Proceedings of the 2nd READI Workshop @ LREC2022, pages 31–38
Marseille, 24 June 2022

The Swedish Simplification Toolkit: Designed with Target Audiences in Mind

Evelina Rennes, Marina Santini, Arne Jönsson
Linköping University, RISE, Linköping University
Linköping, Sweden, Stockholm, Sweden, Linköping, Sweden
evelina.rennes@liu.se, marina.santini@ri.se, arne.jonsson@liu.se

Abstract

In this paper, we present the current version of *The Swedish Simplification Toolkit*. The toolkit includes computational and empirical tools that have been developed along the years to explore a still neglected area of NLP, namely the simplification of "standard" texts to meet the needs of target audiences. Target audiences, such as people affected by dyslexia, aphasia, autism, but also children and second language learners, require different types of text simplification and adaptation. For example, while individuals with aphasia have difficulties in reading compounds (such as *arbetsmarknadsdepartement*, eng. ministry of employment), second language learners struggle with cultural-specific vocabulary (e.g. *konflikträdd*, eng. afraid of conflicts). The toolkit allows user to selectively select the types of simplification that meet the specific needs of the target audience they belong to. The Swedish Simplification Toolkit is one of the first attempts to overcome the one-fits-all approach that is still dominant in Automatic Text Simplification, and proposes a set of computational methods that, used individually or in combination, may help individuals reduce reading (and writing) difficulties.

Keywords: automatic text simplification, easy-to-read, automatic text adaptation

1. Introduction

Poor readers come in many forms and include those affected by cognitive disabilities (e.g. individuals with dyslexia), but also those who have not yet developed the skills to master the language (e.g. children and second language learners). Poor readers from these different target groups have more or less widespread cognitive and language difficulties which selectively impair different aspects of reading comprehension. To meet the demands of accessible text of poor readers, a number of initiatives have attempted to adapt texts and make them more comprehensible. Examples of such initiatives in Sweden are the recommendations issued by The Swedish Agency for Accessible Media (Swedish: Myndigheten för tillgängliga medier, (MTM, 2021) MTM and the initiative "Comprehensible text" (Swedish: Begriplig text) (Begriplig Text, 2019). Internationally, the most influential set of guidelines on Easy Language is Plain text (PLAIN, 2011). Such recommendations and guidelines commonly have a one-size-fits-all approach, which means that they have a generalist approach and sometimes overlook the factors underlying different types of reading difficulties. It has been already been pointed out that recommendations and guidelines are often based on common sense assumptions rather than empirical testing on groups of poor readers (Wengelin, 2015).

Although the one-size-fits-all approach is an important first step, individuals who struggle with reading have deficits in cognitive and language skills which make their reading process qualitatively different. Therefore, it is essential to consider the different target audiences when developing recommendations and guidelines, but also when implementing Automatic Text Simplification (ATS) systems. The readers' needs cannot, and should not, be taken out of the equation, rather their needs should be the cornerstone of ATS.

The rationale of the *Swedish Simplification Toolkit* is then to start addressing the one-fits-all bias that still exists in Automatic Text Simplification by putting the linguistic needs of the target audience in the limelight. The set of computational methods underpinning the toolkit address some of these needs and at the same time reflects the current state-of-the-art in ATS for the Swedish language.

The toolkit is the concrete answer to the two research questions that drive our work, namely:

1. What types of linguistic simplification are needed, and which ones are implementable for the Swedish language?
2. Can ATS be conceived, designed and implemented to meet the needs of different target audiences?

In the next sections, we provide the background and illustrate our approach that we have implemented for the Swedish language.

2. Profiling Selected Target Audiences

In the short description of the target audiences provided below, we single out different linguistic phenomena that can be used to characterise audience-specific simplification needs. Simplification may be needed at lexical level (e.g. for individuals with dyslexia), at syntactic levels (e.g. for the individuals affected by aphasia) or at discourse levels (e.g. for people with Autism Spectrum Disorder). Table 1 summarizes the target audiences and their main simplification needs.

Dyslexia. In the International Statistical Classification of Diseases and Related Health Problems - Eleventh

Table 1: Target audiences and their simplification needs

	Lexical Simpification	**Syntactic Simplification**	**Discourse-level Simplification**
Dyslexia	Long words Less frequent words Homophones Words that are orthographically similar New words Non-words		
Aphasia	Information density Noun compounds	Long sentences Long sequences of adjectives Passive voice Object relative clauses Comparison of word meaning	
Intellectual Disability (ID)	Limited vocabulary		
Deaf and Hard-of-Hearing	Limited vocabulary	Complex sentences Morphology Syntax	
Austism Spectrum Disorder	Words related to emotions		Figurative language Texts that require little social knowledge
Second Language Learners	Limited vocabulary		Tight text structure
Children	New words Limited vocabulary		

Revision (ICD-11), developmental dyslexia is categorised under F81.0 Specific reading disorder. An article discussing the major findings of the research on dyslexia during the last decades (Vellutino et al., 2004) showed that word identification inadequacies were the most basic cause of reading difficulties. Individuals with dyslexia experience a wide range of difficulties, such as problems with long words and less frequent words (Hyönä and Olson, 1995; Rello et al., 2013). Except for long and unfamiliar words, other issues that individuals with dyslexia may encounter have been listed, such as homophones, words that are orthographically similar, new words, and nonwords (Rello et al., 2013).

Aphasia is a language impairment caused by brain damage acquired by for example stroke, trauma to the head, neuro-degenerative diseases or brain surgery. Common difficulties experience by individuals with aphasia include high information density, long sentences, long sequences of adjectives, passive voice and noun compounds (Carroll et al., 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).

Intellectual disability (ID) is characterised by low IQ and limitations in many cognitive abilities, such as working memory and executive functions (Danielsson et al., 2012). Individuals with ID have a delay in reading as compared to typical readers which is manifested in capabilities concerning decoding and reading comprehension (Nilsson et al., 2021b; Nilsson et al., 2021a). Using simple texts in order to enhance reading skills is a common strategy in education targeting individuals with ID. Due to the limited amount of textual resources, the teachers face a challenge when choosing accurate educational material, and they often adapt the texts themselves, for example by the use of readability metrics and metrics for reading level estimation, or by writing completely new texts.

Deaf and Hard-of-Hearing. One group of people that may struggle with reading is the deaf or hard-of-hearing. It is established that childhood hearing loss deeply affects language development, and the language deficits may also affect other cognitive developmental areas related to language negatively, such as the development of literacy (Lederberg et al., 2013). Children that are deaf and hard-of-hearing especially struggle with grammar (ibid.), most prominently syntactically complex sentences (Siddharthan, 2003) and grammatical morphology, as well as a limited vocabulary (Fabbretti et al., 1998).

Autism Spectrum Disorder (ASD). With respect to reading comprehension skills, the ASD audience is diverse. The reading difficulties are less straightforward to describe than those exhibited in most of the other

target audiences, due to the large variety of symptoms included in the diagnosis. Difficulties 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 et al., 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 depend on text type (Brown et al., 2013). Generally, individuals with ASD perform better when reading texts that require little social knowledge. However, they also highlighted the fact that ASD covers a variety of symptoms and deficits, and that the diagnosis in itself does not imply any reading comprehension difficulties.

Second Language Learners. This audience differs from many of the other groups, since learners of a new language do not necessarily have any impairment that hinders reading or understanding, but may 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. Knowing a language's vocabulary has proved to be an important factor for learning a new language. Knowing a language's vocabulary has proved to be an important factor for learning a new language.

Children. Although not having any physical or cognitive disability, a possible target audience for text adaptation or simplification is children. As the Internet is becoming the dominating source of textual information, there is a growing need for text adapted to the reading level of children of different ages. 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).

3. Simplification techniques

In this section we briefly present techniques that address the different levels of simplification.

Lexical simplification refers to the automatic simplification at word level. It aims at identifying and replacing complex words (or phrases) by an easier-to-read alternative. Regardless of how we define simple words, the substitution of words that are more complex to simpler words with the same meaning is a rather well-studied area in ATS (e.g. (Paetzold and Specia, 2017) for an overview), and although it is a challenging task with many non-trivial subtasks (identifying complex words, disambiguating word senses, etc.), the guidelines of substituting complex words and compounds can be considered possible to automate. The avoidance of jargon and technical terms can be solved with specialised term lists, such as the black list (Stadsrådsberedningen) used by the Swedish public authorities. To not split words on two lines, and how to write (or not write) numerical expressions are other guidelines that are relatively easy to automate. Abbreviations should be avoided and this is also a task that can be automated.

Syntactic simplification refers to the automatic simplification at text level. It aims at simplifying the text by restructuring the words of the sentences, and/or rewriting it into smaller sentences. The issue of keeping the text brief can be addressed through different kinds of ATS. 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, one must be aware of the risk that relevant text information might be deleted in the process, which could cause confusion or impair the experienced reading flow of the reader. One guideline is to keep one proposition per sentence. To follow this guideline is slightly more complicated, as it requires some semantic parsing. One possible solution to this could be event-based simplification (see for example (Štajner et al., 2016)), that identifies mentions of factual events and delete sentences or parts of sentences that are irrelevant to these event mentions. Such simplification approaches could enhance text comprehension by deleting irrelevant information and highlighting the main information, but will naturally also 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 more typographic reasons, i.e. that features of the text layout makes the text less appealing to read. One guideline suggests to mix long and short sentences. This could be considered as a parameter when applying guidelines that intend to write as brief as possible. Guidelines that change negative statements to positive statements are possible to automate, but require a mechanism for identifying such structures, as well as a set of rewriting suggestions. For relatively simple cases (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 et al., 2015; Mukherjee et al., 2017). It is generally recommended to use personal pronouns, and to address the reader directly. Such linguistic simplification strategies have previously been, at least partially, implemented in a rule-based simplifi-

cation system (Rennes and Jönsson, 2015). In the same system, a rule for reordering sentences so that they keep a straight word order, with subject, verb and object kept close to each other, was implemented.

Discourse-level simplification. Syntactic skills facilitate access of meaning from grammatical structures, which is a fundamental process in gaining text meaning at any level of reading comprehension. Discourse skills allow readers to understand the cohesive interlinks within and between sentences and is important for a macro level of passage understanding. The macro level simplification could be easily implemented in an ATS tool with existing techniques. For example, to make the main information easy to find could include the automatic extraction of keywords and present them in clear ways (boldface, headlines, bullet lists, etc.), as well as providing an automatic summary of the text (Hahn and Mani, 2000). Keyword extraction and extractive summarisation (to extract the most important sentences of a text) are techniques that could be relatively easily implemented, whereas abstractive summarisation (to rewrite the summary from scratch) require more sophisticated methods and data for training (Monsen and Jönsson, 2021). The guideline to let the general and the most important information be presented in the beginning could possibly be approached using the same keyword extraction and summarisation techniques.

4. The Swedish Simplification Toolkit

The Swedish Simplification Toolkit is a modular text simplification system covering a wide range of simplification and summarization functions. It is important to point out that in this context summarization is used as simplification tools as explained in Smith and Jönsson (2011a).

The Swedish Simplification Toolkit can be used via two interfaces that offer complementary functions, namely FRIENDLYREADER (more targeted to reading) and TECST (more targeted to writing), see Figure 1.

Both FRIENDLYREADER and TECST leverage on four modules: STILLETT (Rennes and Jönsson, 2015), TEXTCOMP (Falkenjack, 2018; Falkenjack et al., 2013), COGSUM (Smith and Jönsson, 2011a; Smith and Jönsson, 2011b) and JULIUSUM (Monsen and Jönsson, 2021). SAPIS (Fahlborg and Rennes, 2016; Falkenjack et al., 2017) is a REST API aiming to make the services readily available and is used by both FRIENDLYREADER and TECST. All the core modules have been evaluated according to the criteria of intrinsic evaluation and all results can be found in the references given below.

In the rest of the section, first we describe the core modules and than the user interfaces through which the core modules are deployed.

STILLETT (Rennes and Jönsson, 2015) is a rule-based automatic text simplification tool for Swedish. It started off as a Java application, partly built on COGFLUX (Rybing et al., 2010), with a dynamic structure of processes and modules, where each process runs a number of modules. 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 originally combined the word pairs from the Synlex lexicon (Kann, 2004) and frequency information. The Synlex lexicon includes 82,000 word pairs including an annotation of level of synonymity. This score was calculated by 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 was based on a corpus of texts in simple Swedish, and the other method was based on theories from the field of cognitive linguistics where hypernyms with characteristics of basic-level words were found to be useful for the task of lexical simplification (Rennes and Jönsson, 2021).

STILLETT has undergone several improvements (see for instance Johansson and Rennes (2016)) since the first implementation and is today built on Python3, and uses dep_tregex 1(Dvorkovich et al., 2016), for reordering the dependency trees using rules inspired by Tregex (Levy and Andrew, 2006). 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 preprocessor is accessed through the REST API SAPIS and runs the Swedish pipeline with EFSELAB (Östling, 2018) and MaltParser (Nivre et al., 2007) version 1.9.0.

TEXTCOMP is a collection of text complexity measures. The main part of the included measures consists of the SCREAM (Swedish Compound REAdability Metric) features (Sjöholm, 2012). SCREAM features include surface, lexical and structural features (the complete list of SCREAM features can be found in Falkenjack (2018) and Falkenjack et al. (2013)). More recently, cohesion-related measures have been included, namely the Coh-Metrix measurements, translated to Swedish from the orginal English version (Graesser et al., 2004).

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, 2011; Hassel, 2007) 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 (Monsen and Jönsson, 2021) is an auto-

matic abstractive summariser. An abstractive summarisation differs from an extractive summarisation in that the words and sentences are not directly extracted from the text, but instead generated based on a pre-trained model. This means that abstractive summaries can contain completely novel words and sentences not present in the source text, while maintaining the key content of the text. JULIUSUM was trained utilising the methodology proposed by Rothe et al. (2020), using a pretrained Swedish BERT model (Malmsten et al., 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.

FRIENDLYREADER, Figure 1 left, is a customizable interface that can be adjusted to the specific needs of different target audiences. The idea is that the interface should contain the entire palette of simplification techniques, including both linguistic and layout simplifications, and that the user can adapt the text completely to their individual needs. FRIENDLYREADER is under constant development. In addition to the modules for simplification and summarization, FRIENDLYREADER also contains text-to-speech functionality, which lets the reader listen to the text. The simplification related to text layout is the possibility to change font size, line spacing, font and line length. In its current state, the user pastes the text into a large text field and presses Run. The view in the left of Figure 1 is then presented to the user. The layout consists of three parts. The main field is the middle field, where the the text is presented to the reader. The left-hand side contains a menu with various types of text simplification. 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. 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 simplification options related to text visualization.

TECST (Text Complexity and Simplification Toolkit), Figure 1 right, is a a 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 simplification 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 simplification process quicker and cheaper, without overlooking the expertise and unique competence provided by the human writer.

The TECST layout, presented in the right of Figure 1, consists of two fields: the editor, which makes up the main part of the tool layout, and the simplification 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 simplification and visualisation field, on the right-hand side, presents information regarding the current complexity and simplification suggestions of the text. It has three tabs: visualisation, text information and text simplification. In the visualisaton tab, a text complexity visualisation in the form of a radar chart is presented. 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.

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.

5. Discussion

The Swedish Simplification Toolkit meets the needs of the target audiences described in Section 2. There is a close match between the characterization of target audience presented in Table 1 and the options provided in the user interfaces. For example, in FRIENDLYREADER, people affected by ID, as well as non-native speakers and children, can click the button "Visa Synonymer" (Show synonyms) to fill up the gap of their limited vocabulary. People affected by aphasia have the possibility to convert passive voice into active voice by checking the box "passiv till aktiv form" (pas-

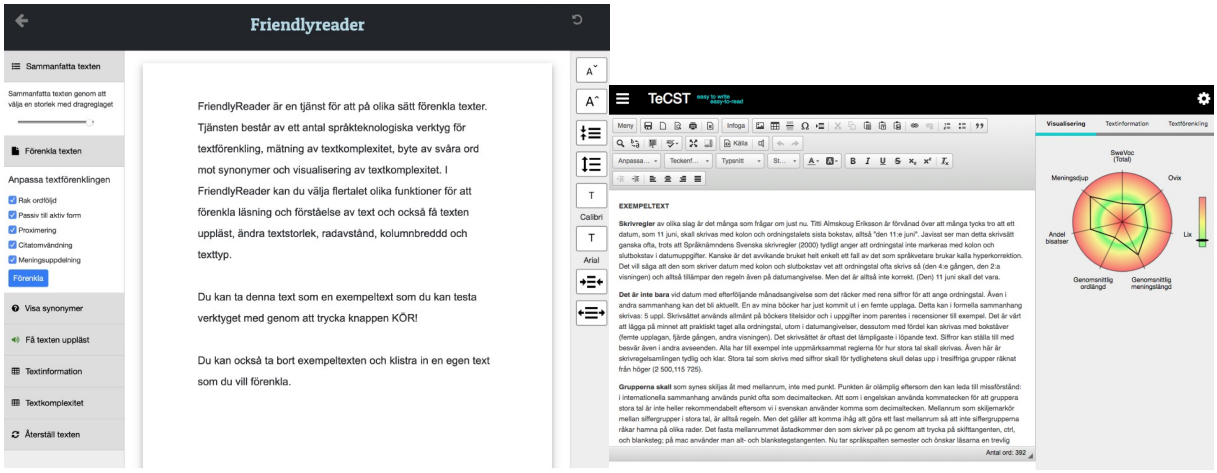

Figure 1: FRIENDLYREADER, left and TECST, right

sive to active form). The aphasic can also shorten long sentences that are difficult for the to process by using the summarization functions. In TECST, it is possible to visualize the spikes of the complexity of a text via a radar chart and then decide what types of simplifications needed. For instance, people with dyslexia might decide to choose shorter synonyms of long words, but leave the syntactic complexity untouched since this type of complexity does not hinder their reading comprehension.

At the time of this publication, no options have been implemented to convert figurative language into demetaphorized language. Also, no functionality has been created yet to compensate for the social knowledge that autistic people might miss from a standard text.

As mentioned above, the core modules have been evaluated, but the usability of the toolkit has not yet been tested on target audiences. Such evaluations include two activities, to assess the interface and interaction with the various simplification tools, and assessment of the various simplification techniques and text complexity measures.

The latter is currently done in three studies, one involving students with dyslexia, one involving students having an intellectual disability and one with teachers for students with reading difficulties. The first two studies use texts that are adapted using the toolkit. The students read them on paper and assess the usability from various perspectives. The study with teachers investigate the use of text complexity measures and visualisation of text complexity and is conducted in focus groups where the teachers are presented a variety of text complexity measures and visualisations. The reason for using paper and not the interface is, of course, that we want to focus on the usability of the techniques for text simplification and complexity measure, not the usability of the interface.

The answers to the research questions are:

1. What types of linguistic simplification are needed, and which ones are implementable for the Swedish language?

   Answer: lexical, syntactic and discourse level simplification are needed for the target audiences that we have explored in this paper. However for the Swedish language, many areas are unexplored, as seen from the empty cells in Table 1, especially for discourse-level simplification.

2. Can ATS be conceived, designed and implemented to meet the needs of different target audiences?

   Answer: Absolutely yes. We have presented an approach (core modules + interfaces) that shows how target audiences can adapt standard text to their needs.

6. Conclusion and Future Work

In this paper we presented the Swedish Simplification Toolkit, conceived and designed with the target audiences in mind. The toolkit is the outcome of years of theoretical and computational research. However, much remains to be done. Our next step will be the validation of the adaptations on readers from the target audiences. We are currently testing the effects of specific simplification operations on individuals with dyslexia and intellectual disabilities, and the results of this study will provide a starting point for further development of more customized text simplification. Future work includes the validation of the usability of the interfaces directly by target audiences.

7. Acknowledgements

This work has been funded by The Swedish Research Council (VR) and Sweden's innovation agency (VINOVA).

8. References

Begriplig Text. (2019). *19 råd för att skriva begripligt*. Dyslexiförbundet.

Brown, H. M., Oram-Cardy, J., and Johnson, A. (2013). A meta-analysis of the reading comprehension skills of individuals on the autism spectrum. *Journal of Autism and Developmental Disorders*, 43(4):932–955.

Burgers, C., Beukeboom, C. J., Sparks, L., and Diepeveen, V. (2015). How (not) to inform patients about drug use: use and effects of negations in dutch patient information leaflets. *Pharmacoepidemiology and drug safety*, 24(2):137–143.

Carroll, J., Minnen, G., Pearce, D., Canning, Y., Devlin, S., and Tait, J. (1999). Simplifying Text for Language-Impaired Readers. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 269–270.

Chatterjee, N. and Mohan, S. (2007). Extraction-Based Single-Document Summarization Using Random Indexing. In *Proceedings of the 19th IEEE international Conference on Tools with Artificial intelligence – (ICTAI 2007)*, pages 448–455.

Danielsson, H., Henry, L., Messer, D., and Rönnberg, J. (2012). Strengths and weaknesses in executive functioning in children with intellectual disability. *Research in developmental disabilities*, 33(2):600–607.

De Belder, J. and Moens, M.-F. (2010). Text simplification for children. In *Proceedings of the SIGIR workshop on accessible search systems*, pages 19–26. ACM; New York.

Decker, A. (2003). Towards automatic grammatical simplification of Swedish text. Master's thesis, Stockholm University.

Dvorkovich, A., Gubanov, S., and Galinskaya, I. (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*, volume 2, pages 281–288, Berlin, Germany.

Fabbretti, D., Volterra, V., and Pontecorvo, C. (1998). Written language abilities in deaf italians. *The Journal of Deaf Studies and Deaf Education*, 3(3):231–244.

Fahlborg, D. and Rennes, E. (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*.

Falkenjack, J., Heimann Mühlenbock, K., and Jönsson, A. (2013). Features Indicating Readability in Swedish Text. In *Proceedings of the 19th Nordic Conference of Computational Linguistics (NoDaLiDa-2013), Oslo, Norway*, number 085 in NEALT Proceedings Series 16, pages 27–40. Linköping University Electronic Press.

Falkenjack, J., Rennes, E., Fahlborg, D., Johansson, V., and Jönsson, A. (2017). Services for text simplification and analysis. In *Proceedings of the 21st Nordic Conference on Computational Linguistics, Gothemburg, Sweden*.

Falkenjack, J. (2018). *Towards a model of general text complexity for Swedish*. Licentiate thesis, Linköping University Electronic Press.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., and Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, 36(2):193–202.

Hahn, U. and Mani, I. (2000). The Challenges of Automatic Summarization. *Computer*, 33(11):29–36.

Hassel, M. (2007). *Resource Lean and Portable Automatic Text Summarization*. Ph.D. thesis, ISRN-KTH/CSC/A–07/09-SE, KTH, Sweden.

Hassel, M. (2011). Java Random Indexing toolkit, 1. `http://www.csc.kth.se/~xmartin/java/`.

Hillis, A. E. (2007). Aphasia: Progress in the last quarter of a century. *Neurology*, 69:200–213.

Hyönä, J. and Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6):1430.

Johansson, V. and Rennes, E. (2016). Automatic extraction of synonyms from an easy-to-read corpus. In *Proceedings of the Sixth Swedish Language Technology Conference (SLTC-16), Umeå, Sweden*.

Kalandadze, T., Norbury, C., Nærland, T., and Næss, K.-A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*, 22(2):99–117.

Kann, V. (2004). Folkets användning av Lexin – en resurs. In *Lexinkonferens 2004*, Stockholm, Sweden.

Lederberg, A. R., Schick, B., and Spencer, P. E. (2013). Language and literacy development of deaf and hard-of-hearing children: successes and challenges. *Developmental psychology*, 49(1):15.

Levy, R. and Andrew, G. (2006). Tregex and Tsurgeon: tools for querying and manipulating tree data structures. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*.

Malmsten, M., Börjeson, L., and Haffenden, C. (2020). Playing with words at the national library of Sweden–making a Swedish BERT. *arXiv preprint arXiv:2007.01658*.

Monsen, J. and Jönsson, A. (2021). A method for building non-english corpora for abstractive text summarization. In *Proceedings of CLARIN Annual Conference 2021*.

MTM. (2021). Att skriva lättläst. `https://www.mtm.se/var-verksamhet/lattlast/att-skriva-lattlast/`. Accessed: 2021-10-05.

Mukherjee, P., Leroy, G., Kauchak, D., Rajanarayanan, S., Diaz, D. Y. R., Yuan, N. P., Pritchard, T. G., and

Colina, S. (2017). Negait: A new parser for medical text simplification using morphological, sentential and double negation. *Journal of biomedical informatics*, 69:55–62.

Nilsson, K., Danielsson, H., Elwér, Å., Messer, D., Henry, L., and Samuelsson, S. (2021a). Decoding Abilities in Adolescents with Intellectual Disabilities: The Contribution of Cognition, Language, and Home Literacy. *Journal of Cognition*, 4(1):1–16.

Nilsson, K., Danielsson, H., Elwér, Å., Messer, D., Henry, L., and Samuelsson, S. (2021b). Investigating reading comprehension in adolescents with intellectual disabilites: Evaluating the simple view of reading. *Journal of Cognition*, 4(1):1–16.

Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S., and Marsi, E. (2007). MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Paetzold, G. H. and Specia, L. (2017). A survey on lexical simplification. *Journal of Artificial Intelligence Research*, 60:549–593.

PLAIN. (2011). Federal Plain Language Guidelines.

Rello, L., Baeza-Yates, R., Dempere-Marco, L., and Saggion, H. (2013). Frequent words improve readability and short words improve understandability for people with dyslexia. In *IFIP Conference on Human-Computer Interaction*, pages 203–219. Springer.

Rennes, E. and Jönsson, A. (2015). A tool for automatic simplification of swedish texts. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa-2015), Vilnius, Lithuania*.

Rennes, E. and Jönsson, A. (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)*, pages 259–267.

Rothe, S., Narayan, S., and Severyn, A. (2020). Leveraging pre-trained checkpoints for sequence generation tasks.

Rybing, J., Smith, C., and Silvervarg, A. (2010). Towards a Rule Based System for Automatic Simplification of Texts. In *Swedish Language Technology Conference, SLTC, Linköping, Sweden*.

Siddharthan, A. (2003). *Syntactic Simplification and Text Cohesion*. Ph.d. thesis, University of Cambridge, UK.

Sjöholm, J. (2012). Probability as readability: A new machine learning approach to readability assessment for written Swedish. Master's thesis, Linköping University.

Smith, C. and Jönsson, A. (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*.

Smith, C. and Jönsson, A. (2011b). Enhancing extraction based summarization with outside word space. In *Proceedings of the 5th International Joint Conference on Natural Language Processing, Chiang Mai, Thailand*.

Štajner, S., Popovic, M., and Béchara, H. (2016). Quality estimation for text simplification. In *Proceedings of the LREC Workshop on Quality Assessment for Text Simplification (QATS), Portoroz, Slovenia*.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., and Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of child psychology and psychiatry*, 45(1):2–40.

Wengelin, Å. (2015). Mot en evidensbaserad språkvård? en kritisk granskning av några svenska klarspråksråd i ljuset av forskning om läsbarhet och språkbearbetning. *Sakprosa*, 7(2).

Östling, R. (2018). Part of speech tagging: Shallow or deep learning? *North European Journal of Language Technology*, 5:1–15.