.070) summaries, $W(111) = 577$, $p < .05$, with a small to medium effect size ($d = 0.271$). In Figure 11.1, this difference is illustrated.

As given by Tables 11.8 and 11.9, statistically significant differences were found between extractive and abstractive summaries regarding perceived adequacy in all four questions (extractive summaries having higher adequacy). When combining these four questions by averaging the scores, the difference was still statistically significant.

The difference in adequacy between extractive and abstractive summaries is illustrated by Figure 11.2. The plot shows the combined adequacy measure, i.e. the average of questions (b)–(e).

Regarding simplicity, no statistically significant difference between extractive (M = 3.55, SD = 1.241) and abstractive (M = 3.41, SD = 1.232) sum-

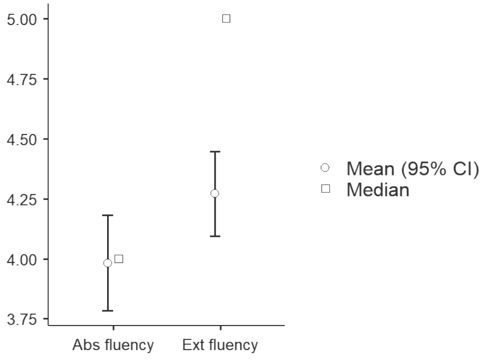


Figure 11.1: Difference in fluency between extractive and abstractive summaries.

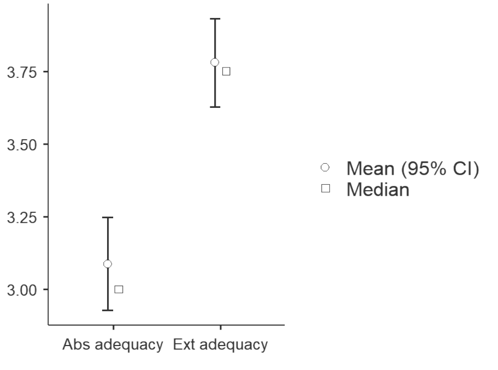


Figure 11.2: Difference in adequacy between extractive and abstractive summaries.

maries was found, $W(111) = 1037$, $p = 0.394$ ($d = 0.094$). The difference is illustrated in Figure 11.3.

On the question of which summary the participant regarded to have better quality, the extractive summary was preferred in 74 cases, the abstractive summary in 28 cases, and in 9 cases, there was no perceived difference in terms of their quality.

A few distinct and recurring themes stood out in the free-text answers. The most common reasons for choosing the extractive summary as the preferred summary were that the extractive summary was more in line with the facts presented in the article and that the abstractive summaries contained incorrect facts. The main reason for choosing the abstractive summary as the

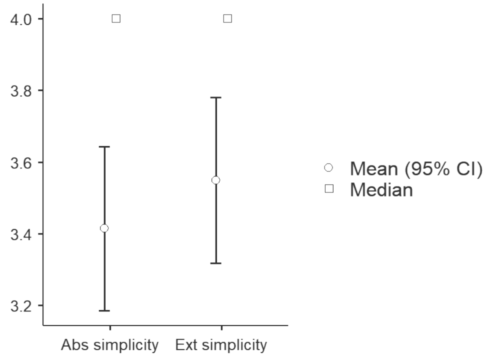


Figure 11.3: Difference in simplicity between extractive and abstractive summaries.

preferred summary was that the extractive summary was poorly structured and therefore hard to follow, or that it lacked essential information.

Discussion of Results

The survey indicated that there were some differences between the summaries regarding perceived text quality and readability.

Fluency

Extractive summaries scored higher at perceived *fluency* than abstractive summaries. It is not surprising that the extractive summaries are perceived as more fluent (i.e. more grammatical) as complete sentences are extracted from the (presumably) well-formed text. Abstractive summaries, on the other hand, generate completely new text segments which might have an effect on the perceived fluency.

Adequacy

The extractive summaries also scored higher on perceived *adequacy*. The largest effect size was found in question (b), whether the meaning of the summary conformed to the meaning of the original text. This is intuitive, as abstractive summaries are prone to introduce factual errors, resulting in a text that might deviate from the original in terms of content.

A medium effect size was present in question (c), whether all the important information of the original text was contained in the summary. One interesting finding here is that, although the difference was statistically significant, this question revealed the lowest mean value for both the extractive and abstractive summaries. This indicates that both summary types had problems preserving content. However, this could also be a factor of summary length, and it is possible that the scores would have been higher if we had allowed the summaries to be longer. We found a small to medium effect size for question (d), whether the summary contained superfluous information. The relatively high mean values for both summary types indicate that superfluous information was not common in any of the summary types.

Question (e), whether the summary contained words that did not fit the context, revealed a small effect size, which indicates that abstractive summaries are more likely to contain words that do not fit the context. This can, again, be explained by the fact that extractive summaries are derived from texts written by human writers and no new words are introduced in the summary, whereas the abstractive summaries are generated and may introduced previously unseen words. The difference was statistically significant, but the mean scores were low for both summary types, which indicates that summaries rarely contained words that did not fit into the context.

Simplicity

No statistically significant difference was found regarding perceived *simplicity*. Both summary types scored above 3, meaning that they can be regarded as relatively simple. This could mean that both types of automatic summarisation could be good candidates for producing more accessible text.

General Preference

We noted a general preference for the extractive summaries. This could be explained in part by the factual errors sometimes introduced by the abstractive summarisation model, see for instance the abstractive summary sentence *It is now the Russian hockey league,* ***NHL****, ...* in Appendix A. This is supported by the free-text answers.

We also observed that the extractive summaries were sometimes considered to be poorly structured or missing vital information. This connects to the study described earlier in this chapter and the cohesion errors that extractive summaries are prone to produce. For instance, one extractive summary used in the survey began with the sentence *"At the same time, the Iraqi government has..."*, see also the first extractive summarisation sentence in Appendix A. Compared to factual errors, these cohesion-related errors were fewer and did not seem to matter as much in most cases.

In the end, both summary types received relatively high scores on most aspects, and only two questions had an average score below 3.0 for the abstractive summaries. This indicates that the summaries were generally considered to have good quality. Thus, automatic text summarisation seems like a promising technique for provision of adapted text for end-users.

11.3 Chapter Summary

In this chapter, extractive and abstractive automatic text summarisation were assessed as Easy Language adaptation techniques. We explored cohesion errors made by an extractive summariser and their effect on reading, and found that cohesion errors affect the experience of reading a summary negatively, and that participants experienced difficulties particularly when the extractive summary resulted in missing context or when the anaphoric expression, a pronoun, referred to a missing antecedent.

We then compared the perceived readability and text quality in an extractive and an abstractive summarisation system by asking participants to rate the summary outputs according to fluency, adequacy, and simplicity, and found that extractive summaries scored higher regarding fluency and adequacy, and that the extractive summaries were preferred overall. The general preference for extractive summaries seems to be explained in part by the factual errors introduced by the abstractive summariser.

Summarising a text can be a way of adapting the macrostructure of a text (as described in Chapter 7). It is often described as a good way to help poor readers get a quick overview of the key content in a text. Automatic text summarisation is implementable and often gives relatively good results, but there are some issues that should be considered when using automatic text summarisation as an adaptation for Easy Language. Extractive summarisation techniques shorten the text by removing information, which could result in reduced cohesion. Abstractive summarisation techniques, on the other hand, may introduce factual errors into the text. The studies described in this chapter indicate that extractive and abstractive summaries are similarly simple (or difficult), but that extractive summaries are preferred in other ways. Moreover, in the first study, some participants said that they would be more tolerant with the texts if they knew that they were automatically produced. This could indicate that erroneous summaries could be accepted to some extent, but this remains to be explored.

However, the evaluations were conducted on participants without any reading disability, and it is viable that erroneous summaries could have a larger impact when the reader is struggling with reading comprehension. For this reason, it is important to perform additional testing with participants from the various target audiences in order to reveal the severity of such errors for this group of people.

CHAPTER 12

Tools and Services

The language technology research conducted within the scope of this thesis, and in surrounding projects, has been implemented in various tools for making texts easier to understand or to aid professional writers in their Easy Language writing process. The large variety in experienced reading comprehension issues and challenges led us to believe that the one-size-fits-all approach to automatic text adaptation is not the best way forward. Thus, our idea is to include our available text adaptation techniques in the end user tools, and let the needs and demands of the user decide which adaptations to apply.

As of today, there are two main end user tools: (FriendlyReader and TeCST) and three back end tools (SAPIS, CogSum and JuliuSum). A few modules handling various types of linguistic adaptations and text complexity measurement are integrated in SAPIS.

12.1 Back End Tools

There are three back end tools. SAPIS is an API enabling access to tools and techniques developed within our research projects. CogSum and JuliuSum are tools for extractive and abstractive text summarisation.

SAPIS

SAPIS[1], illustrated in Figure 12.2 is an API service for distribution of the developed tools and techniques. It is a RESTful web service, now based on Python3.

[1]StilLett SCREAM API Service. The middle part of the name, API, refers to the service being an API, whereas the two S's refer to StilLett and SCREAM.

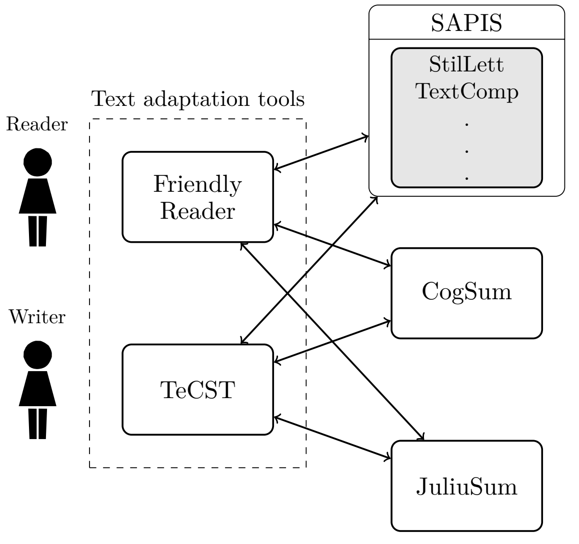


Figure 12.1: The integration of modules in the two different automatic text adaptation systems.

SAPIS interprets options and input data as variables in an input JSON object, which is passed to the SAPIS service in a HTTP request. A client can specify which of the services to run by passing instructions with the options variable in the input JSON object. The resulting metrics for text analysis and simplification suggestions are merged and returned to the client as one JSON object.

The modules developed are StilLett (automatic text simplification) and TextComp (text complexity measurement). Each module is briefly described in this section.

StilLett

StilLett (Rennes and Jönsson, 2015) is a module for rule-based syntactic text simplification. The first implementation of StilLett, further described in Chapter 10 was a Java application, partly built on CogFlux (Rybing, Smith, and Silvervarg, 2010).

In its original implementation, StilLett included rules for rewriting to passive-to-active, quotation inversion, rearranging to straight word order, sentence split, and synonym replacement, in addition to the original rule sets proposed by Decker (2003). The synonym replacement module implemented in StilLett was originally developed by Abrahamsson (2011), and combined

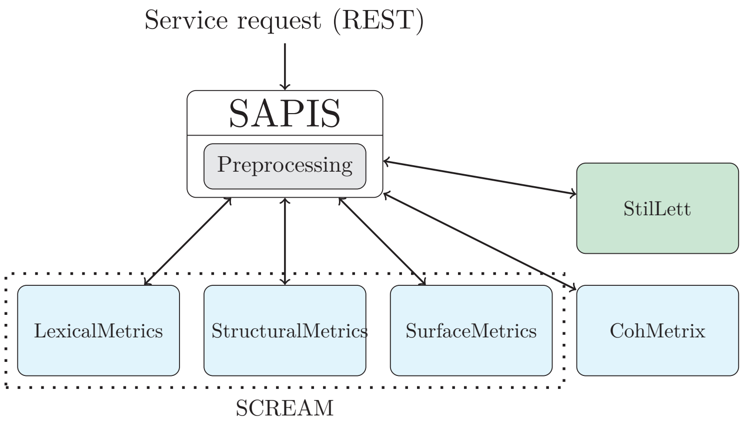


Figure 12.2: SAPIS architecture. Blocks in blue are concerned with text complexity calculations and the box in green is concerned with automatic text simplification.

the word pairs from the Synlex lexicon (Kann, 2004) and frequency information.

StilLett has undergone several improvements (see for instance Johansson (2017)) since the first implementation. Today, StilLett is based on Python3, and uses dep_tregex[2] (Dvorkovich, Gubanov, and Galinskaya, 2016), for reordering the dependency trees.

The included rule set still contains the original rules for rewriting to passive-to-active, quotation inversion, rearranging to straight word order, and sentence split, but the rules are further refined.

The pre-processor is accessed through SAPIS and runs the Swedish pipeline with Efselab (Östling, 2018) and MaltParser (Nivre, Hall, Nilsson, Chanev, Eryigit, Kübler, Marinov, and Marsi, 2007) version 1.9.0.

TextComp

TextComp is a collection of text complexity measures. The main part of the included measures consists of the SCREAM (Swedish Compound READability Metric) (Sjöholm, 2012) features, but other text complexity features have been included.

SCREAM provides text complexity features. The features are described in Chapter 6 and the complete list of SCREAM features can be found in Falkenjack (2018b) and Falkenjack, Heimann Mühlenbock, and Jönsson (2013). The implemented SCREAM services are:

[2]https://github.com/yandex/dep_tregex

- SurfaceMetrics includes measures that can be calculated after tokenisation. This includes LIX, OVIX, Nominal ratio, Average sentence length and Average word length.
- LexicalMetrics includes measures that can be calculated after lemmatisation. This includes a categorised frequency analysis from word occurrences in the basic Swedish vocabulary SweVoc (Heimann Mühlenbock and Johansson Kokkinakis, 2012b).
- StructuralMetrics includes measures that can be calculated after parsing. This includes the syntactic and morpho-syntactic features described in Falkenjack, Heimann Mühlenbock, and Jönsson (2013).

Except for the SCREAM features, there is one supplementary set of text complexity measures. Coh-Metrix is the latest addition to the TextComp module, and this service includes cohesion measures from Coh-Metrix (Graesser, McNamara, Louwerse, and Cai, 2004a), translated to Swedish. This work is ongoing, and measures included as of today are listed in Appendix B.

CogSum

CogSum (Smith and Jönsson, 2011a; Smith and Jönsson, 2011b) is an automatic extractive summariser, which means that it extracts the most important sentences in order to create a shorter version of the text. CogSum uses the Random Indexing (RI) (Hassel, 2007; Hassel, 2011) word space model with pre-trained word vectors, and a modified version of the PageRank algorithm to rank the sentences (Chatterjee and Mohan, 2007). Evaluations have shown that CogSum performs at an average ROUGE-1 score of 0.6.

JuliuSum

JuliuSum (Monsen and Jönsson, 2021) is an automatic abstractive summariser. JuliuSum was trained utilising the methodology proposed by Rothe, Narayan, and Severyn (2020), using a pre-trained Swedish BERT model (Malmsten, Börjeson, and Haffenden, 2020) to warm-start an encoder-decoder model. The data used for training consisted of news articles published in Sweden's largest morning newspaper *Dagens Nyheter (DN)* during the years 2000–2020. JuliuSum was evaluated and compared to an extractive equivalent in Chapter 11.

12.2 End User Tools

There are two front-end solutions utilising the functionality of the developed back end tools. The tools differ mainly in that they are intended for different user groups. FriendlyReader is a tool for readers in need of Easy Language adaptations, and TeCST is a tool for supporting writers in their Easy Language writing process.

FriendlyReader

FriendlyReader[3] is a tool targeting end users, i.e. readers in need of Easy Language text. The idea is that the tool should contain the entire palette of adaptation techniques, including both linguistic adaptations as well as adaptations related to text layout and design, and that the user can adapt the text completely to their individual needs.

FriendlyReader is under constant development, but as of March 2022, the tool consists of two summarisation modules (CogSum and JuliuSum), one simplification module (StilLett), and the set of text complexity measures provided in SAPIS (TextComp).

Except for the modules for linguistic adaptation, FriendlyReader also contains text-to-speech functionality, which lets the reader listen to the text. The adaptations related to text layout and design are the possibility to change font size, line spacing, font and line length.

Layout

FriendlyReader has undergone several design procedures, but in its current state, the user pastes the text into a large text field and presses *Run*. The view in Figure 12.3 is then presented to the user. The layout consists of three parts. The main field is the middle field, where the text is presented to the reader.

The left-hand side contains a menu with various types of text adaptations. The user is presented with a number of options:

1. **Summarise**: The user can summarise the text using a slider that outputs summaries of different lengths.

2. **Simplify**: The user can simplify the text using the syntactic simplification operations of StilLett. There are check boxes that lets the user choose what operations to make, and the rules are applied directly to the text.

3. **Synonyms**: By clicking Synonyms, the user can activate the exhibition of synonyms of words in the text. Words with available synonyms are highlighted in the text, and by clicking any such word, the user is presented to a list of possible synonyms.

[3] *FriendlyReader* was the former name of the extractive summariser now called *CogSum*. This might be slightly confusing to any reader reading old publications describing the systems.

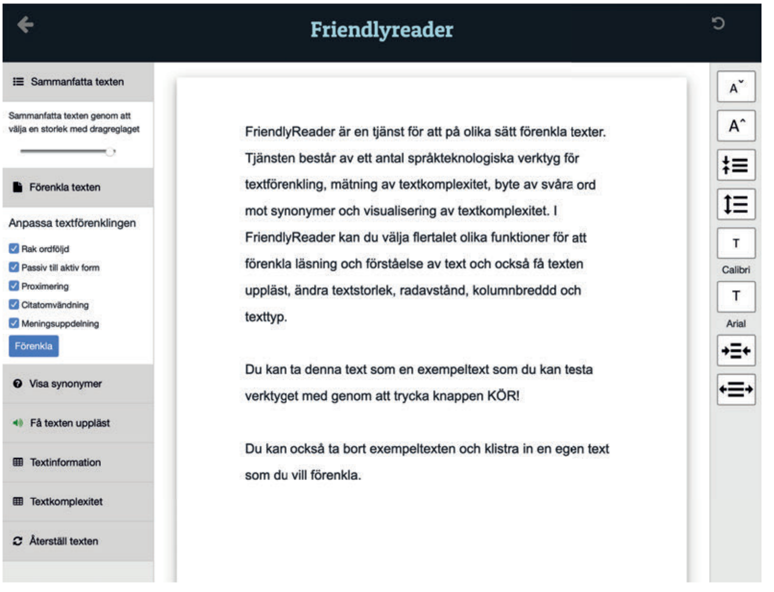


Figure 12.3: FriendlyReader

4. **Text-to-speech**: The user can have the text read out loud by activating the text-to-speech functionality.

5. **Text complexity**: The user can see basic text complexity measures, such as LIX and OVIX. The user is also presented to a visualisation of the complexity of the text presented in a radar chart.

The right-hand side contains a menu with various adaptation options related to text design. The user can change the font size, increase or decrease line spacing, choose between two different fonts, and alter the margin width so that the text line length is increased or decreased.

TeCST

TeCST (Text Complexity and Simplification Toolkit) is a web-based tool developed for web editors and writers of Easy Language texts, but could be used by anyone interested in calculating the complexity of a text, as well as applying various text adaptation techniques. The intuition behind this tool is that providing the Easy Language text writers with advanced techniques for measuring and visualising complexity, identifying complex linguistic structures, and give advice on how such structures should be adapted to suit the needs of various target audiences, is one way of making the text adaptation process quicker and cheaper, without overlooking the expertise and unique competence provided by the human writer.

As of November 2021, TeCST consists of two summarisation modules (CogSum and JuliuSum), one simplification module (StilLett), and subset of text complexity measures provided by SAPIS (SCREAM and CohMetrix).

Layout

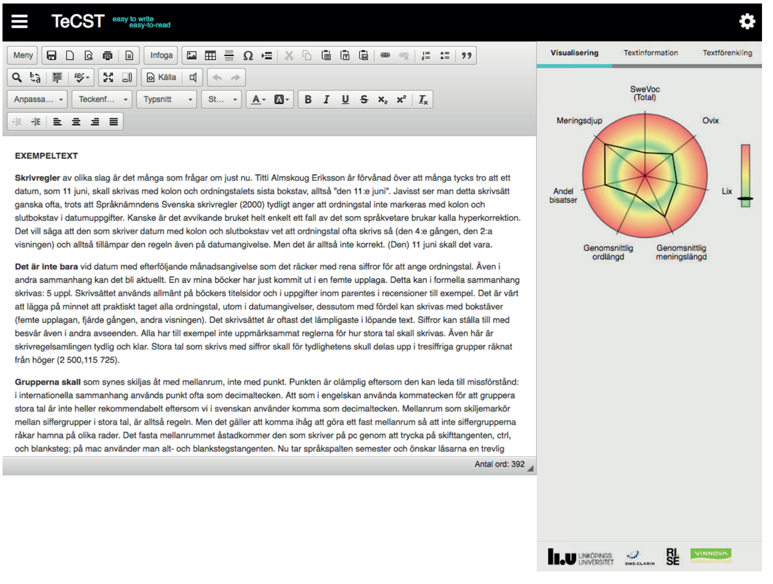


Figure 12.4: TeCST

The TeCST layout, presented in Figure 12.4, consists of two fields: the editor, which makes up the main part of the tool layout, and the adaptation and visualisation field. The editor allows the writer to customise the text using different fonts, font sizes, bold face, bullet point lists, and similar features often included in text editing tools. The adaptation and visualisation field, on the right-hand side, presents information regarding the current complexity and adaptation suggestions of the text. It has three tabs: *visualisation*, *text information* and *text simplification*.

In the visualisation tab, a text complexity visualisation (described in depth in Chapter 8, Section 8.2) in the form of a radar chart is presented. The different axes of the chart are replaceable with any text complexity measure available in SAPIS.

In the text information tab, the writer can choose to see a summary of the text, as well as some general information about the current text, such as the text length in words and sentences, as well as a subset of the text complexity measures. Similarly to the features presented in the visualisation, the subset of text complexity features shown under the text information tab is customisable.

The third tab, text simplification, allows the writer to make adaptations to the text. There are four options here.

1. **Summarisation**: The user can summarise the text, by the use of a slider that regulates the length of the resulting summary.
2. **Synonyms**: The user can use a check box to highlight the words of the text that have available synonyms, and customise the synonym replacement functionality to mark long words, i.e. words longer than some length chosen by the user.
3. **Markings**: The user can use check boxes to let the tool identify and highlight different features of the text, such as long words, long sentences, and numbers. The number of characters that make up a long word is customisable, as well as the number of words that make up a long sentence.
4. **Text simplification suggestions**: The user can get suggested simplifications of complex sentences. The simplification module identifies sentences that are flagged by any of the rules of the StilLett rule set and suggest a rewritten sentence. There is an option that lets the user choose which rules to include in the simplification suggestion check.

12.3 Chapter Summary

The variety in experienced reading comprehension issues and challenges made us believe that the one-size-fits-all approach to automatic text adaptation is not the best solution. For this reason, we have chosen to develop a palette of tools and techniques that can be combined to make texts easier to understand or to aid professional writers in their Easy Language writing process. Our idea is that the inclusion of text adaptation techniques in end user tools will allow for individual customisation of the tools.

Our research on automatic text adaptation has resulted in a number of tools and services, described in this chapter. There are two main tools targeting end users, FriendlyReader and TeCST, where the former targets readers of Easy Language texts, and the latter targets writers of Easy Language texts.

There are three back end tools. SAPIS is a REST API aiming to make the services readily available. SAPIS includes modules for calculating a large set of text complexity features, as well as giving text simplification suggestions generated by StilLett. CogSum and JuliuSum are the extractive and abstractive summarisation systems developed within our projects.

These tools and services will be further developed and evaluated. Although some work has already been done on enhancing the usability of FriendyReader and TeCST, they remain to be tested on readers from various reader audiences, and professional writers, in order to be further adapted to the needs and wants of the users.

Part IV

Discussion and Conclusion

CHAPTER 13

Summary of Thesis Work

This thesis is long and comprises many different pieces of work. This chapter intends to summarise the thesis to provide a more condensed overview of the work.

13.1 Part I

This part provided the theoretical background relevant to the thesis.

Chapter 2

We began by describing the needs of the various target audiences in need of adapted Easy Language texts. The chapter described the diversity of experienced difficulties associated with each reader audience, and concluded that these differences, as well as differences within the reader audiences, must be accounted for in the text adaptation process.

Chapter 3

In this chapter, we described the characteristics of *Easy Language*, based on the advice given by various initiatives that provide guidelines for writing Easy Language texts. We also reviewed what is known about the effect of Easy Language texts on poor readers.

Chapter 4

In this chapter, we reviewed earlier research on automatic text simplification and automatic text summarisation. We concluded that automatic text adaptation techniques often claim a reader audience perspective, but that the approaches are often general in their nature.

Chapter 5

In this chapter, we discussed the role of corpora in text adaptation research, and then described the main resources available today, with special focus on corpora for automatic text simplification. Available resources for Swedish text adaptation were described, as well as corpora targeting different reader audiences.

Chapter 6

This chapter introduced text complexity assessment, a research field closely related to automatic text adaptation and Easy Language. The most prominent readability measures were described, as well as modern data-driven text complexity features, and earlier efforts on visualising text complexity.

13.2 Part II

This part explored Easy Language texts from various perspectives.

Chapter 7

In this chapter, we mapped available Easy Language guidelines for Swedish and English, and their connection to a theoretical framework of reading comprehension. By adopting this perspective, we put more emphasis on the actual strengths and weaknesses of the individual readers, and that we must take the various (dis)abilities into account when adapting text automatically. We saw that most guidelines could be connected to the different levels of the model of reading comprehension and that guidelines were represented at all levels of the model, but that there were guidelines that were either vague or relying on the common sense of the writer. Such guidelines are for obvious reasons very difficult to implement in an automatic adaptation system. Other guidelines that were found to be more challenging to implement concerned, for instance, clarification of connections between sentences. In this chapter, we also concluded that, on the basis of the literature review on the topic, we still do not know much about the actual effect that linguistic text adaptations for Easy Language text have on reading comprehension for poor readers, and if this effect would hold for all types of readers. More research is needed in order to gain insights in reader-specific challenges connected to text comprehension.

Chapter 8

It is possible to capture numerous aspects of the characteristics by the use of text complexity features, but since such features often are derived from computational models, they are not always comprehensible for the general public. If we want text complexity assessment methods to be useful for an end user, whether it is a consumer of Easy Language text or a producer of Easy Language text, the features must be represented in a more comprehensible way. This thesis has approached this problem from two ways: by clustering individual text complexity features into larger chunks of features, and visualising them in radar diagrams.

In this chapter, we showed that a component-based text complexity analysis can be used to classify texts in genres, and might say something about a text's complexity. We visualised text complexity facets using radar diagrams, and argued that such visualisation could add to the understanding of text characteristics by the use of shapes corresponding to combinations of text complexity facets.

Chapter 9

This chapter described our work on collecting a corpus that would fill a void in Swedish text adaptation research: an aligned resource of parallel complex-simple sentence pairs. The chapter described the implementation of two corpus collection procedures and tested various algorithms for sentence alignment. A corpus comprising text from websites of Swedish public authorities and municipalities and their Easy Language counterparts was collected and aligned at the sentence level using an algorithm which combined similarities between word vectors into a sentence-similarity measure indicating whether the sentence pairs were semantically similar or not. As the sentence pairs were extracted from this corpus of standard and simple documents, we assumed that the corpus comprised good representatives of standard and Easy Language sentences. The corpus of web texts from Swedish public authorities and municipalities was the first corpus of its kind for Swedish. We found that the resulting corpus turned out to be unbalanced as well as limited in size, and thus, not ideal for input for data-driven text simplification or adaptation, but it has been used in other projects for comparison of text categorisation feature representations and text classification.

We also presented NyLLex, a novel lexical resource extracted from Easy Language literature. The lexicon is annotated with reading proficiency levels and can be used in, for instance, applications for synonym replacement or text complexity assessment.

13.3 Part III

This part described the different text adaptation techniques explored in our research projects.

Chapter 10

This chapter treated the development of techniques for automatic text adaptation. In this chapter, a platform for rule-based syntactic simplification was presented, including rules for rewriting from passive to active tense, quotation inversion, rearrangement to straight word order, and splitting a long and complex sentence into two. The chapter also described various methods for lexical simplification. The first methods made use of a bilingual dictionary for extracting more comprehensible synonyms from an Easy Language corpus, and another approach traversed word hierarchies upwards in order to find near-synonymous words that exhibited traits of being at the basic level.

Chapter 11

This chapter described our work on evaluating automatic text summarisation systems as Easy Language text adaptation techniques. First, we explored cohesion errors made by an extractive summariser and found that cohesion errors do affect the experience of reading a summary negatively, and that the types of errors which were considered particularly difficult regarded missing context or a missing antecedent to an anaphoric expression. Second, we compared the perceived readability and text quality in an extractive and an abstractive summarisation, and found that extractive summaries scored higher regarding fluency and adequacy, and that the extractive summaries were overall preferred by the participants.

Chapter 12

The adaptation techniques conducted within the scope of this doctoral thesis were implemented in a number of tools. This palette of tools is described in this chapter. There are two main tools targeting end users (readers and writers): FriendyReader and TeCST. There are three back end tools. SAPIS is a REST API aiming to make the services readily available. SAPIS calculates a large set of text complexity features, and gives text simplification suggestions generated by StilLett. CogSum and JuliuSum are the extractive and abstractive summarisation systems developed within our projects.

CHAPTER 14

Research Approach

The work presented in this thesis is oriented with the approaches of the technological sciences.

14.1 Technological Science

Technological science aims to produce *new and better artefacts* (Solheim and Stølen, 2007). Technological science differs from the traditional scientific approaches (Hansson, 2007). The perhaps most prominent difference is that the study objects of the technological sciences are human-made objects, rather than natural objects, but other differences are that the design and construction could be a part of the scientific process. In technological sciences, the study objects tend to be defined on the basis of their function, and—as a consequence of this—evaluated according to function. For instance, in order to define a spoon, we would need to figure out whether or not it has the function of a spoon. How good of a spoon it is would be determined by evaluating how well the spoon satisfies the expectations we have of a spoon. The aim of the technological sciences, thus, does not need to be reaching a truth value, but could be, if we hold on to our example, to construct a better spoon. Regarding the theories, models and explanations of the technological sciences, there are two main characteristics that distinguish technological sciences from other types of sciences (Hansson, 2007). First, the models of natural sciences are built to study the phenomenon separately from surrounding factors. Second, the technological sciences do not necessarily seek the ideal mathematical solution, but could do with a solution that is a sufficiently good approximation.

When the classical science intends to gain knowledge about the world, the hypothesis of technological science is whether or not the artefact satisfies the need (Solheim and Stølen, 2007). The research process is cyclical and consists of three stages. First, there is the *problem analysis stage* establishing what is the potential need. Then, there is the *innovation stage*: how can we make an artefact that satisfies the need? The third step is the *evaluation stage*: how can we show that the artefact satisfies the need?

The *problem analysis stage* maps the needs for an improved artefact. As described in Chapter 1 and Chapter 3, there is a clear motivation for why we need systems for automatic text adaptation. Previous research in the area has either been conducted on languages other than Swedish, or adopted either a very narrow or very general approach, from the target audience perspective. Knowledge about the differences of the various groups of poor readers would help developing a better artefact. The *innovation stage* is concerned with how the artefact can be constructed so that it satisfies the need. This stage can essentially be mapped to the second research question of the thesis and in this thesis, it is approached mainly by a literature survey on reading comprehension for various audiences of Easy Language texts. It was also assumed that previous knowledge about how Easy Language texts have been written in the past would provide a good source of knowledge about how such artefacts should be developed, which motivated the need for consulting corpora of available Easy Language texts, and, when such resources were not available: collecting our own corpus. The *evaluation stage* aims to control whether the predictions about the artefact are true. In this thesis, this is approached by evaluating the various subparts of the artefact, such as how summary output and synonym candidates are conceived by readers.

14.2 Alternative Method

Natural language processing is a research field that combines linguistics with computer science and artificial intelligence, and can be approached from different directions. Approaching automatic text adaptation through the traditions of theoretical computer science would entail, for instance, applying a pure machine learning perspective on the task, under the assumption that everything which we need to know to solve this task can be derived from data. The work would then be concerned with enhancing the model performance. Although such methods are common in modern natural language processing, this would require large amounts of high-quality Easy Language data, which—at the time—is not available for Swedish. Another drawback of applying such approaches is that the model performance is dependent on the data it has been trained on. On the basis of the assumption that the various reader audiences have different needs and challenges, this would require data of the different types of texts, and finding such reader-specific data is not plausible. The available corpora of Easy Language texts often adopt a broader approach to cover as many audiences as possible, and the models trained on such data will be equally broad. Although this could be considered *good-enough*, this thesis seeks to innovate an artefact which could be adapted to a specific reader audience, or even a specific reader.

14.3 An Ethical and Sustainable Perspective on Automatic Text Adaptation

Developing techniques for ensuring quick access to Easy Language texts is motivated by, for example, the UN Convention on the Rights of Persons with Disabilities, Article 21. However, there are some considerations concerning ethical and sustainable values that should be addressed and weighed against the benefit of increased inclusion.

Correctness of Information

When a source text is adapted to an Easy Language version, it is inevitable that text is changed, moved or removed. As a result of this, there is a risk that information is lost or altered. In the normal case, this gives rise to relatively harmless mistakes, but there is obviously a risk that erroneous information—or information loss—can be truly problematic. How can we make sure that the information assimilated by the reader is correct and that the semantic content is preserved? This issue has been indirectly addressed in the thesis. In the evaluations of automatic text summarisation systems presented in Chapter 11, we investigated 1) how various types of cohesion errors affected reading, and 2) how well extractive and abstractive summarisation systems preserve content. The question of meaning preservation is important to consider when evaluating the performance of the systems in future work. It is possible that the more conservative adaptations are suitable for a reader tool, such as FriendlyReader, whereas an adaptation tool for producing Easy Language text, such as TeCST, could benefit from adaptations that make more changes to the texts, as the adaptations are then applied under the supervision of professional writers.

Informed Consent

Some of our studies have included user evaluations and experiments with human participants. However, the evaluations and experiments have never requested any sensitive personal information. The participants have never been asked to submit any information about ethnicity, political opinions, religious or political beliefs or health. No economic compensation was offered to the participants in any of the studies. In the case where we requested information beyond rating of text, for instance in the eye tracking experiment performed in Rennes and Jönsson (2014) and presented in Chapter 11, the participants were informed prior to the experiment that the participation was completely voluntary, that they were going to be anonymous and that they were allowed to terminate the experiment if they did not want to continue. In the studies where we asked participants to rate aligned pairs of sentences or evaluate synonyms, the forms were created so that the participants were anonymous and able to cancel the evaluation at any point.

Sustainability

The issue of sustainability in natural language processing is gaining much attention lately. Modern artificial intelligence techniques used in natural language processing, such as deep learning, require large amounts of data and computational power, and the trend of constantly improving benchmark performance of language models has an inevitable environmental impact. The AI field as a whole is growing rapidly, and the carbon footprint is a significant issue which cannot be ignored. In the light of this, rule-based approaches could be considered as an alternative for training large-scale models, especially for well-delimited tasks with limited data resources. If rule-based approaches perform results that are good-enough for the intended application, they could provide a satisfactory alternative to large-scale models, with lower carbon emissions.

The projects included in this thesis did not themselves use massive model training, but the automatic text summarisation models described in Chapter 11 did use such models, trained in other projects, and fine-tuned to suit our purposes. The computations carried out in the reported projects, for instance the process of aligning sentence pairs, were comparatively computationally cheap.

Evaluation on Target Audiences

The automatic text adaptation systems have not yet been evaluated on readers with reading impairment. However, we are planning on doing this, and have considered several ethical aspects. First, it is essential to ensure the integrity of the participants, as the study might contain sensitive data about, for instance, the participants' disabilities. To handle this, we will ensure anonymity in various ways. Only the researchers of the project will be involved in collection, analysis and storage of data, and the results will be presented at statistical group level. In the case where individual responses from, for instance, interviews are reported, the responses will be deidentified so that they cannot be connected to any person.

Every participant will be assigned a unique ID including numbers representing information about which group the participant belongs to, which study the participant belongs to, and a sequence number to represent the individual. This ID will be the only identification method used for all tests and interviews. The code key will be stored in a locked cabinet, and only accessed by the researcher responsible for the experiments. The code key will be stored separately from the data. The interviews will be recorded and transcribed, and the sound recordings will be deleted after transcription.

There is a risk that the texts or the tasks will be experienced to be difficult, which by extension could lead to negative emotions towards oneself. This will be handled by a number of measures. For instance, the tests will have stop criteria to ensure that the tests are interrupted if the participants experience them to be too challenging. Furthermore, it will be clearly stated that the aim of the tests is not to assess the performance of the participants, but rather to find what they find challenging in order to develop better adaptation techniques.

14.4 Chapter Summary

Returning to the key words and phrases of the UN Convention on the Rights of Persons with Disabilities, Article 21: *disabilities*, *accessible formats*, *technologies*, and *timely manner and without additional cost*, it can be concluded that techniques for automatic text adaptation for creating Easy Language texts do serve an important role in the mission of increasing inclusion by creating easily accessed information. To be able to read is important, both at a societal level, as well as for the individual, and the benefits that the techniques can provide should be appreciated and exploited. However, it is similarly important that the work we do to ensure such adaptation techniques are performed with care for the ethical and sustainable values presented above. The conducted research should not interfere with the individual's right to integrity, or cause any negative emotions towards the abilities of the individual. The research should also be carefully planned not to use models or techniques that are computationally expensive beyond the actual needs and requirements of the application.

CHAPTER 15

Results and Discussion

This chapter summarises the research conducted within the scope of the thesis, and discusses the main contributions in relation to the research questions. Finally, possible ways forward are discussed.

15.1 Revisiting the Research Questions

This doctoral thesis was built around two main research questions:

RQ1 What linguistic adaptations are needed in an automatic adaptation system for simple Swedish?

RQ2 How can automatic text adaptation be implemented and conceived in order to meet the needs of different target audiences?

Answering RQ1

Regarding the first research question (**RQ1**), there are several subquestions: Which are the linguistic adaptations available? Which of the adaptations are needed? Which of the available linguistic adaptations are feasible to implement in an automatic implementation? These questions were mainly addressed in Chapter 7, where Easy Language guidelines were reviewed and connected to a model of reading comprehension. The results from this work were:

- The description of available guidelines recommended for writing Easy Language texts, and the mapping of the guidelines to a theoretical model of reading comprehension.

- The review of the current state of knowledge regarding the impact of such guidelines on the reading comprehension of poor readers.

Another way of answering **RQ1** is to adopt a data-driven approach and consult texts written by professional writers of Easy Language texts. In Chapter 9, such corpora of Easy Language texts were collected in order to create a resource from which knowledge about Easy Language texts can be extracted. The results from this work were:

- The description of a method for collecting such corpora.
- The collection of a new resource, filling a void in Swedish text adaptation research.
- The testing and evaluation of different methods for aligning sentence pairs.

Answering RQ2

Based on our knowledge about the target audiences, the second research question (**RQ2**) aimed to find out how the adaptations are best realised. Answering this question requires us to know *who are the target audiences of adapted Easy Language texts?* and *what do the various reader audiences need in terms of adaptation operations?* These questions were addressed in Chapter 2 and Chapter 3 respectively. We found that:

- There are several groups of people that can and should be considered target audiences for adapted Easy Language texts: individuals with dyslexia, aphasia, intellectual disabilities, deaf and hard-of-hearing, second language learners, individuals with autism spectrum disorder, etc. Each group has unique challenges and difficulties that should be considered when working on automatic text adaptation targeting real readers.
- There is still knowledge missing about if and how text adaptations are useful for poor readers.

Another subquestion of **RQ2** could be how we know if we have successfully managed to adapt the texts. This could be in part answered by our work on making sense of the text complexity features. Text complexity assessment can be useful for getting a quick overview of a text, and could be used by, for instance, readers in need of Easy Language texts, or by teachers in a school context. However, to the common man, features of text complexity could be difficult to interpret, and we addressed this in our work on making text complexity comprehensible by clustering and visualisation in Chapter 8. The results from this work were:

- The grouping of text complexity features into components to reduce the number of complexity dimensions.
- The visualisation of text complexity using radar diagrams.

When researching these issues, we established that the capabilities and needs of the different reader groups vary greatly, and that the one-size-fits-all approach might not be the best way forward in this field of research. While more general guidelines can be useful to aid human writers in their writing process, the natural language processing solutions do not necessarily have to be delimited to the one-size-fits-all approach. By offering a wide range of adaptation solutions, it is possible to tailor the reading or writing tool to the actual reader audience, or even the specific individual. This was the reason for developing a palette of different text adaptation techniques, as described in Chapter 10–12. The tools we developed as a result of this were:

- FriendlyReader, an end-user tool for readers in need of adapted Easy Language texts.
- TeCST, an end-user tool for producers of Easy Language texts.

15.2 The Work in a Wider Context

This doctoral thesis has positioned the research field of automatic text adaptation for Easy Language texts in relation to a theoretical model of reading comprehension. This connection encourages a clearer structure in how adaptation research should be framed: around the needs and challenges of the targeted reader audience.

The perhaps most common approach for automatic text adaptation for creation of Easy Language text is to adopt the one-size-fits-all approach, even though many Easy Language guidelines state that it is vital to think about the targeted reader of a text. The audiences experience a variety of reading-related challenges, and this variation in experienced difficulties must be recognised in the development of adaptation techniques targeting specific audiences. The one-size-fits-all approach to adaption can be useful to some extent, and for some applications, but there will always be reader-specific issues that should not be neglected. Even within the target audiences, there are variations depending on the type of dyslexia, type of aphasia, age, first language, and severity of symptoms. In short: it is likely that the individuals of the different target audiences cannot be considered one large homogeneous group.

Studying the literature, we noted that there is still knowledge missing about if and how text adaptations are useful for poor readers. And this is where text adaptation research should start: investigating the impact of actual adaptation of texts targeting the intended reader audience.

The user-based perspective of automatic text adaptation has implications on various parts of the research field. One such example is the data we use for developing the models and techniques. Our model is only as good as our data, and if we want to consider different target audiences, we should make sure that the data that we use for constructing our model mirrors the characteristics of texts written for that specific target group. Adapting the general view of Easy Language as one single phenomenon will result in general models that do not take special needs into account. This might not be a problem! General models can be useful, but it is important be aware of the limitations this might imply. Another example is evaluation. Adopting a user-based perspective means that we must make sure that the evaluation methods correspond to the actual reading performance of that specific target group. Automatic metrics are useful for comparing systems, but in the end, we need to test the performance on the actual readers.

The adaptations implemented within the scope of this thesis are a few of many possible adaptations. Our way of dealing with the one-size-does-**not**-fit-all problem is to provide a buffet of tools and techniques, so that a tool can be tailor-made to suit a specific reader, reader audience, or writer. This is an ambitious goal, and the work on implementing new techniques is ongoing. The collection of possible adaptations discussed in Chapter 7 is a good starting point for adaptations to include in a system, but each adaptation (and combination of adaptations) should be carefully evaluated on real users.

This thesis described work on assembling resources for adaptation research. We explored various ways of compiling parallel corpora that can be used for automatic text adaptation, including implementation and evaluation of several methods for sentence alignment. The final resource was the first of its kind, and even though it turned out to be too limited in size to be truly useful for our purposes, the applied methods showed promising results. It is possible that the resulting resource would be larger if a similar project was conducted today, as the number of available websites with Easy Language material might be higher. The aligned corpus was used in other projects related to Easy Language and readability. We also compiled NyLLex, a lexical resource of words derived from Easy Language literature and annotated with six reading proficiency levels. This resource can have different applications, such as text complexity assessment or lexical simplification.

We presented work on how to make sense of text complexity features, by clustering and visualising them. These new ways of organising and visualising text complexity features could make it possible for readers and writers to get a more nuanced overview of the complexity of a text. The visual representation of text complexity facets could guide readers with different types of cognitive impairment to select the most adequate text for their specific reading needs. It could also be useful for teachers who search for texts with a specific linguistic profile, or producers of Easy Language texts who want to get a quick overview of the text they are writing.

15.3 Limitations of the Thesis Work

Although this doctoral thesis in many ways highlights the importance of taking the needs and challenges of target audiences into account, the developed tools and techniques have not yet reached this evaluation stage. Most adaptation techniques have been evaluated on readers without any kind of reading impairment, and the evaluation results should, thus, be interpreted with care. Another limitation is that the evaluations are often performed on a rather small group of participants, which further delimits the generalisability of the results. The adaptation techniques described in this thesis have been developed with focus on the technical challenges, and the tools and techniques have been immature for target audience evaluation.

A similar issue concerns the corpora used for development and evaluation of the various adaptation techniques: they are not specific to any reader audience. This is a problem, both for this project as well as for the field as a whole, and there is no easy solution. It is probably easier to gather audience-specific data for some of the reader audiences. For instance, compiling large-scale resources of texts written for children or second-language learners is probably more viable than constructing resources comprising texts for individuals with intellectual disabilities, simply due to the availability of such texts. For such groups, it is possible that the data-driven methods are not the best way forward. We have approached this issue from two angles. First, we have tried to use the second-best option: to compile and use resources of general Easy Language texts. Since such texts are assumed to be simple in a general way, they can probably go a long way. Second, we have planned to include the reader audiences in testing and development of techniques. At the time of writing this thesis, we are conducting evaluations of text adaptations for two of the target audiences: individuals with dyslexia and individuals with intellectual disability. Participants of the different audiences will be tested in a series of iterated studies, where the overall aim is to isolate the best adaptation strategies for each target audience, based on a theoretical knowledge of the audience-specific challenges as well as iterations of text comprehension tests.

15.4 Future Work

There are several research tracks to follow from here. Regarding the text adaptation techniques, this thesis has implemented a handful of adaptation techniques, but there are still many more ways in which text can be adapted, and more work needs to be done here. Except for the work on automatic text summarisation, this thesis has implemented text adaptations at the surface level of text comprehension (according to the CI model, described in Chapter 7), and there is still much work left to do on the other levels.

Moreover, we still do not have knowledge about how the adaptations should be combined in the best way and how they complement each other.

This is another interesting research track that should be addressed in future work. Studying the effect of simultaneous adaptations is rather complex, as it could be difficult to pinpoint the reasons behind a given outcome. One suggestion could be to study the different *levels* of adaptations, rather than single operations, by the use of, for instance, the levels of the CI model. Such an approach could throw light on the potency of text adaptations in relation to the various cognitive processes of reading.

This thesis has not included any evaluation of the complete adaptation systems. This is obviously an important issue, as we still do not know anything about how well the systems work. A common practice is to compare the performance of the simplification or summarisation system to similar systems by the use of automatic measures. This is obviously a good way to see how well the system performs in comparison with other systems, but it does not capture the subjective experience of the reader. In this thesis, it has been lifted that the reader cannot be removed from the equation when developing and evaluating adaptation techniques, and this stresses the relevance of including actual readers, with and without reading impairment, when evaluating the techniques. This kind of evaluation is expensive in terms of time and work effort, but could provide unique insights into the actual reading experience. Reading comprehension is too complex to be reduced to automatic evaluation measures or text complexity metrics, and future work should address the reader perspective to automatic text adaptation.

Reading is not only the processing of linguistic units, but it also includes design-related aspects. In the Easy Language guidelines, it is highlighted that the actual presentation of a text, i.e. what a text looks like to a reader, can be significant for the motivation and attitude towards the given text. Although we began looking into how we can visualise a text's complexity, we have not included other design-related challenges in this thesis, and there is still much to be done to enhance such aspects of our tools.

Similarly, the work on how text complexity should be visualised is still in its infancy. It is possible, and even probable, that how Easy Language readers want and need text complexity differs from the wants and needs of an Easy Language writer, and this should be further investigated, for instance by conducting workshops with Easy Language readers, similar to the web editor workshops described in Chapter 8.

CHAPTER 16

Conclusion

This doctoral thesis aimed to figure out how automatic text adaptation can and should be implemented to meet the needs of different target audiences, as a step towards a more inclusive world. The work conducted within the scope of the thesis concerns, among other things, if and how Easy Language text adaptations are useful for different groups of poor readers, how resources for automatic text adaptation can be assembled and how the complexity of a text can be visualised. The overall work also resulted in a palette of tools and services for analysis and adaptation of texts to target readers.

Our contributions to the overall goal of automatic adaptation for increased inclusion can be summarised in the following four items:

- Easy language guidelines were mapped to a theoretical model of reading comprehension, with a goal of obtaining information about if and how Easy Language text adaptations are useful for different groups of poor readers.
- A palette of tools and services for analysis and adaptation of texts were provided to target readers.
- The work on creating parallel corpora that can be used for automatic text adaptation purposes was presented. Among other things, this work includes implementation and evaluation of several methods for sentence alignment.
- Additional work was presented on how to make sense of text complexity features, by clustering and visualising them.

Some Final Words

Returning to the quotation which Einstein probably never uttered:

> *"Everything should be made as simple as possible, but no simpler."*

We could take this as an example of what we risk when we automatically adapt text for Easy Language purposes: we might produce a text that is comprehensible, but at the same time seems to have lost some of its meaning in the adaptation process. Or, we could claim that this captures how we *should* adapt text to Easy Language, creating text that is as simple as possible, with care for the strengths and challenges of the respective reader audience, while not becoming so simple that we lose information, alter the meaning of a text, or prevent the reader from getting the chance to improve their reading skills.

Today, large parts of our daily lives depend on textual data, and the ability to read is, thus, essential for participation in modern society. But the good thing about the times we live in, is that they also come with rapid technical and computational development that, if used correctly, could help overcome some of the challenges which poor readers may experience.

Bibliography

Abrahamsson, Emil, Timothy Forni, Maria Skeppstedt, and Maria Kvist (2014). "Medical text simplification using synonym replacement: Adapting assessment of word difficulty to a compounding language." In: *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR) @ EACL 2014*. Gothenburg, Sweden, pp. 57–65.

Abrahamsson, Peder (2011). *Mer Lättläst - Påbyggnad av ett automatiskt omskrivningsverktyg till lätt svenska*. Bachelor's thesis, Linköping University.

Albertsson, Sarah, Evelina Rennes, and Arne Jönsson (2016). "Similarity-Based Alignment of Monolingual Corpora for Text Simplification." In: *Coling 2016 Workshop on Computational Linguistics for Linguistic Complexity (CL4LC), Osaka, Japan*.

Alfter, David (2021). "Exploring natural language processing for single-word and multi-word lexical complexity from a second language learner perspective." PhD thesis. Department of Swedish, University of Gothenburg, Gothenburg, Sweden.

Alonzo, Oliver, Matthew Seita, Abraham Glasser, and Matt Huenerfauth (2020). "Automatic Text Simplification Tools for Deaf and Hard of Hearing Adults: Benefits of Lexical Simplification and Providing Users with Autonomy." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI, USA: Association for Computing Machinery, pp. 1–13.

Alsobhi, Aisha Yaquob, Nawaz Khan, and Harjinder Rahanu (2014). "Toward Linking Dyslexia Types and Symptoms to the Available Assistive Technologies." In: *2014 IEEE 14th International Conference on Advanced Learning Technologies*. IEEE, pp. 597–598.

Aluísio, Sandra Maria and Caroline Gasperin (2010). "Fostering digital inclusion and accessibility: the PorSimples project for simplification of Portuguese texts." In: *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*. Association for Computational Linguistics, pp. 46–53.

Alva-Manchego, Fernando, Joachim Bingel, Gustavo Henrique Paetzold, Carolina Scarton, and Lucia Specia (2017). "Learning How to Simplify from

Explicit Labeling of Complex-Simplified Text Pairs.” In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Taipei, Taiwan: Asian Federation of Natural Language Processing, pp. 295–305.

Alva-Manchego, Fernando, Carolina Scarton, and Lucia Specia (Jan. 2020). “Data-Driven Sentence Simplification: Survey and Benchmark.” In: *Computational Linguistics*, pp. 1–87.

Anderson, Richard C. and Peter Freebody (1979). “Vocabulary Knowledge.” In: *Technical Report No. 136*.

Arfé, Barbara, Lucia Mason, and Inmaculada Fajardo (2018). “Simplifying informational text structure for struggling readers.” In: *Reading and Writing* 31.9, pp. 2191–2210.

Barbu, Eduard, Maria Teresa Martín-Valdivia, and Luis Alfonso Ureña-López (2013). “Open Book: a tool for helping ASD users’ semantic comprehension.” In: *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility*, pp. 11–19.

Barlacchi, Gianni and Sara Tonelli (2013). “Ernesta: A sentence simplification tool for children’s stories in italian.” In: *International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 476–487.

Barzilay, Regina and Noemie Elhadad (2003). “Sentence Alignment for Monolingual Comparable Corpora.” In: *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pp. 25–32. doi: 10.3115/1119355.1119359. url: http://dx.doi.org/10.3115/1119355.1119359.

Beck, Isabel L, Margaret G McKeown, Gale M Sinatra, and Jane A Loxterman (1991). “Revising social studies text from a text-processing perspective : Evidence of improved comprehensibility.” In: *Reading Research Quarterly* 26, pp. 251–276.

Begriplig Text (2019). *19 råd för att skriva begripligt*. Dyslexiförbundet.

Bergström Kousta, Lia (2016). “Lättläst lättare - Utformningen av ett grafiskt gränssnitt för en automatisk textförenklare profilerat mot webbredaktörer.” Bachelor’s thesis. Linköping University.

Biber, Douglas (1988). *Variation across speech and writing*. Cambridge: Cambridge Univ. Press, 1988. isbn: 0521320712.

Billami, Mokhtar, Thomas François, and Núria Gala (2018). “ReSyf: a French lexicon with ranked synonyms.” In: *27th International Conference on Computational Linguistics (COLING 2018)*.

Björnsson, Carl Hugo (1968). *Läsbarhet*. Stockholm: Liber.

Borin, Lars, Marcus Forsberg, and Lennart Lönngren (2008). “SALDO 1.0 (Svenskt associationslexikon version 2).” In: *Språkbanken, Göteborgs universitet*.

Borin, Lars and Markus Forsberg (2014). “Swesaurus; or, The Frankenstein approach to Wordnet construction.” In: *Proceedings of the Seventh Global Wordnet Conference*, pp. 215–223.

Borin, Lars, Markus Forsberg, and Lennart Lönngren (2013). "SALDO: a touch of yin to WordNet's yang." In: *Language resources and evaluation* 47.4, pp. 1191–1211.

Bott, Stefan, Luz Rello, Biljana Drndarević, and Horacio Saggion (2012). "Can Spanish be Simpler? LexSiS: Lexical Simplification for Spanish." In: *Proceedings of COLING 2012*, pp. 357–374.

Bott, Stefan and Horacio Saggion (2011a). "An Unsupervised Alignment Algorithm for Text Simplification Corpus Construction." In: *Proceedings of the Workshop on Monolingual Text-To-Text Generation*. MTTG '11. Portland, Oregon: Association for Computational Linguistics, pp. 20–26. ISBN: 9781937284053. URL: http://dl.acm.org/citation.cfm?id=2107679.2107682.

— (2011b). "Spanish text simplification: An exploratory study." In: *Revista de Procesamiento de Lenguaje Natural* 47, pp. 87–95.

Bott, Stefan, Horacio Saggion, and Simon Mille (2012). "Text Simplification Tools for Spanish." In: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*.

Branco, António, João Rodrigues, Francisco Costa, Joao Silva, and Rui Vaz (2014). "Rolling out text categorization for language learning assessment supported by language technology." In: *International Conference on Computational Processing of the Portuguese Language*. Springer, pp. 256–261.

Brouwers, Laetitia, Delphine Bernhard, Anne-Laure Ligozat, and Thomas François (2014). "Syntactic sentence simplification for French." In: *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)@ EACL 2014*, pp. 47–56.

Brown, Heather M., Janis Oram-Cardy, and Andrew Johnson (2013). "A meta-analysis of the reading comprehension skills of individuals on the autism spectrum." In: *Journal of Autism and Developmental Disorders* 43.4, pp. 932–955.

Brunato, Dominique, Felice Dell'Orletta, Giulia Venturi, and Simonetta Montemagni (2015). "Design and annotation of the first Italian corpus for text simplification." In: *Proceedings of The 9th Linguistic Annotation Workshop*, pp. 31–41.

Burgers, Christian, Camiel J Beukeboom, Lisa Sparks, and Vera Diepeveen (2015). "How (not) to inform patients about drug use: use and effects of negations in Dutch patient information leaflets." In: *Pharmacoepidemiology and drug safety* 24.2, pp. 137–143.

Cain, Kate, Jane Oakhill, and Peter Bryant (2000). "Investigating the causes of reading comprehension failure: The comprehension-age match design." In: *Reading and Writing*, pp. 31–40. URL: http://link.springer.com/article/10.1023/A:1008058319399.

Canning, Yvonne and John Tait (1999). "Syntactic Simplification of Newspaper Text for Aphasic Readers." In: *ACM SIGIR'99 Workshop on Customised Information Delivery*.

Carenini, Giuseppe and Jackie C. K. Cheung (June 2008). "Extractive vs. NLG-based Abstractive Summarization of Evaluative Text: The Effect of Corpus Controversiality." In: *Proceedings of the Fifth International Natural Language Generation Conference*. Salt Fork, Ohio, USA: Association for Computational Linguistics, pp. 33–41. URL: https://aclanthology.org/W08-1106.

Carlson, Sarah E, Paul van den Broek, Kristen McMaster, David N Rapp, Catherine M Bohn-Gettler, Panayiota Kendeou, and Mary Jane White (2014). "Effects of comprehension skill on inference generation during reading." In: *International Journal of Disability Development and Education* 61.3, pp. 258–274. DOI: 10.1080/1034912X.2014.934004.

Carrell, Patricia L. (1984). "The effects of rhetorical organization on ESL readers." In: *TESOL quarterly* 18.3, pp. 441–469.

Carroll, John, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait (1998). "Practical simplification of English newspaper text to assist aphasic readers." In: *Proceedings of the AAAI98 Workshop on Integrating Artificial Intelligence and Assistive Technology*. Vol. 1. Citeseer, pp. 7–10.

Carroll, John, Guido Minnen, Darren Pearce, Yvonne Canning, Siobhan Devlin, and John Tait (1999). "Simplifying Text for Language-Impaired Readers." In: *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 269–270.

Chall, Jeanne S. and Edgar Dale (1995). *Readability revisited: The new Dale–Chall readability formula*. Cambride, MA: Brookline Books.

Chandrasekar, Raman, Christine Doran, and Bangalore Srinivas (1996). "Motivations and Methods for Text Simplification." In: *Proceedings of the Sixteenth International Conference on Computational Linguistics (COLING '96)*.

Chandrasekar, Raman and Bangalore Srinivas (1997). "Automatic induction of rules for text simplification." In: *Knowledge-Based Systems* 10.3, pp. 183–190. ISSN: 0950-7051. DOI: 10.1016/S0950-7051(97)00029-4. URL: http://www.sciencedirect.com/science/article/pii/S0950705197000294.

Chatterjee, Nilhadri and Shiwali Mohan (2007). "Extraction-Based Single-Document Summarization Using Random Indexing." In: *Proceedings of the 19th IEEE international Conference on Tools with Artificial intelligence – (ICTAI 2007)*, pp. 448–455.

Chung, Jin-Woo, Hye-Jin Min, Joonyeob Kim, and Jong C Park (2013). "Enhancing Readability of Web Documents by Text Augmentation for Deaf People." In: *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics*, pp. 1–10.

Coster, William and David Kauchak (2011a). "Learning to simplify sentences using Wikipedia." In: *Proceedings of the workshop on monolingual text-to-text generation*. Association for Computational Linguistics, pp. 1–9.

— (2011b). “Simple English Wikipedia: A New Text Simplification Task.” In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 665–669. URL: http://www.aclweb.org/anthology/P11-2117.

— (2011c). “Simple English Wikipedia: a new text simplification task.” In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, pp. 665–669.

Daelemans, Walter, Anja Höthker, and Erik F Tjong Kim Sang (2004). “Automatic Sentence Simplification for Subtitling in Dutch and English.” In: *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*.

Dale, Edgar and Jeanne S. Chall (1949). “The Concept of Readability.” In: *Elementary English* 26.23.

Danielsson, Henrik, Lucy Henry, David Messer, and Jerker Rönnberg (2012). “Strengths and weaknesses in executive functioning in children with intellectual disability.” In: *Research in developmental disabilities* 33.2, pp. 600–607.

De Belder, Jan and Marie-Francine Moens (2010). “Text Simplification for Children.” In: *Proceedings of the SIGIR workshop on accessible search systems*. ACM; New York, pp. 19–26.

Decker, Anna (2003). “Towards automatic grammatical simplification of Swedish text.” Master’s thesis. Stockholm University.

Dell’Orletta, Felice, Simonetta Montemagni, and Giulia Venturi (July 2011). “READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification.” In: *Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies*. Edinburgh, Scotland, UK, pp. 73–83.

Den internationella undersökningen av vuxnas färdigheter (2013). Tech. rep. Rapport 2013:2. Accessed: 2021-11-10. SCB, Tema utbildning.

Denoyer, Ludovic and Patrick Gallinari (2006). “The Wikipedia XML Corpus.” In: *Comparative Evaluation of XML Information Retrieval Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 12–19.

Devlin, Siobhan and Gary Unthank (2006). “Helping aphasic people process online information.” In: *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, pp. 225–226.

Dolan, Bill, Chris Quirk, and Chris Brockett (2004). “Unsupervised Construction of Large Paraphrase Corpora: Exploiting Massively Parallel News Sources.” In: *Proceedings of the 20th International Conference on Computational Linguistics (COLING ’04)*, Article No. 350. DOI: 10.3115/1220355.1220406. URL: http://dl.acm.org/citation.cfm?id=1220406.

Drndarević, Biljana and Horacio Saggion (2012). “Towards Automatic Lexical Simplification in Spanish: an Empirical Study.” In: *Proceedings of the First*

Workshop on Predicting and Improving Text Readability for target reader populations, pp. 8–16.

DuBay, William H. (2004). *The Principles of Readability.* Impact Information, Costa Mesa, California.

Duchowski, Andrew T. (2007). *Eye tracking methodology: Theory and practice.* Springer-Verlag, London.

Dvorkovich, Anton, Sergey Gubanov, and Irina Galinskaya (2016). "Yandex School of Data Analysis approach to English-Turkish translation at WMT16 News Translation Task." In: *Proceedings of the First Conference on Machine Translation.* Vol. 2. Berlin, Germany, pp. 281–288.

Dyslexiföreningen (n.d.). *Vad är läs- och skrivsvårigheter/dyslexi?* https://www.dyslexiforeningen.se/vad-ar-dyslexi/. Accessed: 2018-08-06.

Egbert, Jesse and Douglas Biber (2018). "Do all roads lead to Rome?: Modeling register variation with factor analysis and discriminant analysis." In: *Corpus Linguistics and Linguistic Theory* 14.2, pp. 233–273.

Ehmke, Claudia and Stephanie Wilson (2007). "Identifying Web Usability Problems from Eye-Tracking Data." In: *Proceedings of HCI 2007.* Vol. 1, pp. 119–128.

Eika, Evelyn and Frode Eika Sandnes (2016). "Authoring WCAG2. 0-compliant texts for the web through text readability visualization." In: *International Conference on Universal Access in Human-Computer Interaction.* Springer, pp. 49–58.

Einstein, Albert (1934). "On the method of theoretical physics." In: *Philosophy of science* 1.2, pp. 163–169.

Ejerhed, Eva, Gunnel Källgren, and Benny Brodda (2006). *Stockholm Umeå Corpus version 2.0.*

Elwér, Åsa, Janice M. Keenan, Richard K. Olson, Brian Byrne, and Stefan Samuelsson (July 2013). In: *Journal of experimental child psychology* 3, pp. 497–516. DOI: 10.1016/j.jecp.2012.12.001.

Evans, Richard, Constantin Orasan, and Iustin Dornescu (2014). "An evaluation of syntactic simplification rules for people with autism." In: *Proceedings of the Third Workshop on Predicting and Improving Text Readability for Target Reader Populations.*

Evans, Vyvyan (2019). *Cognitive Linguistics(2nd edition).* Edinburgh: Edinburgh University Press.

Fabbretti, Daniela, Virginia Volterra, and Clotilde Pontecorvo (1998). "Written Language Abilities in Deaf Italians." In: *The Journal of Deaf Studies and Deaf Education* 3.3, pp. 231–244.

Fahlborg, Daniel and Evelina Rennes (2016). "Introducing SAPIS - an API service for text analysis and simplification." In: *The second national Swe-Clarin workshop: Research collaborations for the digital age, Umeå, Sweden.*

Fajardo, Inmaculada, Vicenta Avila, Antonio Ferrer, Gema Tavares, Marcos Gomez, and Ana Hernandez (2014). "Easy-to-read Texts for Students with

Intellectual Disability: Linguistic Factors Affecting Comprehension.” In: *Journal of Applied Research in Intellectual Disabilities* 27.
Falk, Lena and Stefan Johansson (2006). *Hur fungerar lättlästa texter på webben? Undersökning av lättlästa texter på offentliga webbplatser*. URL: http : / / www . funka . com / contentassets / 755a917870714fc8b9dd5a34de6a3237/rapport%5C_lattlast.pdf.
Falkenjack, Johan (Nov. 27, 2018a). Personal communication.
— (2018b). “Towards a model of general text complexity for Swedish.” Licentiate thesis. Linköping University Electronic Press.
Falkenjack, Johan, Katarina Heimann Mühlenbock, and Arne Jönsson (2013). “Features Indicating Readability in Swedish Text.” In: *Proceedings of the 19th Nordic Conference of Computational Linguistics (NoDaLiDa-2013), Oslo, Norway*. NEALT Proceedings Series 16 085. Linköping University Electronic Press, pp. 27–40.
Falkenjack, Johan, Evelina Rennes, Daniel Fahlborg, Vida Johansson, and Arne Jönsson (2017). “Services for text simplification and analysis.” In: *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pp. 309–313.
Falkenjack, Johan, Marina Santini, and Arne Jönsson (2016). “An Exploratory Study on Genre Classification using Readability Features.” In: *Proceedings of the The Sixth Swedish Language Technology Conference (SLTC 2016), Umeå, Sweden*.
Fallgren, Per, Jesper Segeblad, and Marco Kuhlmann (2016). “Towards a Standard Dataset of Swedish Word Vectors.” In: *Proceedings of the Sixth Swedish Language Technology Conference (SLTC), Umeå, Sweden*.
Feng, Lijun, Noémie Elhadad, and Matt Huenerfauth (2009). “Cognitively Motivated Features for Readability Assessment.” In: *Proceedings of the 12th Conference of the European Chapter of the ACL*. Athens, Greece.
Flesch, Rudolf (1948). “A new readability yardstick.” In: *Journal of Applied Psychology* 32.3, pp. 221–233.
Forsbom, Eva (2006). “A swedish base vocabulary pool.” In: *Swedish Language Technology conference, Gothenburg*.
Fostick, Leah and Hadas Revah (2018). “Dyslexia as a multi-deficit disorder: Working memory and auditory temporal processing.” In: *Acta psychologica* 183, pp. 19–28.
François, Thomas, Elena Volodina, Ildikó Pilán, and Anaïs Tack (May 2016). “SVALex: a CEFR-graded Lexical Resource for Swedish Foreign and Second Language Learners.” In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association (ELRA), pp. 213–219. URL: https://aclanthology.org/L16-1032.
Gala, Núria, Anaïs Tack, Ludivine Javourey-Drevet, Thomas François, and Johannes C Ziegler (2020). “Alector: A Parallel Corpus of Simplified French

Texts with Alignments of Misreadings by Poor and Dyslexic Readers." In: *Language Resources and Evaluation for Language Technologies (LREC).*

Gala, Núria and Johannes Ziegler (2016). "Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia." In: *Coling 2016 Workshop on Computational Linguistics for Linguistic Complexity (CL4LC), Osaka, Japan.*

Gardner, Dee and Elizabeth C. Hansen (2007). "Effects of Lexical Simplification During Unaided Reading of English Informational Texts." In: *TESL Reporter* 40.2, pp. 27–59.

Gernsbacher, Morton A. (1984). "Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy." In: *Journal of Experimental Psychology: General* 113.2, pp. 256–281.

Glavaš, Goran and Sanja Štajner (2015). "Simplifying lexical simplification: Do we need simplified corpora?" In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers).* Vol. 2, pp. 63–68.

Gonzalez-Dios, Itziar, María Jesús Aranzabe, Arantza Díaz de Ilarraza, and Haritz Salaberri (2014). "Simple or Complex? Assessing the Readability of Basque Texts." In: *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers*, pp. 334–344.

Goto, Isao, Hideki Tanaka, and Tadashi Kumano (2015). "Japanese news simplification: Task design, data set construction, and analysis of simplified text." In: *Proceedings of MT Summit XV* 1, pp. 17–31.

Graesser, Arthur C., Danielle S. McNamara, Max M. Louwerse, and Zhiqiang Cai (2004a). "Coh-Metrix: Analysis of text on cohesion and language." In: *Behavior research methods, instruments, & computers* 36.2, pp. 193–202.

Graesser, Arthus C., Danielle S. McNamara, Max M. Louwerse, and Zhiqiang Cai (2004b). "Coh-Metrix: Analysis of text on cohesion and language." In: *Behavior Research Methods, Instruments, & Computers* 36.2, pp. 193–202.

Guthrie, John T, Allan Wigfield, Nicole M Humenick, Kathleen C Perencevich, Ana Taboada, and Pedro Barbosa (2006). "Influences of Stimulating Tasks on Reading Motivation and Comprehension." In: *The Journal of Educational Research* 99.4, pp. 232–246.

Hahn, Udo and Inderjeet Mani (2000). "The Challenges of Automatic Summarization." In: *Computer* 33.11, pp. 29–36.

Hansson, Karin (2020). *Det finns ett sug efter klarspråk: Bättre stöd till klarspråksansvariga i offentlig verksamhet.*

Hansson, Sven Ove (2007). "What is technological science?" In: *Studies in History and Philosophy of Science Part A* 38.3, pp. 523–527.

Harris, Zellig S. (1970). "Distributional structure." In: *Papers in structural and transformational Linguistics.* Dordrecht: Springer Science+Business Media B.V., pp. 775–794.

Hassel, Martin (2000). "Pronominal Resolution in Automatic Text Summarisation." MA thesis. Master thesis in Computer Science, Department of Computer and Systems Sciences (DSV), Stockholm University, Sweden.

— (2007). "Resource Lean and Portable Automatic Text Summarization." PhD thesis. ISRN-KTH/CSC/A–07/09-SE, KTH, Sweden.

— (Jan. 2011). *Java Random Indexing toolkit*. http://www.csc.kth.se/~xmartin/java/. URL: http://www.csc.kth.se/~xmartin/java.

Hazem, Amir and Emmanuel Morin (May 2016). "Improving Bilingual Terminology Extraction from Comparable Corpora via Multiple Word-Space Models." In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Ed. by Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Portorož, Slovenia: European Language Resources Association (ELRA). ISBN: 978-2-9517408-9-1.

Heimann Mühlenbock, Katarina (2013). "I see what you mean. Assessing readability for specific target groups." Dissertation. Språkbanken, Dept of Swedish, University of Gothenburg. URL: http://hdl.handle.net/2077/32472.

Heimann Mühlenbock, Katarina and Sofie Johansson Kokkinakis (2012a). "SweVoc - a Swedish vocabulary resource for CALL." In: *Proceedings of the SLTC 2012 workshop on NLP for CALL*. Lund: Linköping University Electronic Press, pp. 28–34.

— (Oct. 2012b). "SweVoc – A Swedish vocabulary resource for CALL." In: *Proceedings of the SLTC 2012 workshop on NLP for CALL*. Lund: Linköping University Electronic Press, pp. 28–34. URL: http://www.ep.liu.se/ecp/080/004/ecp12080004.pdf.

Hermann, Karl Moritz, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom (2015). "Teaching Machines to Read and Comprehend." In: *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*. Montreal, Canada: MIT Press.

Hess, Eckhard H. and James M. Polt (1964). "Pupil Size in Relation to Mental Activity during Simple Problem-Solving." In: *Science* 132.3611, pp. 1190–1192.

Hillis, Argye E. (2007). "Aphasia: Progress in the last quarter of a century." In: *Neurology* 69, pp. 200–213.

Hmida, Firas, Mokhtar B. Billami, Thomas François, and Núria Gala (2018). "Assisted Lexical Simplification for French Native Children with Reading Difficulties." In: *Proceedings of the 1st Workshop on Automatic Text Adaptation (ATA)*. Tilburg, the Netherlands: Association for Computational Linguistics, pp. 21–28.

Hobson, R Peter and Anthony Lee (1989). "Emotion-related and abstract concepts in autistic people: Evidence from the British Picture Vocabulary

Scale." In: *Journal of autism and developmental disorders* 19.4, pp. 601–623. doi: 10.1007/BF02212860.

Høien, Torleiv and Ingvar Lundberg (2013). *Dyslexi - från teori till praktik.* 2., [uppda. Stockholm: Natur & Kultur.

Holmer, Daniel and Arne Jönsson (2020). "Comparing the performance of various Swedish BERT models for classification." In: *Eighth Swedish Language Technology Conference (SLTC2020). Organised by University of Gothenburg, Sweden.*

Holmer, Daniel and Evelina Rennes (2022). "NyLLex: A Novel Resource of Swedish Words Annotated with Reading Proficiency Level." Manuscript submitted for publication.

Holmqvist, Kenneth (2011). *Eye Tracking - A Comprehensive Guide to Methods and Measures.* Oxford University Press.

Hörselskadade i siffror 2017. Statistik om hörselskadade och hörapparatutprovningar i Sverige från Hörselskadades Riksförbund (HRF) (2017). Tech. rep. Accessed: 2021-11-10. Hörselskadades Riksförbund (HRF).

Hua, AN and JM Keenan (2014). "The role of text memory in inferencing and in comprehension deficits." In: *Scientific Studies of Reading* 18.6, pp. 415–431. issn: 1088-8438. doi: 10.1080/10888438.2014.926906. url: http://www.tandfonline.com/doi/abs/10.1080/10888438.2014.926906.

Huenerfauth, Matt, Lijun Feng, and Noémie Elhadad (2009). *Comparing Evaluation Techniques for Text Readability Software for Adults with Intellectual Disabilities.*

Hulme, Charles and Margaret J. Snowling (2011). "Children's Reading Comprehension Difficulties." In: *Current Directions in Psychological Science* 20.3, pp. 139–142. issn: 0963-7214. doi: 10.1177/0963721411408673. url: http://journals.sagepub.com/doi/10.1177/0963721411408673.

Hultman, Tor G. and Margareta Westman (1977). *Gymnasistsvenska.* Lund: LiberLäromedel.

Hur många döva finns det i Sverige? (N.d.). https://sdr.org/question/hur-manga-dova-finns-det-i-sverige/. Accessed: 2021-11-10.

Hwang, William, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu (2015). "Aligning Sentences from Standard Wikipedia to Simple Wikipedia." In: *HLT-NAACL*, pp. 211–217.

Ihs Håkansson, Ellinor (2018). *TexSTo - Text Simplification tool Ett läsverktyg med fokus på personer med lässvårigheter.* Internal report.

Inclusion Europe (2014). *Information for all - European standards for making information easy to read and understand.*

— (2020). *Easy-to-read checklist. Checklist to make sure your document is easy to read.* https://www.inclusion-europe.eu/wp-content/uploads/2020/06/Easy-to-read-checklist-Inclusion-Europe.pdf. Accessed: 2021-09-30.

Inrikes och utrikes födda efter region, ålder och kön. År 2000–2020 (2020). http://www.statistikdatabasen.scb.se/sq/117031. Accessed: 2021-11-10.

International Plain Language Federation (n.d.). *International Plain Language Federation*. https://www.iplfederation.org/plain-language/. Accessed: 2021-09-28.

Inui, Kentaro, Atsushi Fujita, Tetsuro Takahashi, Ryu Iida, and Tomoya Iwakura (2003). "Text Simplification for Reading Assistance: a project note." In: *Proceedings of the second international workshop on Paraphrasing*. Association for Computational Linguistics, pp. 9–16.

Jin, Hongyan (2002). "Using Hidden Markov Modeling to Decompose Human-Written Summaries." In: *Computational Linguistics* 28.4, pp. 527–543.

Johansson, Vida (2017). "Depending on VR: Rule-based Text Simplification Based on Dependency Relations." B.Sc. thesis. Linköping University.

Johansson, Vida and Evelina Rennes (2016). "Automatic Extraction of Synonyms from an Easy-to-Read Corpus." In: *Proceedings of the Sixth Swedish Language Technology Conference (SLTC-16), Umeå, Sweden*.

Jönsson, Simon, Evelina Rennes, Johan Falkenjack, and Arne Jönsson (2018). "A component based approach to measuring text complexity." In: *Proceedings of The Seventh Swedish Language Technology Conference 2018 (SLTC-18)*.

Just, M. A. and P. A. Carpenter (1980). "A theory of reading: From eye fixations to comprehension." In: *Psychological Review* 87.4, pp. 329–354.

Kajiwara, Tomoyuki and Mamoru Komachi (2016). "Building a Monolingual Parallel Corpus for Text Simplification Using Sentence Similarity Based on Alignment between Word Embeddings." In: *Proceedings of COLING, Osaka, Japan*, pp. 1147–1158.

Kalandadze, Tamar, Courtenay Norbury, Terje Nærland, and Kari-Anne B Næss (2018). "Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review." In: *Autism* 22.2, pp. 99–117.

Kann, Viggo (2004). "Folkets användning av Lexin – en resurs." In: *Lexinkonferens 2004*. Stockholm, Sweden.

Kann, Viggo and Magnus Rosell (2005). "Free construction of a free Swedish dictionary of synonyms." In: *Proceedings of the 15th NODALIDA conference*. Stockholm, pp. 105–110.

Karmakar, Saurav and Ying Zhu (2010a). "Visualizing multiple text readability indexes." In: *Education and Management Technology (ICEMT), 2010 International Conference on*. IEEE, pp. 133–137.

— (2010b). "Visualizing text readability." In: *Advanced Information Management and Service (IMS), 2010 6th International Conference on*. IEEE, pp. 291–296.

Karreman, Joyce, Thea Van Der Geest, and Esmee Buursink (2007). "Accessible Website Content Guidelines for Users with Intellectual Disabilities." In: *Journal of Applied Research in Intellectual Disabilities* 20.6, pp. 510–518.

Kaspersson, Thomas, Christian Smith, Henrik Danielsson, and Arne Jönsson (2012). "This also affects the context - Errors in extraction based summaries." In: *Proceedings of the eighth international conference on Language Resources and Evaluation (LREC), Istanbul, Turkey.*

Kendeou, Panayiota, Paul Van Den Broek, Anne Helder, and Josefine Karlsson (2014). "A cognitive view of reading comprehension: Implications for reading difficulties." In: *Learning Disabilities Research and Practice* 29.1, pp. 10–16. ISSN: 09388982. DOI: 10.1111/ldrp.12025.

Keskisärkkä, Robin and Arne Jönsson (2013). "Investigations of Synonym Replacement for Swedish." In: *Northern European Journal of Language Technology* 3.3, pp. 41–59.

Kim, Hyoyoung, Jin Wan Park, and Dongsu Seo (2014). "Readability visualization for massive text data." In: *Int. J. Multimedia Ubiquit. Eng* 9, pp. 707–719.

Kincaid, J. Peter, Robert P. Fishburne, Richard L. Rogers, and Brad S. Chissom (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel.* Tech. rep. Millington, TN: U.S. Naval Air Station.

Kintsch, Walter (1988). "The role of knowledge in discourse comprehension: a construction-integration model." In: *Psychological review* 95.2, pp. 163–82.

Kintsch, Walter and Praful Mangalath (2011). "The Construction of Meaning." In: *Topics in Cognitive Science* 3.2, pp. 346–370.

Kintsch, Walter and Teun A Van Dijk (1978). "Toward a Model of Text Comprehension and Production." In: *Psychological review* 85.5, p. 363.

Kjelgaard, Margaret M and Helen Tager-Flusberg (2001). "An investigation of language impairment in autism: Implications for genetic subgroups." In: *Language and cognitive processes* 16.2-3, pp. 287–308.

Kleijn, Suzanne, Henk L.W. Pander Maat, and Ted J.M. Sanders (2019). "Comprehension Effects of Connectives Across Texts, Readers, and Coherence Relations." In: *Discourse Processes* 56.5-6, pp. 447–464. ISSN: 15326950. DOI: 10.1080/0163853X.2019.1605257. URL: https://doi.org/10.1080/0163853X.2019.1605257.

Klerke, Sigrid and Anders Søgaard (2012). "DSim, a Danish Parallel Corpus for Text Simplification." In: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12),Istanbul, Turkey.* European Language Resources Association (ELRA). ISBN: 978-2-9517408-7-7.

Knight, Kevin and Daniel Marcu (2000). "Statistics-Based Summarization - Step One: Sentence Compression." In: *AAAI-00 - 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence*, pp. 703–710. ISSN: 0004-3702. DOI: http://dx.doi.org/10.1016/S0004-3702(02)00222-9. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.2394.

Krashen, Stephen (1989). “We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis.” In: *The modern language journal* 73.4, pp. 440–464.

Kucera, Henry and W. Nelson Francis (1967). *Computational analysis of present-day American English*. University Press of New England.

Lättlästförlaget (n.d.). *Att skriva lättläst – våra bästa tips*. https://ll-forlaget.se/om-oss/att-skriva-lattlast/. Accessed: 2021-09-29.

Lederberg, Amy R., Brenda Schick, and Patricia E. Spencer (2013). “Language and literacy development of deaf and hard-of-hearing children: successes and challenges.” In: *Developmental psychology* 49.1, p. 15.

Lin, Dekang, Shaojun Zhao, Lijuan Qin, and Ming Zhou (2003). “Identifying synonyms among distributionally similar words.” In: *IJCAI'03 Proceedings of the 18th international joint conference on Artificial intelligence*. San Fransisco: Morgan Kaufmann Publishers Inc., pp. 1492–1493.

Liu, Haitao (2008). “Dependency distance as a metric of language comprehension difficulty.” In: *Journal of Cognitive Science* 9.2, pp. 169–191.

Long, Michael H. and Steven Ross (1993). “Modifications That Preserve Language and Content.” In: *Simplification: Theory and application*. Ed. by M. L. Tickoo. Singapore: SEAMEO Regional Language Center, pp. 29–52.

Malmsten, Martin, Love Börjeson, and Chris Haffenden (2020). *Playing with Words at the National Library of Sweden – Making a Swedish BERT*. arXiv: 2007.01658 [cs.CL].

Mani, Inderjeet, Eric Bloedorn, and Barbara Gates (1998). “Using Cohesion and Coherence Models for Text Summarization.” In: *AAAI Technical Report SS-98-06*.

Margarido, Paulo R. A., Thiago A. S. Pardo, Gabriel M. Antonio, Vinícius B. Fuentes, Rachel Aires, Sandra M. Aluísio, and Renata P. M. Fortes (2008). “Automatic Summarization for Text Simplification: Evaluating Text Understanding by Poor Readers.” In: *Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web*.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). “Efficient Estimation of Word Representations in Vector Space.” In: *CoRR* abs/1301.3781.

Misako Nomura, Gyda Skat Nielsen and Bror Tronbacke (2010). *Guidelines for Easy-to-Read materials*. International Federation of Library Associations and Institutions.

Monsen, Julius and Arne Jönsson (2021). “A method for building non-English corpora for abstractive text summarization.” In: *Proceedings of CLARIN Annual Conference 2021*.

Monsen, Julius and Evelina Rennes (2022). “Perceived Text Quality and Readability in Extractive and Abstractive Summaries.” Manuscript submitted for publication.

Morgan, Michelle F. and Karen B. Moni (2008). “Meeting the challenge of limited literacy resources for adolescents and adults with intellectual disabilities.” In: *British Journal of Special Education* 35.2, pp. 91–101.

Morrison, Catriona M and Andrew W Ellis (1995). “Roles of word frequency and age of acquisition in word naming and lexical decision.” In: *Journal of experimental psychology: learning, Memory, and cognition* 21.1, p. 116.

MTM (2021). *Att skriva lättläst*. https://www.mtm.se/var-verksamhet/lattlast/att-skriva-lattlast/. Accessed: 2021-10-05.

Mühlenbock, Katarina (2008). “Readable, Legible or Plain Words – Presentation of an easy-to-read Swedish corpus.” In: *Multilingualism: Proceedings of the 23rd Scandinavian Conference of Linguistics*. Ed. by Anju Saxena and Åke Viberg. Vol. 8. Acta Universitatis Upsaliensis. Uppsala, Sweden: Acta Universitatis Upsaliensis, pp. 327–329.

Mukherjee, Partha, Gondy Leroy, David Kauchak, Srinidhi Rajanarayanan, Damian Y Romero Diaz, Nicole P Yuan, T Gail Pritchard, and Sonia Colina (2017). “NegAIT: A new parser for medical text simplification using morphological, sentential and double negation.” In: *Journal of biomedical informatics* 69, pp. 55–62.

Nakov, Preslav and Torsten Zesch, eds. (2014). *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*.

Nallapati, Ramesh, Bowen Zhou, Cicero dos Santos, Caglar Gulcehre, and Bing Xiang (Aug. 2016). “Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond.” In: *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. Berlin, Germany: Association for Computational Linguistics, pp. 280–290. doi: 10.18653/v1/K16-1028. url: https://aclanthology.org/K16-1028.

Nations, United (2020). *Convention on the Rights of Persons with Disabilities*. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with\-disabilities/convention-on-the-rights\-of-persons-with-disabilities-2.html. (Visited on 06/23/2020).

NHS (n.d.). *Overview: Dyslexia*. https://www.nhs.uk/conditions/dyslexia/. Accessed: 2018-08-06.

Nilsson, Karin, Henrik Danielsson, Åsa Elwér, David Messer, Lucy Henry, and Stefan Samuelsson (2021a). “Decoding Abilities in Adolescents with Intellectual Disabilities: The Contribution of Cognition, Language, and Home Literacy.” In: 4.1, pp. 1–16. doi: 10.5334/joc.191.

— (2021b). “Investigating reading comprehension in adolescents with intellectual disabilites: Evaluating the simple view of reading.” In: *Journal of Cognition* 4.1, pp. 1–16. doi: 10.5334/joc.191.

Nisioi, Sergiu, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P Dinu (2017). “Exploring neural text simplification models.” In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 85–91.

Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülşen Eryigit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi (2007). "MaltParser: A language-independent system for data-driven dependency parsing." In: *Natural Language Engineering* 13.2, pp. 95–135.

Oelke, Daniela, David Spretke, Andreas Stoffel, and Daniel A Keim (2012). "Visual readability analysis: How to make your writings easier to read." In: *IEEE Transactions on Visualization and Computer Graphics* 18.5, pp. 662–674.

Östling, Robert (2013). "Stagger: An open-source part of speech tagger for Swedish." In: *Northern European Journal of Language Technology (NEJLT)* 3, pp. 1–18.

— (2018). "Part of Speech Tagging: Shallow or Deep Learning?" In: *North European Journal of Language Technology* 5, pp. 1–15.

Otterbacher, Jahna C., Dragomir R. Radev, and Airong Luo (2002). "Revisions that Improve Cohesion in Multi-document Summaries: A Preliminary Study." In: *Proceedings of the Workshop on Automatic Summarization (including DUC 2002), Philadelphia*, pp. 27–36.

Paetzold, Gustavo Henrique (2016). "Lexical Simplification for Non-Native English Speakers." Ph.D. thesis. Sheffield, UK: University of Sheffield.

Paetzold, Gustavo Henrique and Lucia Specia (2017). "A Survey on Lexical Simplification." In: *Journal of Artificial Intelligence Research* 60, pp. 549–593.

Palincsar, A. S. and A. L. Brown (1984). "Reciprocal Teaching of comprehension-fostering and comprehension-monitoring activities." In: *Cognition and Instruction* 1, pp. 117–175.

Pavlick, Ellie and Chris Callison-Burch (2016). "Simple PPDB: A Paraphrase Database for Simplification." In: *ACL*.

Pedler, Jennifer (2007). "Computer Correction of Real-Word Spelling Errors in Dyslexic Text." Ph.D. thesis. University of London.

Pilán, Ildikó, David Alfter, and Elena Volodina (2016). "Coursebook texts as a helping hand for classifying linguistic complexity in language learners' writings." In: *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, pp. 120–126.

Pilán, Ildikó, Sowmya Vajjala, and Elena Volodina (2016). "A Readable Read: Automatic Assessment of Language Learning Materials based on Linguistic Complexity." In: *International Journal of Computational Linguistics and Applications* 7, pp. 143–159.

Pinter, Yuval, Robert Guthrie, and Jacob Eisenstein (2017). "Mimicking Word Embeddings using Subword RNNs." In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 102–112.

PLAIN (2011). *Federal Plain Language Guidelines, revision 1*. https://www.plainlanguage.gov/media/FederalPLGuidelines.pdf. Accessed: 2021-10-18.

Priscepov, Roman (2020). "Automatic Text Simplification of Swedish forSecond Language Learners." Master's thesis. Technische Universität München.

Ratz, Christoph and Wolfgang Lenhard (2013). "Reading skills among students with intellectual disabilities." In: *Research in Developmental Disabilities* 34.5, pp. 1740–1748.

Rayner, Keith (1998). "Eye movements in reading and information processing: 20 years of research." In: *Psychological Bulletin* 124.3, pp. 372–422.

Reichle, Erik D., Alexander Pollatsek, Donalt L. Fisher, and Keith Rayner (1998). "Toward a Model of Eye Moement Control in Reading." In: *Psychological Review* 105.1, pp. 125–157.

Rello, Luz, Ricardo Baeza-Yates, Stefan Bott, and Horacio Saggion (2013). "Simplify or help? Text simplification strategies for people with dyslexia." In: *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, pp. 1–10.

Rello, Luz, Ricardo Baeza-Yates, Laura Dempere, and Horacio Saggion (2013). "Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia." In: *Human-Computer Interaction, INTERACT 2013 - 14th IFIP TC 13 International Conference.* Cape Town, pp. 203–219.

Rello, Luz, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion (2013). "Frequent words improve readability and short words improve understandability for people with dyslexia." In: *IFIP Conference on Human-Computer Interaction.* Springer, pp. 203–219.

Rello, Luz, Ricardo Baeza-Yates, and Joaquim Llisterri (2014). "DysList: An Annotated Resource of Dyslexic Errors." In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).* Reykjavik, Iceland, pp. 1289–1296.

Rello, Luz, Susana Bautista, Ricardo Baeza-Yates, Pablo Gervás, Raquel Hervás, and Horacio Saggion (2013). "One half or 50%? An Eye-Tracking Study of Number Representation Readability." In: *IFIP Conference on Human-Computer Interaction.* Springer, pp. 229–245.

Rennes, Evelina (2015). "Improved Automatic Text Simplification by Manual Training." Master's thesis. Linköping University.

— (2018). "An Aligned Resource of Swedish Complex-Simple Sentence Pairs." In: *Proceedings of the Seventh Swedish Language Technology Conference (SLTC2018), Stockholm, Sweden.*

— (2020). "Is it simpler? An Evaluation of an Aligned Corpus of Standard-Simple Sentences." In: *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI@LREC), Marseille, France.* Pp. 6–13.

Rennes, Evelina and Åsa Elwér (2022). "Evidence-based Text Adaptations for Poor Readers." Manuscript submitted for publication.

Rennes, Evelina and Arne Jönsson (2014). "The impact of cohesion errors in extraction based summaries." In: *9th International Conference on Language Resources and Evaluation (LREC).* European Language Resources Association, pp. 1575–1582.

— (2015). "A Tool for Automatic Simplification of Swedish Texts." In: *Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa-2015), Vilnius, Lithuania.*

— (2016). "Towards a Corpus of Easy to Read Authority Web Texts." In: *Proceedings of the Sixth Swedish Language Technology Conference (SLTC-16), Umeå, Sweden.*

— (2021). "Synonym Replacement based on a Study of Basic-level Nouns in Swedish Texts of Different Complexity." In: *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pp. 259–267.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem (1976). "Basic objects in natural categories." In: *Cognitive Psychology.*

Rothe, Sascha, Shashi Narayan, and Aliaksei Severyn (2020). *Leveraging Pre-trained Checkpoints for Sequence Generation Tasks.* arXiv: 1907.12461 [cs.CL].

Rybing, Jonas, Christian Smith, and Annika Silvervarg (2010). "Towards a Rule Based System for Automatic Simplification of Texts." In: *Swedish Language Technology Conference, SLTC, Linköping, Sweden.*

Saggion, Horacio (2017). *Automatic Text Simplification.* Synthesis Lectures on Human Language Technologies Vol. 32. Morgan & Claypool Publishers. ISBN: 9781627058681.

Saggion, Horacio, Montserrat Marimon, and Daniel Ferrés (2015). "Simplificación automática de textos para la accesibilidad de colectivos con discapacidad: experiencias para el español y el inglés." In: *IX Jornadas Científicas Internacionales de Inverstigación sobre Personas con Discapacidad.* Salamanca, Spain.

Saggion, Horacio, Sanja Štajner, Stefan Bott, Simon Mille, Luz Rello, and Biljana Drndarevic (2015). "Making It Simplext: Implementation and Evaluation of a text Simplification System for Spanish." In: *ACM Transactions on Accessible Computing* 6.4, pp. 1–36. ISSN: 19367228. DOI: 10.1145/2738046. URL: http://dl.acm.org/citation.cfm?id=2775084.2738046.

Sanchez, Rebecca Polley, Elizabeth Pugzles Lorch, and Robert F Lorch Jr (2001). "Effects of headings on text processing strategies." In: *Contemporary educational psychology* 26.3, pp. 418–428. DOI: 10.1006/ceps.2000.1056.

Sanchez-Perez, Miguel A., Grigori Sidorov, and Alexander Gelbukh (2014). "The winning approach to text alignment for text reuse detection at PAN 2014: Notebook for PAN at CLEF 2014." In: *CEUR Workshop Proceedings* 1180, pp. 1004–1011. ISSN: 16130073.

Santini, Marina, Benjamin Danielsson, and Arne Jönsson (2019). "Comparing the Performance of Feature Representations for the Categorization of the Easy-to-Read Variety vs Standard Language." In: *Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa'19), Turku, Finland.* Vol. 167, pp. 105–114.

Santini, Marina and Arne Jönsson (2020). "Pinning down text complexity: An Exploratory Study on the Registers of the Stockholm-Umeå Corpus (SUC)." In: *Register Studies* 2.2, pp. 306–349.

Santini, Marina, Arne Jönsson, and Evelina Rennes (2020). "Visualizing Facets of Text Complexity across Registers." In: *Proceedings of the 1st Workshop on Tools and Resources to Empower People with READing DIfficulties (READI@LREC). Marseille, France*, pp. 49–56.

Scarton, Carolina, Gustavo Henrique Paetzold, and Lucia Specia (2018). "Text Simplification from Professionally Produced Corpora." In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).* Miyazaki, Japan: European Language Resources Association (ELRA).

Shardlow, Matthew (2014). "Out in the open: Finding and categorising errors in the lexical simplification pipeline." In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 1583–1590.

Siddharthan, Advaith (2006). "Syntactic simplification and text cohesion." In: *Research on Language & Computation* 4.1, pp. 77–109.

Siddharthan, Advaith and Angrosh Mandya (2014). "Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules." In: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 722–731.

Simmons, Fiona and Chris Singleton (2000). "The Reading Comprehension Abilities of Dyslexic Students in Higher Education." In: *Dyslexia* 6.3, pp. 178–192.

Sjöholm, Johan (2012). " Probability as readability: A new machine learning approach to readability assessment for written Swedish." MA thesis. Linköping University.

Smith, Christian and Arne Jönsson (2011a). "Automatic Summarization As Means Of Simplifying Texts, An Evaluation For Swedish." In: *Proceedings of the 18th Nordic Conference of Computational Linguistics (NoDaLiDa-2010), Riga, Latvia.*

— (2011b). "Enhancing extraction based summarization with outside word space." In: *Proceedings of the 5th International Joint Conference on Natural Language Processing, Chiang Mai, Thailand.*

Solheim, Ida and Ketil Stølen (2007). *Technology research explained.* Tech. rep. A313. SINTEF ICT.

Song, Yangqiu and Dan Roth (2015). "Unsupervised Sparse Vector Densification for Short Text Similarity." In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1275–1280.

Soto, Christian and M Fernanda Rodríguez Poblete (2018). "Exploring the Meta-Comprehension Abilities of Students with Intellectual Disabilities." In: *International journal of special education* 33.2, pp. 233–247.

Souza, Cinthia M., Magali R. G. Meireles, and Paulo E. M. Almeida (2021). "A comparative study of abstractive and extractive summarization techniques to label subgroups on patent dataset." In: *Scientometrics* 126, pp. 135–156.

Specia, Lucia (2010). "Translating from Complex to Simplified Sentences." In: *Proceedings of the 9th international conference on Computational Processing of the Portuguese Language (PROPOR)*, pp. 30–39.

Språkrådet (2014). *Myndigheternas skrivregler*. https://www.isof.se/download/18.17dda5f1791cdbd2873a99/1620030264840/Mynd-skrivreg2014-1.pdf. Accessed: 2021-10-06.

Stadsrådsberedningen, Justitiedepartementet (n.d.). *Svarta listan. Ord och fraser som kan ersättas i författningsspråk. SB PM 2011:1*. https://www.regeringen.se/49b6b2/contentassets/7ea8845f72304d098640272782e5bc26/svarta-listan---ord-och-fraser-som-\kan-ersattas-i-forfattningssprak. Accessed: 2021-10-06.

Stadthagen-Gonzalez, Hans and Colin J Davis (2006). "The Bristol norms for age of acquisition, imageability, and familiarity." In: *Behavior research methods* 38.4, pp. 598–605.

Štajner, Sanja, Richard Evans, and Iustin Dornescu (2014). "Assessing conformance of Manually Simplified Corpora with User Requirements: the Case of Autistic Readers." In: *Proceedings of the Workshop on Automatic Text Simplification-Methods and Applications in the Multilingual Society (ATS-MA 2014)*, pp. 53–63.

Štajner, Sanja and Goran Glavaš (2017). "Leveraging event-based semantics for automated text simplification." In: *Expert Systems with Applications* 82, pp. 383–395.

The Programme for the International Assessment of Adult Competencies (PIAAC) (2012). *OECD Country Note: SWEDEN*. Tech. rep. Accessed: 2021-11-10.

Todirascu, Amalia, Thomas François, Delphine Bernhard, Núria Gala, and Anne-Laure Ligozat (2016). "Are cohesive features relevant for text readability evaluation?" In: *26th International Conference on Computational Linguistics (COLING 2016)*, pp. 987–997.

Utvecklings-störning (n.d.). https://www.fub.se/utvecklings-storning/. Accessed: 2021-11-10.

Vad är Afasi? (N.d.). https://www.hjarnfonden.se/om-hjarnan/diagnoser/afasi/. Accessed: 2021-11-10.

Vajjala, Sowmya and Ivana Lučić (2018). "OneStopEnglish corpus: A new corpus for automatic readability assessment and text simplification." In: *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications*, pp. 297–304.

Vellutino, Frank R (1987). “Dyslexia.” In: *Scientific American* 256.3, pp. 34–41.

Vellutino, Frank R, Jack M Fletcher, Margaret J Snowling, and Donna M Scanlon (2004). “Specific reading disability (dyslexia): What have we learned in the past four decades?” In: *Journal of child psychology and psychiatry* 45.1, pp. 2–40.

Vettori, Chiara and Ornella Mich (2011). “Supporting deaf children’s reading skills: the many challenges of text simplification.” In: *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pp. 283–284.

Volodina, Elena, Ildikó Pilán, Stian Rødven Eide, and Hannes Heidarsson (2014). “You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a Second Language.” In: *Proceedings of the third workshop on NLP for computer-assisted language learning*, pp. 128–144.

Volodina, Elena, Ildikó Pilán, Ingegerd Enström, Lorena Llozhi, Peter Lundkvist, Gunlög Sundberg, and Monica Sandell (2016). “SweLL on the rise: Swedish learner language corpus for European reference level studies.” In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16).*

Volodina, Elena, Ildikó Pilán, Lorena Llozhi, Baptiste Degryse, and Thomas François (2016). “SweLLex: second language learners’ productive vocabulary.” In: *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, pp. 76–84.

Wengelin, Åsa (2015). “Mot en evidensbaserad språkvård? En kritisk granskning av några svenska klarspråksråd i ljuset av forskning om läsbarhet och språkbearbetning.” In: *Sakprosa* 7.2.

Wigfield, Allan and John T. Guthrie (1997). “Relations of Children’s Motivation for Reading to the Amount and Breadth of Their Reading.” In: *Journal of Educational Psychology* 89.3, pp. 420–432.

Wilson, Michael (1988). “MRC psycholinguistic database: Machine-usable dictionary, version 2.00.” In: *Behavior research methods, instruments, & computers* 20.1, pp. 6–10.

Wolska, Magdalena and Yulia Clausen (2017). “Simplifying Metaphorical Language for Young Readers: A corpus study on news text.” In: *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 313–318.

Wubben, Sander, Antal van den Bosch, and Emiel Krahmer (July 2012). “Sentence Simplification by Monolingual Machine Translation.” In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* Jeju Island, Korea: Association for Computational Linguistics, pp. 1015–1024. URL: https://aclanthology.org/P12-1107.

Xu, Wei, Chris Callison-Burch, and Courtney Napoles (2015). "Problems in current text simplification research: New data can help." In: *Transactions of the Association for Computational Linguistics* 3, pp. 283–297.

Xu, Wei, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch (2016). "Optimizing Statistical Machine Translation for Text Simplification." In: *Transactions of the Association for Computational Linguistics* 4, pp. 401–415.

Yaneva, Victoria, Irina Temnikova, and Ruslan Mitkov (2015). "Accessible texts for autism: An eye-tracking study." In: *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pp. 49–57.

— (2016a). "A corpus of text data and gaze fixations from autistic and non-autistic adults." In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 480–487.

— (2016b). "Evaluating the readability of text simplification output for readers with cognitive disabilities." In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 293–299.

Yano, Yasukata, Michael H Long, and Steven Ross (1994). "The Effects of Simplified and Elaborated Texts on Foreign Language Reading Comprehension." In: *Language learning* 44.2, pp. 189–219.

Zaman, Farooq, Matthew Shardlow, Saeed-Ul Hassan, Naif Radi Aljohani, and Raheel Nawaz (2020). "HTSS: A novel hybrid text summarisation and simplification architecture." In: *Information Processing & Management* 57.6, p. 102351. ISSN: 0306-4573.

Zhang, Xingxing and Mirella Lapata (2017). "Sentence Simplification with Deep Reinforcement Learning." In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, pp. 584–594.

Zhang, Yaoyuan, Zhenxu Ye, Yansong Feng, Dongyan Zhao, and Rui Yan (2017). "A Constrained Sequence-to-Sequence Neural Model for Sentence Simplification." In: *CoRR* abs/1704.02312. arXiv: 1704.02312. URL: http://arxiv.org/abs/1704.02312.

Zhu, Zhemin, Delphine Bernhard, and Iryna Gurevych (2010). "A monolingual tree-based translation model for sentence simplification." In: *Proceedings of the 23rd international conference on computational linguistics*. Association for Computational Linguistics, pp. 1353–1361.

APPENDIX A

Examples of Automatically Generated Summaries

This appendix contains examples of automatically generated extractive and abstractive summaries. English translations are provided after each Swedish text.

Original text

Tillsammans med Kanadas nye stjärna Sidney Crosby duellerade 20-årige Alexander Ovetjkin om vem som skulle bli årets rookie i NHL. Poängmässigt vann den tuffe och spelskicklige ryssen i Capitals, laget från USA:s huvudstad. På 81 matcher under sin första säsong i NHL gjorde han 52 mål och passade till 54. 106 poäng en rookiesäsong är bland det absolut bästa en nykomling presterat i ligan. Han debuterade redan som 16-åring i den ryska ligan och trots locktoner om feta dollarbuntar från andra sidan av Atlanten stannade han kvar hemma i Moskva fram till och med förra säsongen. Det gjorde också att Ovetjkin var en redan klar elitspelare när han landade i Washington, där han för övrigt gjorde mål redan i sin debutmatch. Washington som lag räckte dock inte alls till för att ta sig till årets Stanley Cup-slutspel. Det är därför Alexander Ovetjkin nu kan komma och visa upp sig i Globen redan i kväll och därefter förstärka det ryska VM-laget i Riga. Det är alla fall var den ryska lagledningen hoppas på. Ovetjkin skulle egentligen ha anslutit till det ryska laget som på torsdagseftermiddagen kom från Helsingfors. Ryssland spelade sin första match i Hockey Games mot Finland i Helsingfors och vann i onsdags kväll. Nu blev Ovetjkin försenad från USA och landar inte i Stockholm förrän på fredagsmorgonen. Alexander Ovetjkin är bara en i raden av unga framgångsrika ryska hockeyspelare som kommer fram just nu. Tillsammans med bland andra Ilja Kovaltjuk, Atlanta och Jevgenij Malkin, Magnitogorsk i ryska ligan, har rysk hockey fått fram stjärnor som kan ta tillbaka landets hockeylandslag till en nivå vi inte sett sedan början av 1990-talet. Med Ovetjkin i laget blir det ännu svårare för Sverige att dels vinna dagens match i Globen och även att ta hem slutsegern i Euro Hockey Tour, som avgörs på måndag.

Together with Canada's new star Sidney Crosby, 20-year-old Alexander Ovetchkin dueled over who would be this year's rookie in the NHL. In terms of points, the tough and skilled Russian won in Capitals, the team from the US capital. In 81 games during his first season in the NHL, he scored 52 goals and fitted for 54. 106 points in a rookie season is among the absolute best a newcomer has performed in the league. He already made his debut as a 16-year-old in the Russian league and despite allusions to fat dollar bundles from across the Atlantic, he remained at home in Moscow until last season. It also made Ovetchkin an already clear elite player when he landed in Washington, where he scored goals in his debut match. Washington as a team, however, was not enough at all to get to this year's Stanley Cup playoffs. That is why Alexander Ovetchkin can now come and show himself in the Globe already tonight and then strengthen the Russian World Cup team in Riga. In any case, this is what the Russian team leadership hopes. Ovetchkin would actually have joined the Russian team that came from Helsinki on Thursday afternoon. Russia played its first match in the Hockey Games against Finland in Helsinki and won on Wednesday night. Now Ovetchkin was delayed from the US and will not land in Stockholm until Friday morning. Alexander Ovetchkin is just one of a string of young, successful Russian hockey players who are emerging right now. Together with, among others, Ilya Kovalchuk, Atlanta and Yevgeny Malkin, Magnitogorsk in the Russian league, Russian hockey has produced stars who can take the country's hockey national team back to a level we have not seen since the early 1990s. With Ovetchkin in the team, it will be even more difficult for Sweden to win today's match in the Globe and also to take home the final victory in the Euro Hockey Tour, which will be decided on Monday.

Extractive summary

Det är därför Alexander Ovetjkin nu kan komma och visa upp sig i Globen redan i kväll och därefter förstärka det ryska VM-laget i Riga. Alexander Ovetjkin är bara en i raden av unga framgångsrika ryska hockeyspelare som kommer fram just nu. Tillsammans med Kanadas nye stjärna Sidney Crosby duellerade 20-årige Alexander Ovetjkin om vem som skulle bli årets rookie i

NHL. Med Ovetjkin i laget blir det ännu svårare för Sverige att dels vinna dagens match i Globen och även att ta hem slutsegern i Euro Hockey Tour, som avgörs på måndag.

That is why Alexander Ovetchkin can now come and show himself in the Globe already tonight and then strengthen the Russian World Cup team in Riga. Alexander Ovetchkin is just one of a string of young successful Russian hockey players emerging right now. Together with Canada's new star Sidney Crosby, 20-year-old Alexander Ovetchkin dueled over who would become this year's rookie in the NHL. With Ovetchkin in the team, it will be even more difficult for Sweden to win today's match in the Globe and also to take home the final victory in the Euro Hockey Tour, which will be decided on Monday.

Abstractive summary

Alexander Ovetjkin är klar för spel i det ryska VM-laget i ishockey. Det innebär att han kommer att förstärka det ryska laget i kvällens match mot Sverige i Globen. Det blir Ovetjkins första match i Moskva och den blir tuffare än vad han hade tänkt sig innan han lämnade Washington Capitals för att spela i ryska ligan. Det är nu ryska hockeyligan, NHL, är på väg att ta upp sporten på den nivå man inte sett sedan början av 1990-talet. I dag möts Sverige och Ryssland i Globen i en match som kan bli avgörande för om Sverige ska vinna Euro Hockey Tour.

Alexander Ovetchkin is ready to play in the Russian World Cup team in ice hockey. This means that he will strengthen the Russian team in tonight's match against Sweden in the Globe. It will be Ovetchkin's first match in Moscow and it will be tougher than he had imagined before leaving the Washington Capitals to play in the Russian league. It is now the Russian hockey league, NHL, is about to take up the sport at the level not seen since the early 1990s. Today, Sweden and Russia meet in the Globe in a match that can be decisive for whether Sweden will win the Euro Hockey Tour.

APPENDIX B

Implemented Coh-Metrix Measures

The following Coh-Metrix features are implemented in our systems as of Mars 2022. Descriptions are gathered from the Coh-Metrix 3.0 documentation[1].

Referential cohesion

28 CRFNO1 Noun overlap, adjacent sentences, binary, mean

29 CRFAO1 Argument overlap, adjacent sentences, binary, mean

30 CRFSO1 Stem overlap, adjacent sentences, binary, mean

31 CRFNOa Noun overlap, all sentences, binary, mean

32 CRFAOa Argument overlap, all sentences, binary, mean

33 CRFSOa Stem overlap, all sentences, binary, mean

34 CRFCWO1 Content word overlap, adjacent sentences, proportional, mean

36 CRFCWOa Content word overlap, all sentences, proportional, mean

38 CRFANP1 Anaphor overlap, adjacent sentences

39 CRFANPa Anaphor overlap, all sentences

LSA

[1] http://cohmetrix.memphis.edu/cohmetrixhome/documentation_indices.html#General%20overview, last accessed: 2022-03-15.

40 LSASS1 LSA overlap, adjacent sentences, mean

41 LSASS1d LSA overlap, adjacent sentences, standard deviation

42 LSASSp LSA overlap, all sentences in paragraph, mean

43 LSASSpd LSA overlap, all sentences in paragraph, standard deviation

46 LSAGN LSA given/new, sentences, mean

Connectives

52 CNCAll All connectives incidence

53 CNCCaus Causal connectives incidence

55 CNCADC Adversative and contrastive connectives incidence

56 CNCTemp Temporal connectives incidence

58 CNCAdd Additive connectives incidence

Department of Computer and Information Science
Linköpings universitet

Dissertations

Linköping Studies in Science and Technology
Linköping Studies in Arts and Sciences
Linköping Studies in Statistics
Linköping Studies in Information Science

Linköping Studies in Science and Technology

No 14 **Anders Haraldsson:** A Program Manipulation System Based on Partial Evaluation, 1977, ISBN 91-7372-144-1.

No 17 **Bengt Magnhagen:** Probability Based Verification of Time Margins in Digital Designs, 1977, ISBN 91-7372-157-3.

No 18 **Mats Cedwall**: Semantisk analys av process-beskrivningar i naturligt språk, 1977, ISBN 91- 7372-168-9.

No 22 **Jaak Urmi:** A Machine Independent LISP Compiler and its Implications for Ideal Hardware, 1978, ISBN 91-7372-188-3.

No 33 **Tore Risch:** Compilation of Multiple File Queries in a Meta-Database System, 1978, ISBN 91- 7372-232-4.

No 51 **Erland Jungert:** Synthesizing Database Structures from a User Oriented Data Model, 1980, ISBN 91-7372-387-8.

No 54 **Sture Hägglund:** Contributions to the Development of Methods and Tools for Interactive Design of Applications Software, 1980, ISBN 91-7372-404-1.

No 55 **Pär Emanuelson:** Performance Enhancement in a Well-Structured Pattern Matcher through Partial Evaluation, 1980, ISBN 91-7372-403-3.

No 58 **Bengt Johnsson, Bertil Andersson:** The Human-Computer Interface in Commercial Systems, 1981, ISBN 91-7372-414-9.

No 69 **H. Jan Komorowski:** A Specification of an Abstract Prolog Machine and its Application to Partial Evaluation, 1981, ISBN 91-7372-479-3.

No 71 **René Reboh:** Knowledge Engineering Techniques and Tools for Expert Systems, 1981, ISBN 91-7372-489-0.

No 77 **Östen Oskarsson:** Mechanisms of Modifiability in large Software Systems, 1982, ISBN 91- 7372-527-7.

No 94 **Hans Lunell:** Code Generator Writing Systems, 1983, ISBN 91-7372-652-4.

No 97 **Andrzej Lingas:** Advances in Minimum Weight Triangulation, 1983, ISBN 91-7372-660-5.

No 109 **Peter Fritzson:** Towards a Distributed Programming Environment based on Incremental Compilation, 1984, ISBN 91-7372-801-2.

No 111 **Erik Tengvald:** The Design of Expert Planning Systems. An Experimental Operations Planning System for Turning, 1984, ISBN 91-7372- 805-5.

No 155 **Christos Levcopoulos:** Heuristics for Minimum Decompositions of Polygons, 1987, ISBN 91-7870-133-3.

No 165 **James W. Goodwin:** A Theory and System for Non-Monotonic Reasoning, 1987, ISBN 91-7870-183-X.

No 170 **Zebo Peng:** A Formal Methodology for Automated Synthesis of VLSI Systems, 1987, ISBN 91-7870-225-9.

No 174 **Johan Fagerström:** A Paradigm and System for Design of Distributed Systems, 1988, ISBN 91-7870-301-8.

No 192 **Dimiter Driankov:** Towards a Many Valued Logic of Quantified Belief, 1988, ISBN 91-7870-374-3.

No 213 **Lin Padgham:** Non-Monotonic Inheritance for an Object Oriented Knowledge Base, 1989, ISBN 91-7870-485-5.

No 214 **Tony Larsson:** A Formal Hardware Description and Verification Method, 1989, ISBN 91-7870-517-7.

No 221 **Michael Reinfrank:** Fundamentals and Logical Foundations of Truth Maintenance, 1989, ISBN 91-7870-546-0.

No 239 **Jonas Löwgren:** Knowledge-Based Design Support and Discourse Management in User Interface Management Systems, 1991, ISBN 91-7870-720-X.

No 244 **Henrik Eriksson:** Meta-Tool Support for Knowledge Acquisition, 1991, ISBN 91-7870-746-3.

No 252 **Peter Eklund:** An Epistemic Approach to Interactive Design in Multiple Inheritance Hierarchies, 1991, ISBN 91-7870-784-6.

No 258 **Patrick Doherty:** NML3 - A Non-Monotonic Formalism with Explicit Defaults, 1991, ISBN 91-7870-816-8.

No 260 **Nahid Shahmehri:** Generalized Algorithmic Debugging, 1991, ISBN 91-7870-828-1.

No 264 **Nils Dahlbäck:** Representation of Discourse-Cognitive and Computational Aspects, 1992, ISBN 91-7870-850-8.

No 265 **Ulf Nilsson:** Abstract Interpretations and Abstract Machines: Contributions to a Methodology for the Implementation of Logic Programs, 1992, ISBN 91-7870-858-3.

No 270 **Ralph Rönnquist:** Theory and Practice of Tense-bound Object References, 1992, ISBN 91-7870-873-7.

No 273 **Björn Fjellborg:** Pipeline Extraction for VLSI Data Path Synthesis, 1992, ISBN 91-7870-880-X.

No 276 **Staffan Bonnier:** A Formal Basis for Horn Clause Logic with External Polymorphic Functions, 1992, ISBN 91-7870-896-6.

No 277 **Kristian Sandahl:** Developing Knowledge Management Systems with an Active Expert Methodology, 1992, ISBN 91-7870-897-4.

No 281 **Christer Bäckström:** Computational Complexity of Reasoning about Plans, 1992, ISBN 91-7870-979-2.

No 292 **Mats Wirén:** Studies in Incremental Natural Language Analysis, 1992, ISBN 91-7871-027-8.

No 297 **Mariam Kamkar:** Interprocedural Dynamic Slicing with Applications to Debugging and Testing, 1993, ISBN 91-7871-065-0.

No 302 **Tingting Zhang:** A Study in Diagnosis Using Classification and Defaults, 1993, ISBN 91-7871-078-2.

No 312 **Arne Jönsson:** Dialogue Management for Natural Language Interfaces - An Empirical Approach, 1993, ISBN 91-7871-110-X.

No 338 **Simin Nadjm-Tehrani:** Reactive Systems in Physical Environments: Compositional Modelling and Framework for Verification, 1994, ISBN 91-7871-237-8.

No 371 **Bengt Savén:** Business Models for Decision Support and Learning. A Study of Discrete-Event Manufacturing Simulation at Asea/ABB 1968-1993, 1995, ISBN 91-7871-494-X.

No 375 **Ulf Söderman:** Conceptual Modelling of Mode Switching Physical Systems, 1995, ISBN 91-7871-516-4.

No 383 **Andreas Kågedal:** Exploiting Groundness in Logic Programs, 1995, ISBN 91-7871-538-5.

No 396 **George Fodor:** Ontological Control, Description, Identification and Recovery from Problematic Control Situations, 1995, ISBN 91-7871-603-9.

No 413 **Mikael Pettersson:** Compiling Natural Semantics, 1995, ISBN 91-7871-641-1.

No 414 **Xinli Gu:** RT Level Testability Improvement by Testability Analysis and Transformations, 1996, ISBN 91-7871-654-3.

No 416 **Hua Shu:** Distributed Default Reasoning, 1996, ISBN 91-7871-665-9.

No 429 **Jaime Villegas:** Simulation Supported Industrial Training from an Organisational Learning Perspective - Development and Evaluation of the SSIT Method, 1996, ISBN 91-7871-700-0.

No 431 **Peter Jonsson:** Studies in Action Planning: Algorithms and Complexity, 1996, ISBN 91-7871-704-3.

No 437 **Johan Boye:** Directional Types in Logic Programming, 1996, ISBN 91-7871-725-6.

No 439 **Cecilia Sjöberg:** Activities, Voices and Arenas: Participatory Design in Practice, 1996, ISBN 91-7871-728-0.

No 448 **Patrick Lambrix:** Part-Whole Reasoning in Description Logics, 1996, ISBN 91-7871-820-1.

No 452 **Kjell Orsborn:** On Extensible and Object-Relational Database Technology for Finite Element Analysis Applications, 1996, ISBN 91-7871-827-9.

No 459 **Olof Johansson:** Development Environments for Complex Product Models, 1996, ISBN 91-7871-855-4.

No 461 **Lena Strömbäck:** User-Defined Constructions in Unification-Based Formalisms, 1997, ISBN 91-7871-857-0.

No 462 **Lars Degerstedt:** Tabulation-based Logic Programming: A Multi-Level View of Query Answering, 1996, ISBN 91-7871-858-9.

No 475 **Fredrik Nilsson:** Strategi och ekonomisk styrning - En studie av hur ekonomiska styrsystem utformas och används efter företagsförvärv, 1997, ISBN 91-7871-914-3.

No 480 **Mikael Lindvall:** An Empirical Study of Requirements-Driven Impact Analysis in Object-Oriented Software Evolution, 1997, ISBN 91-7871-927-5.

No 485 **Göran Forslund:** Opinion-Based Systems: The Cooperative Perspective on Knowledge-Based Decision Support, 1997, ISBN 91-7871-938-0.

No 494 **Martin Sköld:** Active Database Management Systems for Monitoring and Control, 1997, ISBN 91-7219-002-7.

No 495 **Hans Olsén:** Automatic Verification of Petri Nets in a CLP framework, 1997, ISBN 91-7219-011-6.

No 498 **Thomas Drakengren:** Algorithms and Complexity for Temporal and Spatial Formalisms, 1997, ISBN 91-7219-019-1.

No 502 **Jakob Axelsson:** Analysis and Synthesis of Heterogeneous Real-Time Systems, 1997, ISBN 91-7219-035-3.

No 503 **Johan Ringström:** Compiler Generation for Data-Parallel Programming Languages from Two-Level Semantics Specifications, 1997, ISBN 91-7219-045-0.

No 512 **Anna Moberg:** Närhet och distans - Studier av kommunikationsmönster i satellitkontor och flexibla kontor, 1997, ISBN 91-7219-119-8.

No 520 **Mikael Ronström:** Design and Modelling of a Parallel Data Server for Telecom Applications, 1998, ISBN 91-7219-169-4.

No 522 **Niclas Ohlsson:** Towards Effective Fault Prevention - An Empirical Study in Software Engineering, 1998, ISBN 91-7219-176-7.

No 526 **Joachim Karlsson:** A Systematic Approach for Prioritizing Software Requirements, 1998, ISBN 91-7219-184-8.

No 530 **Henrik Nilsson:** Declarative Debugging for Lazy Functional Languages, 1998, ISBN 91-7219-197-X.

No 555 **Jonas Hallberg:** Timing Issues in High-Level Synthesis, 1998, ISBN 91-7219-369-7.

No 561 **Ling Lin:** Management of 1-D Sequence Data - From Discrete to Continuous, 1999, ISBN 91-7219-402-2.

No 563 **Eva L Ragnemalm:** Student Modelling based on Collaborative Dialogue with a Learning Companion, 1999, ISBN 91-7219-412-X.

No 567 **Jörgen Lindström:** Does Distance matter? On geographical dispersion in organisations, 1999, ISBN 91-7219-439-1.

No 582 **Vanja Josifovski:** Design, Implementation and Evaluation of a Distributed Mediator System for Data Integration, 1999, ISBN 91-7219-482-0.

No 589 **Rita Kovordányi:** Modeling and Simulating Inhibitory Mechanisms in Mental Image Reinterpretation - Towards Cooperative Human-Computer Creativity, 1999, ISBN 91-7219-506-1.

No 592 **Mikael Ericsson:** Supporting the Use of Design Knowledge - An Assessment of Commenting Agents, 1999, ISBN 91-7219-532-0.

No 593 **Lars Karlsson:** Actions, Interactions and Narratives, 1999, ISBN 91-7219-534-7.

No 594 **C. G. Mikael Johansson:** Social and Organizational Aspects of Requirements Engineering Methods - A practice-oriented approach, 1999, ISBN 91-7219-541-X.

No 595 **Jörgen Hansson:** Value-Driven Multi-Class Overload Management in Real-Time Database Systems, 1999, ISBN 91-7219-542-8.

No 596 **Niklas Hallberg:** Incorporating User Values in the Design of Information Systems and Services in the Public Sector: A Methods Approach, 1999, ISBN 91-7219-543-6.

No 597 **Vivian Vimarlund:** An Economic Perspective on the Analysis of Impacts of Information Technology: From Case Studies in Health-Care towards General Models and Theories, 1999, ISBN 91-7219-544-4.

No 598 **Johan Jenvald:** Methods and Tools in Computer-Supported Taskforce Training, 1999, ISBN 91-7219-547-9.

No 607 **Magnus Merkel:** Understanding and enhancing translation by parallel text processing, 1999, ISBN 91-7219-614-9.

No 611 **Silvia Coradeschi:** Anchoring symbols to sensory data, 1999, ISBN 91-7219-623-8.

No 613 **Man Lin:** Analysis and Synthesis of Reactive Systems: A Generic Layered Architecture Perspective, 1999, ISBN 91-7219-630-0.

No 618 **Jimmy Tjäder:** Systemimplementering i praktiken - En studie av logiker i fyra projekt, 1999, ISBN 91-7219-657-2.

No 627 **Vadim Engelson:** Tools for Design, Interactive Simulation, and Visualization of Object-Oriented Models in Scientific Computing, 2000, ISBN 91-7219-709-9.

No 637 **Esa Falkenroth:** Database Technology for Control and Simulation, 2000, ISBN 91-7219-766-8.

No 639 **Per-Arne Persson:** Bringing Power and Knowledge Together: Information Systems Design for Autonomy and Control in Command Work, 2000, ISBN 91-7219-796-X.

No 660 **Erik Larsson:** An Integrated System-Level Design for Testability Methodology, 2000, ISBN 91-7219-890-7.

No 688 **Marcus Bjäreland:** Model-based Execution Monitoring, 2001, ISBN 91-7373-016-5.

No 689 **Joakim Gustafsson:** Extending Temporal Action Logic, 2001, ISBN 91-7373-017-3.

No 720 **Carl-Johan Petri:** Organizational Information Provision - Managing Mandatory and Discretionary Use of Information Technology, 2001, ISBN 91-7373-126-9.

No 724 **Paul Scerri:** Designing Agents for Systems with Adjustable Autonomy, 2001, ISBN 91-7373-207-9.

No 725 **Tim Heyer:** Semantic Inspection of Software Artifacts: From Theory to Practice, 2001, ISBN 91-7373-208-7.

No 726 **Pär Carlshamre:** A Usability Perspective on Requirements Engineering - From Methodology to Product Development, 2001, ISBN 91-7373-212-5.

No 732 **Juha Takkinen:** From Information Management to Task Management in Electronic Mail, 2002, ISBN 91-7373-258-3.

No 745 **Johan Åberg:** Live Help Systems: An Approach to Intelligent Help for Web Information Systems, 2002, ISBN 91-7373-311-3.

No 746 **Rego Granlund:** Monitoring Distributed Teamwork Training, 2002, ISBN 91-7373-312-1.

No 757 **Henrik André-Jönsson:** Indexing Strategies for Time Series Data, 2002, ISBN 917373-346-6.

No 747 **Anneli Hagdahl:** Development of IT-supported Interorganisational Collaboration - A Case Study in the Swedish Public Sector, 2002, ISBN 91-7373-314-8.

No 749 **Sofie Pilemalm:** Information Technology for Non-Profit Organisations - Extended Participatory Design of an Information System for Trade Union Shop Stewards, 2002, ISBN 91-7373-318-0.

No 765 **Stefan Holmlid:** Adapting users: Towards a theory of use quality, 2002, ISBN 91-7373-397-0.

No 771 **Magnus Morin:** Multimedia Representations of Distributed Tactical Operations, 2002, ISBN 91-7373-421-7.

No 772 **Pawel Pietrzak:** A Type-Based Framework for Locating Errors in Constraint Logic Programs, 2002, ISBN 91-7373-422-5.

No 758 **Erik Berglund:** Library Communication Among Programmers Worldwide, 2002, ISBN 91-7373-349-0.

No 774 **Choong-ho Yi:** Modelling Object-Oriented Dynamic Systems Using a Logic-Based Framework, 2002, ISBN 91-7373-424-1.

No 779 **Mathias Broxvall:** A Study in the Computational Complexity of Temporal Reasoning, 2002, ISBN 91-7373-440-3.

No 793 **Asmus Pandikow:** A Generic Principle for Enabling Interoperability of Structured and Object-Oriented Analysis and Design Tools, 2002, ISBN 91-7373-479-9.

No 785 **Lars Hult:** Publika Informationstjänster. En studie av den Internetbaserade encyklopedins bruksegenskaper, 2003, ISBN 91-7373-461-6.

No 800 **Lars Taxén:** A Framework for the Coordination of Complex Systems´ Development, 2003, ISBN 91-7373-604-X.

No 808 **Klas Gäre:** Tre perspektiv på förväntningar och förändringar i samband med införande av informationssystem, 2003, ISBN 91-7373-618-X.

No 821 **Mikael Kindborg:** Concurrent Comics - programming of social agents by children, 2003, ISBN 91-7373-651-1.

No 823 **Christina Ölvingson:** On Development of Information Systems with GIS Functionality in Public Health Informatics: A Requirements Engineering Approach, 2003, ISBN 91-7373-656-2.

No 828 **Tobias Ritzau:** Memory Efficient Hard Real-Time Garbage Collection, 2003, ISBN 91-7373-666-X.

No 833 **Paul Pop:** Analysis and Synthesis of Communication-Intensive Heterogeneous Real-Time Systems, 2003, ISBN 91-7373-683-X.

No 852 **Johan Moe:** Observing the Dynamic Behaviour of Large Distributed Systems to Improve Development and Testing – An Empirical Study in Software Engineering, 2003, ISBN 91-7373-779-8.

No 867 **Erik Herzog:** An Approach to Systems Engineering Tool Data Representation and Exchange, 2004, ISBN 91-7373-929-4.

No 872 **Aseel Berglund:** Augmenting the Remote Control: Studies in Complex Information Navigation for Digital TV, 2004, ISBN 91-7373-940-5.

No 869 **Jo Skåmedal:** Telecommuting's Implications on Travel and Travel Patterns, 2004, ISBN 91-7373-935-9.

No 870 **Linda Askenäs:** The Roles of IT - Studies of Organising when Implementing and Using Enterprise Systems, 2004, ISBN 91-7373-936-7.

No 874 **Annika Flycht-Eriksson:** Design and Use of Ontologies in Information-Providing Dialogue Systems, 2004, ISBN 91-7373-947-2.

No 873 **Peter Bunus:** Debugging Techniques for Equation-Based Languages, 2004, ISBN 91-7373-941-3.

No 876 **Jonas Mellin:** Resource-Predictable and Efficient Monitoring of Events, 2004, ISBN 91-7373-956-1.

No 883 **Magnus Bång:** Computing at the Speed of Paper: Ubiquitous Computing Environments for Healthcare Professionals, 2004, ISBN 91-7373-971-5.

No 882 **Robert Eklund:** Disfluency in Swedish human-human and human-machine travel booking dialogues, 2004, ISBN 91-7373-966-9.

No 887 **Anders Lindström:** English and other Foreign Linguistic Elements in Spoken Swedish. Studies of Productive Processes and their Modelling using Finite-State Tools, 2004, ISBN 91-7373-981-2.

No 889 **Zhiping Wang:** Capacity-Constrained Production-inventory systems - Modelling and Analysis in both a traditional and an e-business context, 2004, ISBN 91-85295-08-6.

No 893 **Pernilla Qvarfordt:** Eyes on Multimodal Interaction, 2004, ISBN 91-85295-30-2.

No 910 **Magnus Kald:** In the Borderland between Strategy and Management Control - Theoretical Framework and Empirical Evidence, 2004, ISBN 91-85295-82-5.

No 918 **Jonas Lundberg:** Shaping Electronic News: Genre Perspectives on Interaction Design, 2004, ISBN 91-85297-14-3.

No 900 **Mattias Arvola:** Shades of use: The dynamics of interaction design for sociable use, 2004, ISBN 91-85295-42-6.

No 920 **Luis Alejandro Cortés:** Verification and Scheduling Techniques for Real-Time Embedded Systems, 2004, ISBN 91-85297-21-6.

No 929 **Diana Szentivanyi:** Performance Studies of Fault-Tolerant Middleware, 2005, ISBN 91-85297-58-5.

No 933 **Mikael Cäker:** Management Accounting as Constructing and Opposing Customer Focus: Three Case Studies on Management Accounting and Customer Relations, 2005, ISBN 91-85297-64-X.

No 937 **Jonas Kvarnström:** TALplanner and Other Extensions to Temporal Action Logic, 2005, ISBN 91-85297-75-5.

No 938 **Bourhane Kadmiry:** Fuzzy Gain-Scheduled Visual Servoing for Unmanned Helicopter, 2005, ISBN 91-85297-76-3.

No 945 **Gert Jervan:** Hybrid Built-In Self-Test and Test Generation Techniques for Digital Systems, 2005, ISBN 91-85297-97-6.

No 946 **Anders Arpteg:** Intelligent Semi-Structured Information Extraction, 2005, ISBN 91-85297-98-4.

No 947 **Ola Angelsmark:** Constructing Algorithms for Constraint Satisfaction and Related Problems - Methods and Applications, 2005, ISBN 91-85297-99-2.

No 963 **Calin Curescu:** Utility-based Optimisation of Resource Allocation for Wireless Networks, 2005, ISBN 91-85457-07-8.

No 972 **Björn Johansson:** Joint Control in Dynamic Situations, 2005, ISBN 91-85457-31-0.

No 974 **Dan Lawesson:** An Approach to Diagnosability Analysis for Interacting Finite State Systems, 2005, ISBN 91-85457-39-6.

No 979 **Claudiu Duma:** Security and Trust Mechanisms for Groups in Distributed Services, 2005, ISBN 91-85457-54-X.

No 983 **Sorin Manolache:** Analysis and Optimisation of Real-Time Systems with Stochastic Behaviour, 2005, ISBN 91-85457-60-4.

No 986 **Yuxiao Zhao:** Standards-Based Application Integration for Business-to-Business Communications, 2005, ISBN 91-85457-66-3.

No 1004 **Patrik Haslum:** Admissible Heuristics for Automated Planning, 2006, ISBN 91-85497-28-2.

No 1005 **Aleksandra Tešanovic:** Developing Reusable and Reconfigurable Real-Time Software using Aspects and Components, 2006, ISBN 91-85497-29-0.

No 1008 **David Dinka:** Role, Identity and Work: Extending the design and development agenda, 2006, ISBN 91-85497-42-8.

No 1009 **Iakov Nakhimovski:** Contributions to the Modeling and Simulation of Mechanical Systems with Detailed Contact Analysis, 2006, ISBN 91-85497-43-X.

No 1013 **Wilhelm Dahllöf:** Exact Algorithms for Exact Satisfiability Problems, 2006, ISBN 91-85523-97-6.

No 1016 **Levon Saldamli:** PDEModelica - A High-Level Language for Modeling with Partial Differential Equations, 2006, ISBN 91-85523-84-4.

No 1017 **Daniel Karlsson:** Verification of Component-based Embedded System Designs, 2006, ISBN 91-85523-79-8

No 1018 **Ioan Chisalita:** Communication and Networking Techniques for Traffic Safety Systems, 2006, ISBN 91-85523-77-1.

No 1019 **Tarja Susi:** The Puzzle of Social Activity - The Significance of Tools in Cognition and Cooperation, 2006, ISBN 91-85523-71-2.

No 1021 **Andrzej Bednarski:** Integrated Optimal Code Generation for Digital Signal Processors, 2006, ISBN 91-85523-69-0.

No 1022 **Peter Aronsson:** Automatic Parallelization of Equation-Based Simulation Programs, 2006, ISBN 91-85523-68-2.

No 1030 **Robert Nilsson:** A Mutation-based Framework for Automated Testing of Timeliness, 2006, ISBN 91-85523-35-6.

No 1034 **Jon Edvardsson:** Techniques for Automatic Generation of Tests from Programs and Specifications, 2006, ISBN 91-85523-31-3.

No 1035 **Vaida Jakoniene:** Integration of Biological Data, 2006, ISBN 91-85523-28-3.

No 1045 **Genevieve Gorrell:** Generalized Hebbian Algorithms for Dimensionality Reduction in Natural Language Processing, 2006, ISBN 91-85643-88-2.

No 1051 **Yu-Hsing Huang:** Having a New Pair of Glasses - Applying Systemic Accident Models on Road Safety, 2006, ISBN 91-85643-64-5.

No 1054 **Åsa Hedenskog:** Perceive those things which cannot be seen - A Cognitive Systems Engineering perspective on requirements management, 2006, ISBN 91-85643-57-2.

No 1061 **Cécile Åberg:** An Evaluation Platform for Semantic Web Technology, 2007, ISBN 91-85643-31-9.

No 1073 **Mats Grindal:** Handling Combinatorial Explosion in Software Testing, 2007, ISBN 978-91-85715-74-9.

No 1075 **Almut Herzog:** Usable Security Policies for Runtime Environments, 2007, ISBN 978-91-85715-65-7.

No 1079 **Magnus Wahlström:** Algorithms, measures, and upper bounds for Satisfiability and related problems, 2007, ISBN 978-91-85715-55-8.

No 1083 **Jesper Andersson:** Dynamic Software Architectures, 2007, ISBN 978-91-85715-46-6.

No 1086 **Ulf Johansson:** Obtaining Accurate and Comprehensible Data Mining Models - An Evolutionary Approach, 2007, ISBN 978-91-85715-34-3.

No 1089 **Traian Pop:** Analysis and Optimisation of Distributed Embedded Systems with Heterogeneous Scheduling Policies, 2007, ISBN 978-91-85715-27-5.

No 1091 **Gustav Nordh:** Complexity Dichotomies for CSP-related Problems, 2007, ISBN 978-91-85715-20-6.

No 1106 **Per Ola Kristensson:** Discrete and Continuous Shape Writing for Text Entry and Control, 2007, ISBN 978-91-85831-77-7.

No 1110 **He Tan:** Aligning Biomedical Ontologies, 2007, ISBN 978-91-85831-56-2.

No 1112 **Jessica Lindblom:** Minding the body - Interacting socially through embodied action, 2007, ISBN 978-91-85831-48-7.

No 1113 **Pontus Wärnestål:** Dialogue Behavior Management in Conversational Recommender Systems, 2007, ISBN 978-91-85831-47-0.

No 1120 **Thomas Gustafsson:** Management of Real-Time Data Consistency and Transient Overloads in Embedded Systems, 2007, ISBN 978-91-85831-33-3.

No 1127 **Alexandru Andrei:** Energy Efficient and Predictable Design of Real-time Embedded Systems, 2007, ISBN 978-91-85831-06-7.

No 1139 **Per Wikberg:** Eliciting Knowledge from Experts in Modeling of Complex Systems: Managing Variation and Interactions, 2007, ISBN 978-91-85895-66-3.

No 1143 **Mehdi Amirijoo:** QoS Control of Real-Time Data Services under Uncertain Workload, 2007, ISBN 978-91-85895-49-6.

No 1150 **Sanny Syberfeldt:** Optimistic Replication with Forward Conflict Resolution in Distributed Real-Time Databases, 2007, ISBN 978-91-85895-27-4.

No 1155 **Beatrice Alenljung:** Envisioning a Future Decision Support System for Requirements Engineering - A Holistic and Human-centred Perspective, 2008, ISBN 978-91-85895-11-3.

No 1156 **Artur Wilk:** Types for XML with Application to Xcerpt, 2008, ISBN 978-91-85895-08-3.

No 1183 **Adrian Pop:** Integrated Model-Driven Development Environments for Equation-Based Object-Oriented Languages, 2008, ISBN 978-91-7393-895-2.

No 1185 **Jörgen Skågeby:** Gifting Technologies - Ethnographic Studies of End-users and Social Media Sharing, 2008, ISBN 978-91-7393-892-1.

No 1187 **Imad-Eldin Ali Abugessaisa:** Analytical tools and information-sharing methods supporting road safety organizations, 2008, ISBN 978-91-7393-887-7.

No 1204 **H. Joe Steinhauer:** A Representation Scheme for Description and Reconstruction of Object Configurations Based on Qualitative Relations, 2008, ISBN 978-91-7393-823-5.

No 1222 **Anders Larsson:** Test Optimization for Core-based System-on-Chip, 2008, ISBN 978-91-7393-768-9.

No 1238 **Andreas Borg:** Processes and Models for Capacity Requirements in Telecommunication Systems, 2009, ISBN 978-91-7393-700-9.

No 1240 **Fredrik Heintz:** DyKnow: A Stream-Based Knowledge Processing Middleware Framework, 2009, ISBN 978-91-7393-696-5.

No 1241 **Birgitta Lindström:** Testability of Dynamic Real-Time Systems, 2009, ISBN 978-91-7393-695-8.

No 1244 **Eva Blomqvist:** Semi-automatic Ontology Construction based on Patterns, 2009, ISBN 978-91-7393-683-5.

No 1249 **Rogier Woltjer:** Functional Modeling of Constraint Management in Aviation Safety and Command and Control, 2009, ISBN 978-91-7393-659-0.

No 1260 **Gianpaolo Conte:** Vision-Based Localization and Guidance for Unmanned Aerial Vehicles, 2009, ISBN 978-91-7393-603-3.

No 1262 **AnnMarie Ericsson:** Enabling Tool Support for Formal Analysis of ECA Rules, 2009, ISBN 978-91-7393-598-2.

No 1266 **Jiri Trnka:** Exploring Tactical Command and Control: A Role-Playing Simulation Approach, 2009, ISBN 978-91-7393-571-5.

No 1268 **Bahlol Rahimi:** Supporting Collaborative Work through ICT - How End-users Think of and Adopt Integrated Health Information Systems, 2009, ISBN 978-91-7393-550-0.

No 1274 **Fredrik Kuivinen:** Algorithms and Hardness Results for Some Valued CSPs, 2009, ISBN 978-91-7393-525-8.

No 1281 **Gunnar Mathiason:** Virtual Full Replication for Scalable Distributed Real-Time Databases, 2009, ISBN 978-91-7393-503-6.

No 1290 **Viacheslav Izosimov:** Scheduling and Optimization of Fault-Tolerant Distributed Embedded Systems, 2009, ISBN 978-91-7393-482-4.

No 1294 **Johan Thapper:** Aspects of a Constraint Optimisation Problem, 2010, ISBN 978-91-7393-464-0.

No 1306 **Susanna Nilsson:** Augmentation in the Wild: User Centered Development and Evaluation of Augmented Reality Applications, 2010, ISBN 978-91-7393-416-9.

No 1313 **Christer Thörn:** On the Quality of Feature Models, 2010, ISBN 978-91-7393-394-0.

No 1321 **Zhiyuan He:** Temperature Aware and Defect-Probability Driven Test Scheduling for System-on-Chip, 2010, ISBN 978-91-7393-378-0.

No 1333 **David Broman:** Meta-Languages and Semantics for Equation-Based Modeling and Simulation, 2010, ISBN 978-91-7393-335-3.

No 1337 **Alexander Siemers:** Contributions to Modelling and Visualisation of Multibody Systems Simulations with Detailed Contact Analysis, 2010, ISBN 978-91-7393-317-9.

No 1354 **Mikael Asplund:** Disconnected Discoveries: Availability Studies in Partitioned Networks, 2010, ISBN 978-91-7393-278-3.

No 1359 **Jana Rambusch**: Mind Games Extended: Understanding Gameplay as Situated Activity, 2010, ISBN 978-91-7393-252-3.

No 1373 **Sonia Sangari**: Head Movement Correlates to Focus Assignment in Swedish, 2011, ISBN 978-91-7393-154-0.

No 1374 **Jan-Erik Källhammer**: Using False Alarms when Developing Automotive Active Safety Systems, 2011, ISBN 978-91-7393-153-3.

No 1375 **Mattias Eriksson**: Integrated Code Generation, 2011, ISBN 978-91-7393-147-2.

No 1381 **Ola Leifler**: Affordances and Constraints of Intelligent Decision Support for Military Command and Control – Three Case Studies of Support Systems, 2011, ISBN 978-91-7393-133-5.

No 1386 **Soheil Samii**: Quality-Driven Synthesis and Optimization of Embedded Control Systems, 2011, ISBN 978-91-7393-102-1.

No 1419 **Erik Kuiper**: Geographic Routing in Intermittently-connected Mobile Ad Hoc Networks: Algorithms and Performance Models, 2012, ISBN 978-91-7519-981-8.

No 1451 **Sara Stymne**: Text Harmonization Strategies for Phrase-Based Statistical Machine Translation, 2012, ISBN 978-91-7519-887-3.

No 1455 **Alberto Montebelli**: Modeling the Role of Energy Management in Embodied Cognition, 2012, ISBN 978-91-7519-882-8.

No 1465 **Mohammad Saifullah**: Biologically-Based Interactive Neural Network Models for Visual Attention and Object Recognition, 2012, ISBN 978-91-7519-838-5.

No 1490 **Tomas Bengtsson**: Testing and Logic Optimization Techniques for Systems on Chip, 2012, ISBN 978-91-7519-742-5.

No 1481 **David Byers**: Improving Software Security by Preventing Known Vulnerabilities, 2012, ISBN 978-91-7519-784-5.

No 1496 **Tommy Färnqvist**: Exploiting Structure in CSP-related Problems, 2013, ISBN 978-91-7519-711-1.

No 1503 **John Wilander**: Contributions to Specification, Implementation, and Execution of Secure Software, 2013, ISBN 978-91-7519-681-7.

No 1506 **Magnus Ingmarsson**: Creating and Enabling the Useful Service Discovery Experience, 2013, ISBN 978-91-7519-662-6.

No 1547 **Wladimir Schamai**: Model-Based Verification of Dynamic System Behavior against Requirements: Method, Language, and Tool, 2013, ISBN 978-91-7519-505-6.

No 1551 **Henrik Svensson**: Simulations, 2013, ISBN 978-91-7519-491-2.

No 1559 **Sergiu Rafiliu**: Stability of Adaptive Distributed Real-Time Systems with Dynamic Resource Management, 2013, ISBN 978-91-7519-471-4.

No 1581 **Usman Dastgeer**: Performance-aware Component Composition for GPU-based Systems, 2014, ISBN 978-91-7519-383-0.

No 1602 **Cai Li**: Reinforcement Learning of Locomotion based on Central Pattern Generators, 2014, ISBN 978-91-7519-313-7.

No 1652 **Roland Samlaus**: An Integrated Development Environment with Enhanced Domain-Specific Interactive Model Validation, 2015, ISBN 978-91-7519-090-7.

No 1663 **Hannes Uppman**: On Some Combinatorial Optimization Problems: Algorithms and Complexity, 2015, ISBN 978-91-7519-072-3.

No 1664 **Martin Sjölund**: Tools and Methods for Analysis, Debugging, and Performance Improvement of Equation-Based Models, 2015, ISBN 978-91-7519-071-6.

No 1666 **Kristian Stavåker**: Contributions to Simulation of Modelica Models on Data-Parallel Multi-Core Architectures, 2015, ISBN 978-91-7519-068-6.

No 1680 **Adrian Lifa**: Hardware/Software Codesign of Embedded Systems with Reconfigurable and Heterogeneous Platforms, 2015, ISBN 978-91-7519-040-2.

No 1685 **Bogdan Tanasa**: Timing Analysis of Distributed Embedded Systems with Stochastic Workload and Reliability Constraints, 2015, ISBN 978-91-7519-022-8.

No 1691 **Håkan Warnquist**: Troubleshooting Trucks – Automated Planning and Diagnosis, 2015, ISBN 978-91-7685-993-3.

No 1702 **Nima Aghaee**: Thermal Issues in Testing of Advanced Systems on Chip, 2015, ISBN 978-91-7685-949-0.

No 1715 **Maria Vasilevskaya**: Security in Embedded Systems: A Model-Based Approach with Risk Metrics, 2015, ISBN 978-91-7685-917-9.

No 1729 **Ke Jiang**: Security-Driven Design of Real-Time Embedded System, 2016, ISBN 978-91-7685-884-4.

No 1733 **Victor Lagerkvist**: Strong Partial Clones and the Complexity of Constraint Satisfaction Problems: Limitations and Applications, 2016, ISBN 978-91-7685-856-1.

No 1734 **Chandan Roy**: An Informed System Development Approach to Tropical Cyclone Track and Intensity Forecasting, 2016, ISBN 978-91-7685-854-7.

No 1746 **Amir Aminifar**: Analysis, Design, and Optimization of Embedded Control Systems, 2016, ISBN 978-91-7685-826-4.

No 1747 **Ekhiotz Vergara**: Energy Modelling and Fairness for Efficient Mobile Communication, 2016, ISBN 978-91-7685-822-6.

No 1748 **Dag Sonntag**: Chain Graphs – Interpretations, Expressiveness and Learning Algorithms, 2016, ISBN 978-91-7685-818-9.

No 1768 **Anna Vapen**: Web Authentication using Third-Parties in Untrusted Environments, 2016, ISBN 978-91-7685-753-3.

No 1778 **Magnus Jandinger**: On a Need to Know Basis: A Conceptual and Methodological Framework for Modelling and Analysis of Information Demand in an Enterprise Context, 2016, ISBN 978-91-7685-713-7.

No 1798 **Rahul Hiran**: Collaborative Network Security: Targeting Wide-area Routing and Edge-network Attacks, 2016, ISBN 978-91-7685-662-8.

No 1813 **Nicolas Melot**: Algorithms and Framework for Energy Efficient Parallel Stream Computing on Many-Core Architectures, 2016, ISBN 978-91-7685-623-9.

No 1823 **Amy Rankin**: Making Sense of Adaptations: Resilience in High-Risk Work, 2017, ISBN 978-91-7685-596-6.

No 1831 **Lisa Malmberg**: Building Design Capability in the Public Sector: Expanding the Horizons of Development, 2017, ISBN 978-91-7685-585-0.

No 1851 **Marcus Bendtsen**: Gated Bayesian Networks, 2017, ISBN 978-91-7685-525-6.

No 1852 **Zlatan Dragisic**: Completion of Ontologies and Ontology Networks, 2017, ISBN 978-91-7685-522-5.

No 1854 **Meysam Aghighi**: Computational Complexity of some Optimization Problems in Planning, 2017, ISBN 978-91-7685-519-5.

No 1863 **Simon Ståhlberg**: Methods for Detecting Unsolvable Planning Instances using Variable Projection, 2017, ISBN 978-91-7685-498-3.

No 1879 **Karl Hammar**: Content Ontology Design Patterns: Qualities, Methods, and Tools, 2017, ISBN 978-91-7685-454-9.

No 1887 **Ivan Ukhov**: System-Level Analysis and Design under Uncertainty, 2017, ISBN 978-91-7685-426-6.

No 1891 **Valentina Ivanova**: Fostering User Involvement in Ontology Alignment and Alignment Evaluation, 2017, ISBN 978-91-7685-403-7.

No 1902 **Vengatanathan Krishnamoorthi**: Efficient HTTP-based Adaptive Streaming of Linear and Interactive Videos, 2018, ISBN 978-91-7685-371-9.

No 1903 **Lu Li**: Programming Abstractions and Optimization Techniques for GPU-based Heterogeneous Systems, 2018, ISBN 978-91-7685-370-2.

No 1913 **Jonas Rybing**: Studying Simulations with Distributed Cognition, 2018, ISBN 978-91-7685-348-1.

No 1936 **Leif Jonsson**: Machine Learning-Based Bug Handling in Large-Scale Software Development, 2018, ISBN 978-91-7685-306-1.

No 1964 **Arian Maghazeh**: System-Level Design of GPU-Based Embedded Systems, 2018, ISBN 978-91-7685-175-3.

No 1967 **Mahder Gebremedhin**: Automatic and Explicit Parallelization Approaches for Equation Based Mathematical Modeling and Simulation, 2019, ISBN 978-91-7685-163-0.

No 1984 **Anders Andersson**: Distributed Moving Base Driving Simulators – Technology, Performance, and Requirements, 2019, ISBN 978-91-7685-090-9.

No 1993 **Ulf Kargén**: Scalable Dynamic Analysis of Binary Code, 2019, ISBN 978-91-7685-049-7.

No 2001 **Tim Overkamp**: How Service Ideas Are Implemented: Ways of Framing and Addressing Service Transformation, 2019, ISBN 978-91-7685-025-1.

No 2006 **Daniel de Leng**: Robust Stream Reasoning Under Uncertainty, 2019, ISBN 978-91-7685-013-8.

No 2048 **Biman Roy**: Applications of Partial Polymorphisms in (Fine-Grained) Complexity of Constraint Satisfaction Problems, 2020, ISBN 978-91-7929-898-2.

No 2051 **Olov Andersson**: Learning to Make Safe Real-Time Decisions Under Uncertainty for Autonomous Robots, 2020, ISBN 978-91-7929-889-0.

No 2065 **Vanessa Rodrigues**: Designing for Resilience: Navigating Change in Service Systems, 2020, ISBN 978-91-7929-867-8.

No 2082 **Robin Kurtz**: Contributions to Semantic Dependency Parsing: Search, Learning, and Application, 2020, ISBN 978-91-7929-822-7.

No 2108 **Shanai Ardi**: Vulnerability and Risk Analysis Methods and Application in Large Scale Development of Secure Systems, 2021, ISBN 978-91-7929-744-2.

No 2125 **Zeinab Ganjei**: Parameterized Verification of Synchronized Concurrent Programs, 2021, ISBN 978-91-7929-697-1.

No 2153 **Robin Keskisärkkä**: Complex Event Processing under Uncertainty in RDF Stream Processing, 2021, ISBN 978-91-7929-621-6.

No 2168 **Rouhollah Mahfouzi**: Security-Aware Design of Cyber-Physical Systems for Control Applications, 2021, ISBN 978-91-7929-021-4.

No 2205 **August Ernstsson**: Pattern-based Programming Abstractions for Heterogeneous Parallel Computing, 2022, ISBN 978-91-7929-195-2.

No 2219 **Evelina Rennes**: Automatic Adaption of Swedish Text for Increased Inclusion, 2022, ISBN 978-91-7929-269-0.

No 2220 **Yuanbin Zhou**: Synthesis of Safety-Critical Real-Time Systems, 2022, ISBN 978-91-7929-271-3.

Linköping Studies in Arts and Sciences

No 504 **Ing-Marie Jonsson:** Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour, 2009, ISBN 978-91-7393-478-7.

No 586 **Fabian Segelström:** Stakeholder Engagement for Service Design: How service designers identify and communicate insights, 2013, ISBN 978-91-7519-554-4.

No 618 **Johan Blomkvist:** Representing Future Situations of Service: Prototyping in Service Design, 2014, ISBN 978-91-7519-343-4.

No 620 **Marcus Mast:** Human-Robot Interaction for Semi-Autonomous Assistive Robots, 2014, ISBN 978-91-7519-319-9.

No 677 **Peter Berggren:** Assessing Shared Strategic Understanding, 2016, ISBN 978-91-7685-786-1.

No 695 **Mattias Forsblad:** Distributed cognition in home environments: The prospective memory and cognitive practices of older adults, 2016, ISBN 978-91-7685-686-4.

No 787 **Sara Nygårdhs:** Adaptive behaviour in traffic: An individual road user perspective, 2020, ISBN 978-91-7929-857-9.

No 811 **Sam Thellman:** Social Robots as Intentional Agents, 2021, ISBN 978-91-7929-008-5.

Linköping Studies in Statistics

No 9 **Davood Shahsavani:** Computer Experiments Designed to Explore and Approximate Complex Deterministic Models, 2008, ISBN 978-91-7393-976-8.

No 10 **Karl Wahlin:** Roadmap for Trend Detection and Assessment of Data Quality, 2008, ISBN 978-91-7393-792-4.

No 11 **Oleg Sysoev:** Monotonic regression for large multivariate datasets, 2010, ISBN 978-91-7393-412-1.

No 13 **Agné Burauskaite-Harju:** Characterizing Temporal Change and Inter-Site Correlations in Daily and Sub-daily Precipitation Extremes, 2011, ISBN 978-91-7393-110-6.

No 14 **Måns Magnusson:** Scalable and Efficient Probabilistic Topic Model Inference for Textual Data, 2018, ISBN 978-91-7685-288-0.

No 15 **Per Sidén:** Scalable Bayesian spatial analysis with Gaussian Markov random fields, 2020, 978-91-7929-818-0.

Linköping Studies in Information Science

No 1 **Karin Axelsson:** Metodisk systemstrukturering- att skapa samstämmighet mellan informationssystem-arkitektur och verksamhet, 1998. ISBN 9172-19-296-8.

No 2 **Stefan Cronholm:** Metodverktyg och användbarhet - en studie av datorstödd metodbaserad systemutveckling, 1998, ISBN 9172-19-299-2.

No 3 **Anders Avdic:** Användare och utvecklare - om anveckling med kalkylprogram, 1999. ISBN 91-7219-606-8.

No 4 **Owen Eriksson:** Kommunikationskvalitet hos informationssystem och affärsprocesser, 2000, ISBN 91-7219-811-7.

No 5 **Mikael Lind:** Från system till process - kriterier för processbestämning vid verksamhetsanalys, 2001, ISBN 91-7373-067-X.

No 6 **Ulf Melin:** Koordination och informationssystem i företag och nätverk, 2002, ISBN 91-7373-278-8.

No 7 **Pär J. Ågerfalk:** Information Systems Actability - Understanding Information Technology as a Tool for Business Action and Communication, 2003, ISBN 91-7373-628-7.

No 8 **Ulf Seigerroth:** Att förstå och förändra systemutvecklingsverksamheter - en taxonomi för metautveckling, 2003, ISBN 91-7373-736-4.

No 9 **Karin Hedström:** Spår av datoriseringens värden – Effekter av IT i äldreomsorg, 2004, ISBN 91-7373-963-4.

No 10 **Ewa Braf:** Knowledge Demanded for Action - Studies on Knowledge Mediation in Organisations, 2004, ISBN 91-85295-47-7.

No 11 **Fredrik Karlsson:** Method Configuration method and computerized tool support, 2005, ISBN 91-85297-48-8.

No 12 **Malin Nordström:** Styrbar systemförvaltning - Att organisera systemförvaltningsverksamhet med hjälp av effektiva förvaltningsobjekt, 2005, ISBN 91-85297-60-7.

No 13 **Stefan Holgersson:** Yrke: POLIS - Yrkeskunskap, motivation, IT-system och andra förutsättningar för polisarbete, 2005, ISBN 91-85299-43-X.

No 14 **Benneth Christiansson, Marie-Therese Christiansson:** Mötet mellan process och komponent - mot ett ramverk för en verksamhetsnära kravspecifikation vid anskaffning av komponent-baserade informationssystem, 2006, ISBN 91-85643-22-X.

FACULTY OF SCIENCE AND ENGINEERING

Linköping Studies in Science and Technology, Dissertation No. 2219
Department of Computer and Information Science

Linköping University
SE-581 83 Linköping, Sweden

www.liu.se

LINKÖPING UNIVERSITY