

# Automatic Text Simplification for Social Good: Progress and Challenges

Sanja Štajner

ReadableAI

Cologne, Germany

sanja.stajner@readableai.com

## Abstract

Since the late 1990s, automatic text simplification (ATS) was promoted as a natural language processing (NLP) task with great potential to make texts more accessible to people with various reading or cognitive disabilities, and enable their better social inclusion. Large multidisciplinary projects showed promising steps in that direction. Since 2010, the field started attracting more attention but at the cost of major shifts in system architecture, target audience, and evaluation strategies. Somewhere along the way, the focus has shifted from ATS for social good towards building complex end-to-end neural architectures that are not aimed at any particular target population. This study presents the trajectory of ATS for social good, the main issues in current ATS trends, and the ways forward that could bring the field back to its initial goals.

## 1 Rationale

The right to accessible information is a fundamental right that should be granted to all people (UN, 2020). It is the key factor for personal empowerment and social inclusion. Nevertheless, textual information found on the web, in the news, health leaflets, and other sources is often linguistically too complex for many people and thus impedes their active participation in the society.

### 1.1 Adult Literacy

In the OECD Adult Literacy Report (OECD, 2013), “literacy is defined as the ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one’s goals, and to develop one’s knowledge and potential”. The literacy scale comprises of six levels:<sup>1</sup>

<sup>1</sup>We shorten the description of the skills at each literacy level and maintain only the information that is most relevant for further discussions. The full description of the literacy levels can be found on the page 64 in (OECD, 2013).

**Below Level 1:** Able to read brief texts on familiar topics and to locate a single piece of specific information (the process requires only basic vocabulary knowledge).

**Level 1:** Able to read very short texts (sometimes non-continuous or mixed) and locate a single piece of information identical or synonymous to the information given in the question/directive (requires only basic vocabulary knowledge). Texts may contain little, if any, competing information.

**Level 2:** Able to integrate two or more pieces of information, use paraphrasing or low-level inferences, compare and contrast or reason about the information requested.

**Level 3:** Able to understand dense and lengthy texts, various rhetorical structures, perform various levels of inference, and disregard irrelevant or inappropriate content.

**Level 4:** Able to integrate, interpret or synthesise information from complex or lengthy texts, perform complex inferences and apply background knowledge, identify and understand one or more non-central idea(s), evaluate subtle evidence-claim or persuasive discourse relationships, discern between the correct and competing information which sometimes appear in almost equal ratio.

**Level 5:** Able to search for and integrate information across multiple dense texts, synthesise various ideas and points of view, evaluate evidence-based arguments, apply and evaluate logical and conceptual models of ideas, evaluate reliability of the sources, selecting relevant information, notice subtle, rhetorical cues and make high-level inferences or use specialised background knowledge.

The percentage of adults (age 16–65) whose literacy is below Level 2 is 16.7%, on average, across 24 countries taken into account (OECD, 2013).<sup>2</sup> In

<sup>2</sup>Participating countries: Australia, Austria, Belgium (Flanders), Canada, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, the

Literacy	US	Spain	Italy	Average (24 countries)	Simplification need
Below Level 2	21.7%	28.3%	28.3%	16.7%	lexical simplification
Below Level 3	≈50%	67.4%	70.3%	50.0%	syntactic simplification
Below Level 4	≈86%	≈95%	≈97%	89.4%	conceptual simplification

Table 1: Percentage of population in need for text simplification (the numbers are inferred from (OECD, 2013)).

Italy, Spain, and United States, this percentage is even higher (Table 1). On average, 0.7% of population has the literacy level 5. According to the definitions of the literacy levels in the OECD report, outlined above, one can conclude that: (1) all people with literacy below Level 4 have difficulties understanding conceptually complex texts; (2) all people with literacy below Level 3 additionally have difficulties understanding syntactically complex texts; and (3) all people with literacy below level 2 have difficulties understanding texts which are linguistically difficult in any sense (lexically, syntactically, or conceptually). This means that, according to the results of the OECD report, approximately 16.7% of population needs lexical simplification of everyday texts, 50% of population needs syntactic simplification, and 89.4% of population needs conceptual simplification (Table 1).

## 1.2 Manual Text Simplification

Text simplification aims to transform original texts into their simpler variants, which are more understandable to the target reader(s), while preserving the original meaning. Some nuances in meaning would inevitably be lost during that process, but the core information should stay the same.

The first guidelines for how to write more accessible texts for international communication, by using reduced vocabulary and restricted number of grammatical rules, were Basic English (Ogden, 1937) and Plain English initiative (Crystal, 1987). Since the late 1990s, many more initiatives have raised awareness about how to write more accessible texts, e.g. how to write for people with intellectual disabilities (Freyhoff et al., 1998; Mencap, 2002; Karreman et al., 2007), how to make public information more accessible to wider audiences (PlainLanguage, 2011), and how to make web content more accessible (W3C, 2008; Cooper et al.,

Netherlands, Norway, Poland, the Slovak Republic, Spain, Sweden, the United Kingdom (England and Northern Ireland), the United States, Cyprus, and the Russian Federation.

2010). Specialized websites that offer easy-to-read news are now common in many countries, e.g. Noticias fácil in Spain<sup>3</sup>, DR in Denmark<sup>4</sup>, News Web Easy in Japan<sup>5</sup>.

## 2 Automatic Text Simplification (ATS)

Manual text simplification, apart from being slow and expensive (requiring trained human editors), cannot keep up-to-date with the new information published online, offer variety of written content, or adaptation at a personal level. That created the need for automatic, or at least semi-automatic (involving a manual post-editing step) TS systems.

The first automatic text simplification (ATS) systems were rule-based, with focus on syntactic simplification in English, and were envisioned mainly as a pre-processing step for various natural language processing (NLP) applications, e.g. parsing, machine translation, summarization, and information retrieval; the clarity of texts for human readers was only one of the possible use cases (Chandrasekar et al., 1996; Chandrasekar and Srinivas, 1997). Subsequent ATS systems were proposed for various languages and were encompassing any number of the following simplification operations: lexical simplification, syntactic simplification, or explanation generation (Shardlow, 2014; Siddharthan, 2014).<sup>6</sup>

### 2.1 ATS Projects for Social Good

Automatic text simplification gained more interest from research community through several national and international projects that secured public funding for building ATS systems for various vulnerable populations and in various languages (Table 2).

<sup>3</sup>[www.noticiasfacil.es](http://www.noticiasfacil.es)

<sup>4</sup><https://www.dr.dk/ligetil/>

<sup>5</sup><https://www3.nhk.or.jp/news/easy/>

<sup>6</sup>For a comprehensive overview of ATS, please see the book written by Saggion (2017), and the more recent surveys on ATS, lexical simplification, and sentence simplification (Paetzold and Specia, 2017b; Alva-Manchego et al., 2020b; Al-Thanyan and Azmi, 2021).

Project	Period	Language	Impairment	Text types
PSET	1996–2000	EN	Aphasia	News
PorSimples	2007–2010	PT (BR)	Low literacy	News and educative articles
Simplext	2010–2013	ES	Down syndrome	News
FIRST	2012–2015	EN, ES, BG	Autism	News, literature, health leaflets
Able To Include	2014–2017	EN, ES, NL	Intellectual	E-mails, short messages

Table 2: Projects proposing automatic text simplification for social good.

All aimed to provide a better social inclusion, thus promoting ATS as an NLP task for social good.

**ATS for Readers with Aphasia.** PSET (Practical Simplification of English Text) was an UK national project (1996–2000) that aimed to build an ATS system that would make newspaper articles more accessible to people who suffer from aphasia (Carroll et al., 1998). This was the first project that promoted the use of ATS for social good. The solution was envisioned as a modular system consisting of two components: an analyser and a simplifier. The analyser would use lexical tagger, morphological analyser, and parser to recognize potentially difficult words and sentences for aphasic readers. The simplifier would use rule-based syntactic simplification systems to convert passive to active voice, extract embedded clauses, and split conjoined sentences, and a lexical simplification system which would use WordNet (Vossen, 2004) for retrieving substitution candidates and the Oxford Psycholinguistic Database (Quinlan, 1992) to choose the simplest one of them. It was envisioned to evaluate the final system with a group of aphasic readers (Carroll et al., 1998). HAPPI (Helping Aphasic People Process Information) was the continuation of the PSET project, this time with the aim of developing web-based ATS system which would make web content more accessible to aphasic readers by simplifying text vocabulary and syntax (Devlin and Unthank, 2006). To the best of our knowledge, there are no publications about the evaluations of the final systems with the end users.

**ATS for Readers with Low Literacy.** The PorSimples (Simplification of Portuguese Text for Digital Inclusion and Accessibility) project (Aluísio and Gasperin, 2010) was a Brazilian national project (2007–2010). Its aim was to produce tools and resources for people with low literacy levels (through the assistive technology systems called FACILITA and Educational FACILITA) and au-

thors that want to produce texts for this audience (through an authoring system called SIMPLIFICA). The FACILITA system was envisioned with the focus on summarization and syntactic simplification, i.e. sentence splitting, change of discourse markers, conversion from passive to active voice, inversion of clause order, subject-verb-object order, topicalization and de-topicalization (Watanabe et al., 2009). Educational FACILITA was a web application prototype which offered lexical elaborations and named entity labelling. A pilot study with low-literacy users reported improved text comprehension, though participants reported to be sometimes confused with the offered list of synonyms that had multiple meanings (Watanabe et al., 2010).

**ATS for People with Cognitive Disabilities.** Simplext (Saggion et al., 2011, 2015) was a Spanish national project (2010–2013) that proposed several modules for automatic simplification of Spanish (news) texts for people with cognitive disabilities (particularly people with Down’s syndrome). The modules combined rule-based and corpus-based techniques for reducing syntactic complexity of sentences, deleting unnecessary information, performing numerical simplification, normalizing reporting verbs, and reducing lexical complexity (Saggion et al., 2015). The final ATS system was tested by 44 people with Down’s syndrome by measuring reading time and text comprehension. More correct answers were obtained for simplified texts, on average, and participants positively rated the existence of such a tool that is available through different technological channels, e.g. computer, smartphone, tablet (Saggion et al., 2015).

The Able to Include project (2014–2017) was an international project that aimed to improve the quality of life of people with intellectual or developmental disabilities, dementia, or any kind of cognitive impairments.<sup>7</sup> Three applications were built during

<sup>7</sup><http://able-to-include.com>

the project: ABLEsocial (media app), ABLEchat (mobility app), and KOLUMBA (ABLEmail app) by using Simplext (ATS technology), PictoText, and Text2Speech web services.<sup>8</sup> This project did not offer significant advances in core ATS research. Its focus was, instead, on bringing existing ATS solutions to life through freely accessible prototypes.

**ATS for Readers with Autism.** The FIRST project (2012–2015) was an EC-funded project under the FP7 ICT call concerning smart and personalised inclusion (Orăsan et al., 2018). It provided OpenBook, a software that can automatically identify a range of language phenomena (e.g. complex syntactic structures, complex words and phrases, ambiguous words, and metaphors) that are problematic for people with high-functioning (IQ>70) autism spectrum disorders (ASD), and replace some of them. It also offered adding illustrative pictures and concise document summaries. The tool supported three languages: English, Spanish, and Bulgarian. Acknowledging that ATS was creating a large number of errors for some texts, and that end users have low tolerance for ungrammatical and erroneous text, OpenBook was also offering powerful post-editing options so that carers can post-edit the texts before showing them to the end users (Orăsan et al., 2018). The architecture was highly modular, with separate modules for each transformation, to allow for highest level of personalization possible. The final evaluation with 243 participants with high-functioning autism from UK, Spain, and Bulgaria was performed on texts which were automatically simplified and post-edited (by the clinical teams) using OpenBook. The participants were more successful at answering multiple-choice questions about the simplified than the original texts, and human-aided text conversion time significantly decreased by using OpenBook (Orăsan et al., 2018).

## 2.2 ATS in Medical Domain

A special type of literacy is health literacy, defined by the European Health Literacy Consortium:

*“Health literacy is linked to literacy and entails people’s knowledge, motivation and competences to access, understand, appraise and apply health information in*

<sup>8</sup>Details about text to pictograph translation and the accessible e-mail client can be found in (Sevens et al., 2017; Saggion et al., 2017).

*order to make judgements and take decisions in everyday life concerning health care, disease prevention and health promotion to maintain or improve quality of life during the life course.”*

According to the European Health Literacy Survey, nearly half of all adults in the eight European countries covered by the survey were found to have low health literacy skills, which may result in less healthy choices, riskier behaviour, poorer health, less self-management and more hospitalization (WHO, 2013). Due to their potential to improve the health of people with low health literacy skills, attempts at building ATS systems focused on simplifying medical content have recently attracted noticeable attention.

In contrast to the ATS systems mentioned so far, the ATS systems in medical domain are particularly focused on translating highly specialized medical expressions into their layman variants, and thus require domain-specific resources and customized algorithms. Kloehn et al. (2018) proposed an algorithm for automatically generating explanations for difficult (medical) terms in English and Spanish. van den Bercken et al. (2019) explored the possibility of training the earlier proposed general purpose neural text simplification model (Nisioi et al., 2017) on medical parallel corpus. Shardlow and Nawaz (2019) used the general purpose neural text simplification model (Nisioi et al., 2017) augmented with the phrase table of complex-simple medical terminology to automatically simplify clinical letters in English. Cardon and Grabar (2020) used similar approach for biomedical texts in French. Emphasizing the need for high-quality simplification in medical domain, Van et al. (2020) explored the possibility of applying pretrained neural language models to the autocomplete process for sentence-level medical text simplification.

## 2.3 ATS Research Trends

Research on ATS in English<sup>9</sup> could roughly be divided in three phases: rule-based systems that focus on specific, well-defined transformations (until 2010); data-driven supervised machine learning systems trained on parallel TS data (2010–2014/2016); and neural text simplification systems (from 2015/2017 onward). The neural trends

<sup>9</sup>The vast majority of ATS research was published for English and thus the discussion here is focused on those. Most of the ATS research for other languages uses similar ideas and adapts them to the particular language.



started earlier for purely lexical simplification systems by leveraging word embeddings (Glavaš and Štajner, 2015; Paetzold and Specia, 2016). Sentence simplification systems embraced neural architectures in 2017 by using neural machine translation (Nisioi et al., 2017) and reinforcement learning (Zhang and Lapata, 2017). Those three years (2010, 2015, and 2017) can clearly be observed in Figures 1 and 2, which show the number of scholarly articles that mention the specific TS terms anywhere in the text, and the number of scholarly articles with those TS terms in their title.<sup>10</sup>

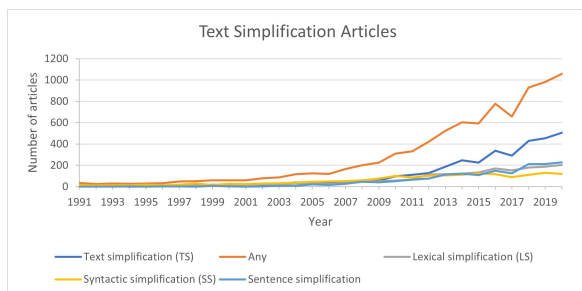


Figure 1: The number of articles in Google Scholar that mention specific text simplification terms.

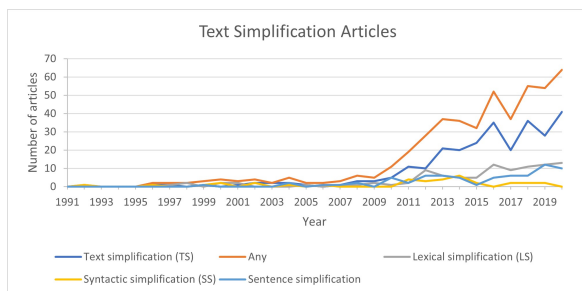


Figure 2: The number of articles in Google Scholar with specific text simplification terms in their title.

Interestingly, the number of articles mentioning syntactic simplification was steadily increasing until 2015, and then again in the period 2017–2019 (Figure 1), while the number of articles with syntactic simplification in their title was increasing until 2014, then noticeably decreased by 2016, to never increase much again (Figure 2). The number of lexical simplification (LS) papers, in turn, has been increasing since 2010, with two peaks, in 2012 and 2016. The peak in 2012 was due to the papers describing systems which participated in the SemEval-2012 English Lexical Simplification task

<sup>10</sup>We excluded the patents and citations from this search, and looked for the exact matches of the TS terms. All returned titles were additionally manually checked.

(Specia et al., 2012), and lexical simplification endeavours in the scope of the Simplext project (Bott et al., 2012; Drndarević and Saggion, 2012a,b). Almost all lexical simplification research published in 2016 was focused on LS for non-native speakers.

So far, conceptual simplification has been tackled only through coreference resolution (Orăsan et al., 2018). Although most works mention “text simplification”, they actually refer to sentence simplification. The state-of-the-art ATS systems published in top tier NLP/CL/AI conferences, e.g. (Nisioi et al., 2017; Zhang and Lapata, 2017; Surya et al., 2019; Kumar et al., 2020; Mallinson et al., 2020), all describe end-to-end systems for sentence simplification, and are not directed towards any particular simplification transformation or target population.

Very few papers proposed ATS methods that operate beyond sentence level (Glavaš and Štajner, 2013; Narayan and Gardent, 2014; Štajner and Glavaš, 2017). The research interests thus do not seem to follow the reported text simplification needs (see Table 1 in Section 1.1).

### 3 Challenges in ATS for Social Good

In spite of the evident increased interest for ATS in the last few years, and the obvious benefits it could bring to the society, no commercial systems have been offered so far, proving thus that automatic text simplification is a very challenging task.

The main reason might lie in the multidisciplinary nature of the task, i.e. the need for combining linguistics, psycholinguistics, computer science, and, to be used in real-world applications, advanced software engineering. Only large multidisciplinary teams can ensure solutions that are technically-advanced and, at the same time, truly help the end users. The absence of such multidisciplinary research leads to usage of non-optimal resources and evaluation procedures which do not involve end users, thus steering the field away from any real-world applications.

#### 3.1 Resources

The number of manually simplified texts increases daily due the number of websites that provide easy-to-read materials (Section 1.2). Unfortunately, those materials and their original (non-simplified) versions, are usually not made available for neither research nor commercial purposes. The exceptions are the English and Spanish Newsela corpora,

Dataset	Language	Quality	Target	Domain	Articles×levels
Britanica (Barzilay and Elhadad, 2003)	EN	●	Children	Encyclopedia	114×2
LiteracyNet (Feng et al., 2009)	EN	●	Low literacy	News	115×2
Zero Hora (Caseli et al., 2009)	PT(BR)	○	Children & Low literacy	News	104×3
Dsim (Klerke and Søgaard, 2012)	DA	●	Reading impaired & L2	News telegrams	3701×2
Wikipedia (Kauchak, 2013)	EN	○	Children & L2	Encyclopedia	≈60,000×2
Newsela (Xu et al., 2015)	EN	●	Children & L2	News	1,130×5
Newsela (Štajner et al., 2015)	ES	●	Children & L2	News	≈250*5
Simplext (Saggion et al., 2015)	ES	●	Intellectual disability	News	200×2
Terence (Brunato et al., 2015)	IT	●	Poor comprehenders	Children stories	32×2
Teacher (Brunato et al., 2015)	IT	○	L2	Literature & Handbooks	24×2
NewsWebEasy (Goto et al., 2015)	JP	○	L2	News	490×2
OneStopEng. (Vajjala and Lučić, 2018)	EN	●	L2	News	189×3
Alector (Gala et al., 2020)	FR	●	Children & L2	Literature & Science	79×2

Table 3: Parallel (comparable in the case of Wikipedia) document-level text simplification datasets (● signifies professional simplification, and ○ non-professional simplification following provided guidelines).

Dataset	Language	Quality	Alignment	Type	Domain	Sentence pairs
Enc. Brit. (Barzilay and Elhadad, 2003)	EN	●	Automatic	all	Encyclopedia	601
PorSimples (Specia, 2010)	PT(BR)	○	Controlled	all	News & Science	4,483
PWKP (Zhu et al., 2010)	EN	○	Automatic	all	Wikipedia	≈108,000
EW-SEW v1.0 (Coster and Kauchak, 2011b)	EN	○	Automatic	all	Wikipedia	≈137,000
EW-SEW v2.0 (Kauchak, 2013)	EN	○	Automatic	all	Wikipedia	≈167,000
EW-SEW (Hwang et al., 2015)	EN	○	Automatic	all	Wikipedia	≈280,000
Newsela (Štajner et al., 2015)	EN	●	Automatic	all	News	≈480,000
NewsWebEasy (Goto et al., 2015)	JP	●	Manual+Auto	all	News	13,386
Simplext (Saggion et al., 2015)	ES	●	Manual	all	News	925
PaCCSS-IT (Brunato et al., 2016)	IT	○	Automatic	structural	Web texts	≈63,000
Newsela (Štajner et al., 2018)	ES	●	Automatic	all	News	Not specified
PorSimplesSent (Leal et al., 2018)	PT(BR)	○	Controlled	all	News & Science	4,888

Table 4: Comparable sentence-level text simplification datasets (‘controlled’ signifies that the alignment was automatic but using the original editing tool, thus corresponding close to a manual alignment).

which are available upon request for research purposes (the portion until 2016).<sup>11</sup>

The largest freely available TS corpora up-to-date are those based on matching the original English Wikipedia articles with the articles of the same title in Simple English Wikipedia, the EW-SEW corpora. They are available in several versions, as document-aligned (Kauchak, 2013)<sup>12</sup> and sentence-aligned (Zhu et al., 2010; Coster and Kauchak, 2011b; Kauchak, 2013; Hwang et al., 2015). Apart from not representing a truly parallel corpus but rather just comparable, as the articles in Simple English Wikipedia might have been written completely independently from the articles with the same title in the original English Wikipedia, many objections have been raised regarding the quality of simplifications (Amancio and Specia, 2014; Štajner et al., 2015; Xu et al., 2015; Štajner et al., 2020).

<sup>11</sup><https://newsela.com/data/>

<sup>12</sup><https://cs.pomona.edu/~dkauchak/simplification/>

The English and Spanish Newsela corpora (Xu et al., 2015; Štajner et al., 2018), in contrast, provide parallel texts on several complexity levels, manually simplified under strict quality control (Xu et al., 2015).

The other available parallel (document-aligned) TS corpora are significantly smaller and do not provide enough material for training ATS models (Table 3). As their texts were simplified for different target populations, they cannot be combined to make larger training datasets for any specific target population, as it is known that there are no universal native simplified–language speaker (Sidharthan, 2014). Nevertheless, those smaller TS corpora could be used for evaluation of ATS systems aimed at particular target populations.

The sentence-aligned TS corpora, apart from those obtained from Newsela and EW-SEW, are still very scarce and limited in their size (Table 4).

The main problem with most sentence-aligned TS corpora is that they were automatically aligned

using various sentence similarity measures with high similarity thresholds. This resulted in high number of pairs of identical, or nearly identical, sentences which, if used for training ATS systems, lead to conservative systems that rarely suggest any changes to the original sentence (Štajner and Saggion, 2015). The newer alignment methods (Hwang et al., 2015; Štajner et al., 2017; Paetzold et al., 2017), offer two types of sentence pairs: full matches and partial matches. In NMT-based TS (Štajner and Nisioi, 2018), the EW-SEW dataset (Hwang et al., 2015) resulted in a model with similarly high percentage of changed sentences as the model trained on Newsela sentence-aligned corpus (Štajner et al., 2017). Those results indicate that the EW-SEW dataset has a potential to be used for training state-of-the-art ATS models, if the sentences are carefully matched and filtered.

For building training datasets for ATS in medical domain, van den Bercken et al. (2019) and Van et al. (2020) used the EW-SEW dataset by retaining only those sentence pairs in which the original sentences contain certain number of medical terms.

One of the biggest challenges in ATS is how to collect a parallel dataset of truly strong paraphrases which are necessary for training ATS systems for people that require higher levels of simplification (e.g. people with cognitive or reading disabilities, and people with low literacy levels).<sup>13</sup> Such sentence pairs are still rarely correctly extracted with automatic alignment methods (Štajner et al., 2018).

The sentence-aligned TS corpora in Brazilian Portuguese, the PorSimples (Specia, 2010) and PorSimplesSent<sup>14</sup> (Leal et al., 2018), compiled under the PorSimples project, are the only ones that did not lose the strong paraphrase pairs due to automatic sentence-alignment, as manual simplifications were performed using a special editing tool which recorded the exact sentence alignments (Caseli et al., 2009). Both datasets are still very small to allow for automatically learning strong paraphrases solely from them. Nevertheless, they demonstrate that by providing editing tools for manual simplification, it is possible to automatically extract strong paraphrases from manually simpli-

<sup>13</sup>Strong paraphrases are those paraphrases where vocabulary and sentence structure were changed beyond isolated lexical simplifications, straightforward sentence splitting and passive to active voice conversion. For examples, see Tables 1, 2, and 5 in (Štajner et al., 2018).

<sup>14</sup><https://github.com/sidleal/porsimplesent>

fied corpora. This approach could thus be used in collaboration with various associations and content editors that perform professional manual simplifications for various target populations, to compile large, high-quality training datasets for building ATS systems aimed for those particular end users.

## 3.2 Evaluation

Ideally, ATS systems should be evaluated for the output quality, as well as for its usability, i.e. whether or not they make texts easier to read and understand for the particular reader.

### 3.2.1 Quality

The quality of ATS output should be evaluated for its grammaticality (and naturalness), meaning preservation (ensuring that, albeit some necessary loss of nuances, the core meaning remains unchanged), and preservation of text coherence and cohesion. The last is particularly important for systems that perform syntactic simplifications and sentence splitting which often break the cohesion links and make texts more complex at a discourse level (Siddharthan, 2003, 2006). Meaning preservation is also often compromised, both by syntactic simplification and by lexical simplification. Some unsupervised LS systems (Glavaš and Štajner, 2015; Paetzold and Specia, 2016) particularly suffer from this issue due to the use of word embeddings for synonyms retrieval, which often returns antonyms that appear in similar contexts (Glavaš and Štajner, 2015). Grammatical errors are less common, and can be easily corrected manually in a post-editing step. Manual correction of change in meaning, in turn, can sometimes require more time than simplifying from scratch. Coherence problems in automatically simplified texts are the most expensive, as they might require heavy restructuring of the whole text.

To measure the quality of the automatically simplified texts, Štajner and Glavaš (2017) proposed to measure the post-editing time needed to restore the text's grammaticality, original meaning and coherence. The output quality of the sentence simplification systems is, in contrast, evaluated only for grammaticality and meaning preservation, usually using a five-point Likert scale by native or non-native speakers with high literacy levels.

### 3.2.2 Usability

Usability of ATS systems should ideally be evaluated by measuring reading time and text com-

prehension by end users, usually via eye-tracking methods and comprehension tests. This type of evaluation requires significant time and special training for preparing, collecting and interpreting the data. Preparation of comprehension tests requires experts, and eye-tracking methods require special equipment and trained humans to analyse and interpret the results. The access to vulnerable populations (e.g. people with autism or cognitive disabilities) usually requires special permissions and training. Another obstacle with vulnerable populations is that laboratory conditions and use of comprehension questionnaires provide unnatural reading scenario that sometimes leads to unreliable results (Yaneva et al., 2015, 2016a).

For all those reasons, ATS systems are rarely evaluated for their usability in this way. Notable exceptions are: the eye-tracking analyses of simplification strategies for dyslexic readers (Rello et al., 2013b,c,a), and readers with autism spectrum disorders (Yaneva et al., 2016b), the comprehension-based evaluations of the Simplex system by people with Down's syndrome (Saggion et al., 2015), the OpenBook software by people with ASD (Orăsan et al., 2018), and several ATS systems by non-native low-pay workers (Angrosh et al., 2014).

### 3.2.3 Simplicity

The usability evaluation with end users requires long time and special expertise. To obtain the results faster, in ATS research papers, it became a common strategy to, instead, evaluate a small portion of automatically simplified sentences for their simplicity using a Likert scale by evaluators who are not the intended end users.

In some studies the evaluators are native speakers, e.g. (Yatskar et al., 2010; Baeza-Yates et al., 2015; Saggion et al., 2015), in some non-native speakers, e.g. (Coster and Kauchak, 2011a; Wubben et al., 2012; Glavaš and Štajner, 2013, 2015; Paetzold and Specia, 2016), in some a mixture of the two (Angrosh et al., 2014; Kumar et al., 2020), and in some not specified, e.g. (Specia, 2010; Xu et al., 2016; Mallinson et al., 2020; Alva-Manchego et al., 2020a). In some of those studies, the evaluators are students, i.e. readers with high literacy levels, and in others, Amazon Mechanical Turk workers whose literacy level is usually unknown. Štajner and Nisioi (2018) found that evaluators with high literacy levels rarely notice any difference in complexity of the original and automatically simplified sentences. Other studies have

shown that native and non-native speakers differently perceive sentence complexity (Štajner, 2018) and have different lexical simplification needs (Yimam et al., 2017a,b, 2018). Therefore, a direct comparison of the ATS systems proposed in studies where simplicity was evaluated by different type of evaluators is not possible.

Apart from the type of evaluators, the evaluation procedures used across the ATS community differ also by evaluation type (absolute or relative) and scale (e.g. 0/1, 1–3, 1–5), thus hindering the possibility to directly compare the results reported in different studies (Štajner, 2018). In the absolute evaluation, evaluators are presented with one sentence at the time and asked to rate its simplicity on a certain scale. If provided with different variants of the same sentence one after another, depending on the guidelines, evaluators may sometimes emphasize even the small differences between them by giving them different scores. Similar may happen in relative evaluation of simplicity gain, if evaluators are explicitly asked to count the number of successful lexical and syntactic paraphrases between the two sentences (original and simplified), as in the work of Xu et al. (2016). These approaches are thus helpful if the goal is to spot even the slightest differences between several models. If the goal is, in contrast, to estimate the usability and helpfulness of the systems in a real-world scenario, in a text comprehension task, then the approach in which evaluators are presented with pairs of sentences (original and simplified, without specifying which one is which) and asked whether or not they find one of them easier to understand, proposed by Nisioi et al. (2017), might be better suited.

To enable a direct comparison of results reported in different ATS studies, a standardized evaluation procedure across the field is needed. To establish the right evaluation procedure, it would be necessary to first investigate if some particular group of evaluators can be a good proxy for assessing the complexity of texts/sentences for intended target population. In the case that none of those groups satisfies the condition, the evaluators would have to be provided with detailed evaluation guidelines and trained for spotting potential obstacles for the intended end users. Furthermore, it would be necessary to systematically investigate which evaluation type and scale give the most reliable assessment of potential usability of the ATS systems in a real-world scenario.



### 3.2.4 Automatic Evaluation

With the goal of offering a more standardized way of measuring ATS systems' performances, several test sets have been compiled for lexical simplification in English (De Belder and Moens, 2012; Specia et al., 2012; Horn et al., 2014; Paetzold and Specia, 2015). De Belder and Moens (2012) propose a test set of 430 sentences, each with one marked word and the list of words that can replace it, ordered by their difficulty. Similarly, Specia et al. (2012) compiled a test set of 1710 sentences, each with one marked word and four substitution candidates ordered by their difficulty. Horn et al. (2014) compiled a set of 500 sentences from English Wikipedia, each with one marked word and 50 substitution candidates collected via crowdsourcing using Amazon Mechanical Turk.

All these datasets rely on suggestions and rankings of either non-native speakers (usually with high education levels), or neurotypical native speakers. Therefore, without systematic investigation of how well they reflect simplification needs of other target populations, it is not clear to which extent, if at all, they could be used to evaluate the performances of ATS systems aimed at them.

For automatic evaluation of sentence splitting modules for English, two datasets are currently available: WebSplit (Narayan et al., 2017) and WikiSplit (Botha et al., 2018). Both datasets were automatically extracted and aligned; the WebSplit from the WebNLG corpus (Gardent et al., 2017), and WikiSplit from Wikipedia edit histories. Recently, a crowdsourced dataset with multiple human simplification suggestions (both syntactic and lexical) for 2,359 original English Wikipedia sentences has been released (Alva-Manchego et al., 2020a). So far, none of them has been tested for how well it could evaluate the usability of real-world ATS systems for any particular target population.

Many ATS systems have been automatically evaluated using the Flesch-Kincaid Grade Level (FKGL) (Kincaid et al., 1975), MT-inspired evaluation metrics, e.g. BLEU (Papineni et al., 2002), METEOR (Denkowski and Lavie, 2011), TER (Snover et al., 2009), or more recently proposed TS-specific metrics: SARI (Xu et al., 2016) and SAMSA (Sulem et al., 2018b). None of these metrics, however, can replace the necessary manual evaluation of grammaticality and meaning preservation (at a sentence level), or cohesion and coherence (at a discourse level). Traditional read-

ability metrics such as FKGL were proposed for human-generated texts and are completely oblivious of meaning, i.e. short, meaningless sentences would score very well on those. The use of BLEU for TS evaluation has been reported to have many drawbacks if used for ATS (Štajner et al., 2015; Xu et al., 2016; Sulem et al., 2018a). Some other MT-based evaluation metrics have shown good correlations with human judgements of grammaticality and meaning preservation (Štajner et al., 2014; Popović and Štajner, 2016), but they all favor simplifications that are close to the original sentence (i.e. offer very few changes). SARI, in turn, favors simplifications that are as different from original as possible, thus often rewarding simple sentences which might have significantly altered the meaning.

## 4 Conclusion and Outlook

In this study, the potential for using ATS for social good and main challenges in the field to achieve that goal were emphasized: the lack of large high-quality TS datasets for training ATS systems for particular target populations, and the lack of standardized evaluation procedures for estimating the usability potential of the proposed ATS systems in real-world scenarios. Due to those, it is not clear if the recently proposed neural ATS systems, published in top tier NLP/CL/AI conferences, e.g. (Nisioi et al., 2017; Zhang and Lapata, 2017; Surya et al., 2019; Kumar et al., 2020; Mallinson et al., 2020), represent a real step forward toward using ATS for social good.

To address the above-mentioned challenges, it seems necessary to: (1) establish multidisciplinary collaborations with associations and content editors which have experience in simplifying texts for specific groups of end users; and (2) prioritize work on highly modular ATS systems, which allow for easy customization of components according to the simplification needs of particular groups of users. Multidisciplinary collaborations would enable compiling of larger high-quality datasets for various target populations, better understanding of which ATS modules require higher prioritization, and the possibility for testing the ATS systems with the real end users.

## Acknowledgements

I would like to thank the three anonymous reviewers for their valuable comments and suggestions.

## Ethics/Impact Statement

As this is a special theme paper and it does not propose any novel systems, the main focus of this section is the discussion about the general ethical considerations for automatic text simplification and its impact on the society.

### Intended Use

Automatic text simplification (ATS) has strong potential to improve social inclusion for many people who are otherwise marginalized in one way or another. While many large-scale projects proved this, the last several years have been dominated by ATS research which does not seem to lead towards practical applications of ATS in the society. The goal of this paper is to raise the awareness among researchers about the original motivations for ATS and its potential for being used for social good, as well as to point out the main challenges in the field, and suggest ways forward.

### Failure Modes

As mentioned in Section 3.2, ATS systems often produce ungrammatical sentences, changes to the original meaning, and break the text coherence. For this reason, to avoid potential unintended harm, it is advisable to add post-editing capabilities in ATS systems, as it was done in the OpenBook software (Orăsan et al., 2018). The unintended changes in meaning produced by ATS systems would be especially dangerous in legal and medical domains.

### Biases

As all machine learning (ML) systems, ML-based ATS systems can suffer from algorithmic biases. If trained on data that represents manual simplifications performed by trained human editors, the system might inherit the biases and preferences of the editor. If trained on automatically sentence-aligned data, the system might learn transformations which might rather reflect the alignment methods than the actual simplification methods used in the pre-aligned corpora.

### Misuse Potential

As any technological advancement, ATS could also be misused, e.g. by learning to make text more complex with the intention of providing linguistically complex texts to hide non-desirable facts, as it is known that high percentage of people have problems to understand complex texts (Section 1.1), or

by unintentionally altering the original meaning thus offering incorrect information.

### Potential Harm to Vulnerable Populations

People with low literacy levels, or any kind of reading or cognitive impairments, which cannot understand original texts and thus fully rely on the simplified versions, can potentially be harmed by being provided with incorrect information due to the use of premature ATS systems where the output is not manually checked for meaning preservation. Therefore, it is important that any ATS system intended for those populations offers post-editing capabilities and that its output passes a rigorous manual check for meaning preservation.

Furthermore, if provided by automatically simplified learning material, which was not carefully manually checked for its grammaticality and natural soundness of sentences, children and language learners might learn incorrect word forms or sentences structures.

### Computing Time

The latest trends in ATS research steered the systems towards very computationally expensive (and environmentally unfriendly) neural architectures. As the current neural state-of-the-art ATS systems are still far behind the traditional modular rule-based ATS systems (especially for syntactic simplification), many research groups and organizations need numerous attempts with neural architectures even just to be able to publish a paper in a top-tier conference. The current best performing neural lexical simplification systems (Glavaš and Štajner, 2015; Paetzold and Specia, 2016, 2017a; Qiang et al., 2020a,b) perform better than the state-of-the-art non-neural lexical simplification systems (Glavaš and Štajner, 2015; Paetzold and Specia, 2017b; Qiang et al., 2020b). Nevertheless, due to using heavy resources, such as BERT (Devlin et al., 2018) and PPDB (Ganitkevitch et al., 2013), or their multiple operations with word embeddings, they require significant computational power and are too slow to be used in real-world applications.

## References

- Suha S. Al-Thanyyan and Aqil M. Azmi. 2021. *Automated text simplification: A survey*. *ACM Comput. Surv.*, 54(2).
- Sandra Maria Aluísio 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*, pages 46–53, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. 2020a. [ASSET: A dataset for tuning and evaluation of sentence simplification models with multiple rewriting transformations](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4668–4679, Online. Association for Computational Linguistics.
- Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2020b. [Data-driven sentence simplification: Survey and benchmark](#). *Computational Linguistics*, 46(1):135–187.
- Marcelo Adriano Amancio and Lucia Specia. 2014. An Analysis of Crowdsourced Text Simplifications. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 123–130.
- Mandya Angrosh, Tadashi Nomoto, and Advait Sidharthan. 2014. Lexico-syntactic text simplification and compression with typed dependencies. In *Proceedings the 25th International Conference on Computational Linguistics, COLING*, pages 1996–2006, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Ricardo A. Baeza-Yates, Luz Rello, and Julia Dembowski. 2015. CASSA: A Context-Aware Synonym Simplification Algorithm. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 1380–1385, Denver, Colorado, USA.
- Regina Barzilay and Noemie Elhadad. 2003. [Sentence alignment for monolingual comparable corpora](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 25–32, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Laurens van den Bercken, Robert-Jan Sips, and Christoph Lofi. 2019. [Evaluating neural text simplification in the medical domain](#). In *The World Wide Web Conference, WWW '19*, page 3286–3292, New York, NY, USA. Association for Computing Machinery.
- Jan A. Botha, Manaal Faruqui, John Alex, Jason Baldridge, and Dipanjan Das. 2018. [Learning to split and rephrase from Wikipedia edit history](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 732–737, Brussels, Belgium. Association for Computational Linguistics.
- Stefan Bott, Luz Rello, Biljana Drndarević, and Horacio Saggion. 2012. Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish. In *Proceedings of the 24th International Conference on Computational Linguistics, COLING*, pages 357–374.
- Dominique Brunato, Andrea Cimino, Felice Dell’Orletta, and Giulia Venturi. 2016. PaCCSS-IT: A Parallel Corpus of Complex-Simple Sentences for Automatic Text Simplification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 351–361, Austin, Texas. Association for Computational Linguistics.
- Dominique Brunato, 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*, pages 31–41, Denver, Colorado, USA. Association for Computational Linguistics.
- Rémi Cardon and Natalia Grabar. 2020. [French biomedical text simplification: When small and precise helps](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 710–716, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical Simplification of English Newspaper Text to Assist Aphasic Readers. In *Proceedings of the AAAI’98 Workshop on Integrating AI and Assistive Technology*, pages 7–10.
- Helena M. Caseli, Tiago F. Pereira, Lucia Specia, Thiago A. S. Pardo, Caroline Gasperin, and Sandra M. Aluísio. 2009. Building a Brazilian Portuguese parallel corpus of original and simplified texts. In *10th Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2009)*.
- R. Chandrasekar, Christine Doran, and B. Srinivas. 1996. Motivations and methods for text simplification. In *16th International Conference on Computational Linguistics*, pages 1041–1044.
- R. Chandrasekar and B. Srinivas. 1997. Automatic induction of rules for text simplification. *Knowledge-Based Systems*, 10:183–190.
- Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, and Ben Caldwell. 2010. Understanding WCAG 2.0. A guide to understanding and implementing Web Content Accessibility Guidelines 2.0. World Wide Web Consortium (W3C).
- William Coster and David Kauchak. 2011a. Learning to Simplify Sentences Using Wikipedia. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1–9.
- William Coster and David Kauchak. 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. Portland, Oregon, USA: Association for Computational Linguistics*, pages 665–669.
- David Crystal. 1987. *The Cambridge Encyclopedia of Language*. Cambridge University Press.
- Jan De Belder and Marie-Francine Moens. 2012. A Dataset for the Evaluation of Lexical Simplification. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 7182 of *Lecture Notes in Computer Science*, pages 426–437. Springer Berlin Heidelberg.
- Michael Denkowski and Alon Lavie. 2011. Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems. In *Proceedings of the EMNLP Workshop on Statistical Machine Translation*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Siobhan Devlin and Gary Unthank. 2006. Helping Aphasic People Process Online Information. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 225–226, New York, NY, USA. ACM.
- Biljana Drndarević and Horacio Saggion. 2012a. Reducing Text Complexity through Automatic Lexical Simplification: an Empirical Study for Spanish. *Procesamiento del Lenguaje Natural*, 49:13–20.
- Biljana Drndarević and Horacio Saggion. 2012b. 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*, PITER, pages 8–16, Montréal, Canada. Association for Computational Linguistics.
- Lijun Feng, Noémie Elhadad, and Matt Huenerfauth. 2009. Cognitively Motivated Features for Readability Assessment. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL, pages 229–237, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Geert Freyhoff, Gerhard Hess, Linda Kerr, Bror Tronbacke, and Kathy Van Der Veken. 1998. *Make it Simple, European Guidelines for the Production of Easy-to-Read Information for People with Learning Disability*. ILSMH European Association, Brussels.
- Núria Gala, 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 *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1353–1361, Marseille, France. European Language Resources Association.
- Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. *PPDB: The paraphrase database*. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 758–764, Atlanta, Georgia. Association for Computational Linguistics.
- Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. *Creating training corpora for NLG micro-planners*. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 179–188, Vancouver, Canada. Association for Computational Linguistics.
- Goran Glavaš and Sanja Štajner. 2013. Event-Centered Simplification of News Stories. In *Recent Advances in Natural Language Processing, RANLP 2013*, 9–11, pages 71–78.
- Goran Glavaš and Sanja Štajner. 2015. *Simplifying Lexical Simplification: Do We Need Simplified Corpora?* In *Proceedings of the ACL&IJCNLP 2015 (Volume 2: Short Papers)*, pages 63–68.
- Isao Goto, Hideki Tanaka, and Tadashi Kumano. 2015. Japanese news simplification: Task design, data set construction, and analysis of simplified text. In *Proceedings of Machine Translation Summit XV, Vol. 1: MT Researchers’ Track*, pages 17–31, Miami, FL.
- Colby Horn, Cathryn Manduca, and David Kauchak. 2014. Learning a Lexical Simplifier Using Wikipedia. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, ACL, pages 458–463.
- William Hwang, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu. 2015. Aligning Sentences from Standard Wikipedia to Simple Wikipedia. In *Proceedings of NAACL&HLT*, pages 211–217.
- Joyce Karreman, Thea van der Geest, and Esmee Bursink. 2007. Accessible website content guidelines for users with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 20:510–518.
- David Kauchak. 2013. *Improving text simplification language modeling using unsimplified text data*. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1537–1546. ACL.
- Peter J. Kincaid, Robert P. Fishburne, Richard L. Rogers, and Brad S. Chissom. 1975. Derivation of new readability formulas for Navy enlisted personnel. *Research Branch Report 8-75*.
- Sigrid Klerke and Anders Sjøgaard. 2012. *DSim, a Danish parallel corpus for text simplification*. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 4015–4018, Istanbul, Turkey. European Language Resources Association (ELRA).



- Nicholas Kloehn, Gondy Leroy, David Kauchak, Yang Gu, Sonia Colina, Nicole P Yuan, and Debra Revere. 2018. Improving consumer understanding of medical text: Development and validation of a new sub-simplify algorithm to automatically generate term explanations in english and spanish. *J Med Internet Res*, 20(8):e10779.
- Dhruv Kumar, Lili Mou, Lukasz Golab, and Olga Vechtomova. 2020. Iterative edit-based unsupervised sentence simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7918–7928, Online. Association for Computational Linguistics.
- Sidney Evaldo Leal, Magali Sanches Duran, and Sandra Maria Aluísio. 2018. A nontrivial sentence corpus for the task of sentence readability assessment in portuguese. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, page 401–413, Santa Fe, New Mexico, USA.
- Jonathan Mallinson, Rico Sennrich, and Mirella Lapata. 2020. Zero-shot crosslingual sentence simplification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5109–5126, Online. Association for Computational Linguistics.
- Mencap. 2002. *Am I making myself clear? Mencap’s guidelines for accessible writing*.
- Shashi Narayan and Claire Gardent. 2014. Hybrid simplification using deep semantics and machine translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 435–445.
- Shashi Narayan, Claire Gardent, Shay B. Cohen, and Anastasia Shimorina. 2017. Split and rephrase. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 606–616, Copenhagen, Denmark. Association for Computational Linguistics.
- Sergiu Nisioi, 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 (ACL)*, pages 85–91.
- OECD. 2013. *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Technical report, OECD Publishing.
- Charles Kay Ogden. 1937. *Basic English: A General Introduction with Rules and Grammar*. Paul Treber, London.
- Constantin Orăsan, Richard Evans, and Ruslan Mitkov. 2018. *Intelligent Text Processing to Help Readers with Autism*, pages 713–740. Springer International Publishing, Cham.
- Gustavo Paetzold and Lucia Specia. 2015. LEXenstein: A Framework for Lexical Simplification. In *Proceedings of the ACL-IJCNLP 2015 System Demonstrations*, pages 85–90, Beijing, China.
- Gustavo Paetzold and Lucia Specia. 2017a. Lexical simplification with neural ranking. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 34–40, Valencia, Spain. Association for Computational Linguistics.
- Gustavo H. Paetzold, Fernando Alva-Manchego, and Lucia Specia. 2017. MASSAlign: Alignment and Annotation of Comparable Documents. In *The Companion Volume of the IJCNLP 2017 Proceedings: System Demonstrations*, pages 1–4.
- Gustavo H. Paetzold and Lucia Specia. 2017b. A survey on lexical simplification. *Journal of Artificial Intelligence Research*, 60(1):549–593.
- Gustavo Henrique Paetzold and Lucia Specia. 2016. Unsupervised lexical simplification for non-native speakers. In *Proceedings of the 30th AAAI*.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL*.
- PlainLanguage. 2011. *Federal plain language guidelines*.
- Maja Popović and Sanja Štajner. 2016. Machine translation evaluation metrics for quality assessment of automatically simplified sentences. In *Proceedings of the LREC Workshop on Quality Assessment for Text Simplification (QATS)*.
- Jipeng Qiang, Yun Li, Zhu Yi, Yunhao Yuan, and Xindong Wu. 2020a. Lexical simplification with pre-trained encoders. *Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 8649—8656.
- Jipeng Qiang, Yun Li, Yi Zhu, Yunhao Yuan, and Xindong Wu. 2020b. *Lsbert: A simple framework for lexical simplification*.
- Philip Quinlan. 1992. *The Oxford Psycholinguistic Database*. Oxford University Press.
- Luz Rello, Ricardo A. Baeza-Yates, Stefan Bott, and Horacio Saggion. 2013a. Simplify or help?: text simplification strategies for people with dyslexia. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility, W4A, Rio de Janeiro, Brazil*.
- Luz Rello, Ricardo A. Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013b. Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia. In *Proceedings of the International Conference on Human-Computer Interaction (Part IV), INTERACT*, pages 203–219, Cape Town, South Africa.

- Luz Rello, Susana Bautista, Ricardo A. Baeza-Yates, Pablo Gervás, Raquel Hervás, and Horacio Saggion. 2013c. One Half or 50%? An Eye-Tracking Study of Number Representation Readability. In *Proceedings of the International Conference on Human-Computer Interaction (Part IV)*, INTERACT, pages 229–245, Cape Town, South Africa.
- Horacio Saggion. 2017. *Automatic text simplification*. Morgan & Claypool Publishers.
- Horacio Saggion, Daniel Ferrés, Leen Sevens, Ineke Schuurman, Marta Ripollés, and Olga Rodríguez. 2017. [Able to read my mail: An accessible e-mail client with assistive technology](#). In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work*, W4A '17, New York, NY, USA. Association for Computing Machinery.
- Horacio Saggion, Elena Gómez Martínez, Alberto Anula, Lorena Bourg, and Esteban Etayo. 2011. Text Simplification in Simplext: Making Text More Accessible. *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 47:341–342.
- Horacio Saggion, Sanja Štajner, Stefan Bott, Simon Mille, Luz Rello, and Biljana Drndarević. 2015. Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish. *ACM Transactions on Accessible Computing (TACCESS)*, 6(4):14.
- Leen Sevens, Vincent Vandeghinste, Ineke Schuurman, and Frank Van Eynde. 2017. Simplified text-to-pictograph translation for people with intellectual disabilities. In *Natural Language Processing and Information Systems*, pages 185–196, Cham. Springer International Publishing.
- Matthew Shardlow. 2014. [A survey of automated text simplification](#). *International Journal of Advanced Computer Science and Applications(IJACSA)*, Special Issue on Natural Language Processing 2014, 4(1).
- Matthew Shardlow and Raheel Nawaz. 2019. [Neural text simplification of clinical letters with a domain specific phrase table](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 380–389, Florence, Italy. Association for Computational Linguistics.
- Advait Siddharthan. 2003. *Syntactic simplification and text cohesion*. Ph.D. thesis, University of Cambridge, UK.
- Advait Siddharthan. 2006. Syntactic simplification and text cohesion. *Research on Language & Computation*, 4(1):77–109.
- Advait Siddharthan. 2014. A survey of research on text simplification. *International Journal of Applied Linguistics*, 165(2):259–298.
- Matthew Snover, Nitin Madnani, Bonnie Dorr, and Richard Schwartz. 2009. Fluency, adequacy, or HTER? Exploring different human judgments with a tunable MT metric. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 259–268, Athens, Greece.
- Lucia Specia. 2010. Translating from complex to simplified sentences. In *Proceedings of the 9th international conference on Computational Processing of the Portuguese Language (PROPOR)*, volume 6001 of *Lecture Notes in Computer Science*, pages 30–39. Springer Berlin Heidelberg.
- Lucia Specia, Sujay Kumar Jauhar, and Rada Mihalcea. 2012. SemEval-2012 task 1: English Lexical Simplification. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, SemEval, pages 347–355, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Elior Sulem, Omri Abend, and Ari Rappoport. 2018a. [BLEU is not suitable for the evaluation of text simplification](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 738–744, Brussels, Belgium. Association for Computational Linguistics.
- Elior Sulem, Omri Abend, and Ari Rappoport. 2018b. [Semantic structural evaluation for text simplification](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 685–696, New Orleans, Louisiana. Association for Computational Linguistics.
- Sai Surya, Abhijit Mishra, Anirban Laha, Parag Jain, and Karthik Sankaranarayanan. 2019. [Unsupervised neural text simplification](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2058–2068, Florence, Italy. Association for Computational Linguistics.
- UN. 2020. [Article 9 – accessibility](#). In *United Nations: Convention on the Rights of Persons with Disabilities (CPRD)*. Accessed: 2020-02-01.
- Sowmya Vajjala and Ivana Lučić. 2018. [On-eStopEnglish 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*, pages 297–304, New Orleans, Louisiana. Association for Computational Linguistics.
- Hoang Van, David Kauchak, and GONDY Leroy. 2020. [AutoMeTS: The autocompletable for medical text simplification](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1424–1434, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Piek Vossen, editor. 2004. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, volume 17. Oxford Univ Press.

- Sanja Štajner. 2018. How to make troubleshooting simpler? assessing differences in perceived sentence simplicity by native and non-native speakers. In *Proceedings of the second LREC workshop on Improving Social Inclusion: Tools, Methods and Resources (ISI-NLP 2)*, Miyazaki, Japan.
- Sanja Štajner, Hannah Bechara, and Horacio Saggion. 2015. [A Deeper Exploration of the Standard PB-SMT Approach to Text Simplification and its Evaluation](#). In *Proceedings of ACL&IJCNLP (Volume 2: Short Papers)*, pages 823–828.
- Sanja Štajner, Marc Franco-Salvador, Simone Paolo Ponzetto, Paolo Rosso, and Heiner Stuckenschmidt. 2017. Sentence alignment methods for improving text simplification systems. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 97–102.
- Sanja Štajner, Marc Franco-Salvador, Paolo Rosso, and Simone Paolo Ponzetto. 2018. CATS: A Tool for Customized Alignment of Text Simplification Corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Sanja Štajner and Goran Glavaš. 2017. [Leveraging event-based semantics for automated text simplification](#). *Expert Systems with Applications*, 82:383–395.
- Sanja Štajner, Ruslan Mitkov, and Horacio Saggion. 2014. One Step Closer to Automatic Evaluation of Text Simplification Systems. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR) at EACL*.
- Sanja Štajner and Sergiu Nisioi. 2018. A detailed evaluation of neural sequence-to-sequence models for in-domain and cross-domain text simplification. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Sanja Štajner, Sergiu Nisioi, and Ioana Hulpus. 2020. [CoCo: A tool for automatically assessing conceptual complexity of texts](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 7179–7186, Marseille, France. European Language Resources Association.
- Sanja Štajner and Horacio Saggion. 2015. [Translating from Original to Simplified Sentences using Moses: When does it Actually Work?](#) In *Proceedings of RANLP 2015*, pages 611–617.
- W3C. 2008. [Web Content Accessibility Guidelines \(WCAG\) 2.0](#).
- William M. Watanabe, Arnaldo Candido Jr., Marcelo A. Amâncio, Matheus De Oliveira, Thiago A. S. Pardo, Renata P. M. Fortes, and Sandra M. Aluísio. 2010. [Adapting web content for low-literacy readers by using lexical elaboration and named entities labeling](#). *New Review of Hypermedia and Multimedia*, 16(3):303–327.
- William Massami Watanabe, Arnaldo Candido Junior, Vinícius Rodriguez Uzêda, Renata Pontin de Matos Fortes, Thiago Alexandre Salgueiro Pardo, and Sandra Maria Aluísio. 2009. [Facilita: Reading assistance for low-literacy readers](#). In *Proceedings of the 27th ACM International Conference on Design of Communication, SIGDOC '09*, page 29–36, New York, NY, USA. Association for Computing Machinery.
- WHO. 2013. [Health literacy: The solid facts](#).
- Sander Wubben, Antal van den Bosch, and Emiel Kraemer. 2012. Sentence Simplification by Monolingual Machine Translation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL, pages 1015–1024, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. [Problems in Current Text Simplification Research: New Data Can Help](#). *Transactions of the Association for Computational Linguistics*, 3:283–297.
- Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. [Optimizing statistical machine translation for text simplification](#). *Transactions of the Association for Computational Linguistics*, 4:401–415.
- Victoria Yaneva, Irina Temnikova, and R. Mitkov. 2016a. A corpus of text data and gaze fixations from autistic and non-autistic adults. In *LREC*.
- Victoria Yaneva, 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, ASSETS '15*, page 49–57, New York, NY, USA. Association for Computing Machinery.
- Victoria Yaneva, Irina Temnikova, and Ruslan Mitkov. 2016b. A Corpus of Text Data and Gaze Fixations from Autistic and Non-autistic Adults. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference, LREC*.
- Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2010. [For the sake of simplicity: unsupervised extraction of lexical simplifications from wikipedia](#). In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10*, pages 365–368, Stroudsburg, PA, USA. Association for Computational Linguistics.

Seid Muhie Yimam, Chris Biemann, Shervin Malmasi, Gustavo Paetzold, Lucia Specia, Sanja Štajner, Anaïs Tack, and Marcos Zampieri. 2018. [A report on the complex word identification shared task 2018](#). In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 66–78, New Orleans, Louisiana. Association for Computational Linguistics.

Seid Muhie Yimam, Sanja Štajner, Martin Riedl, and Chris Biemann. 2017a. [Cwig3g2 - complex word identification task across three text genres and two user groups](#). In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 401–407. Asian Federation of Natural Language Processing.

Seid Muhie Yimam, Sanja Štajner, Martin Riedl, and Chris Biemann. 2017b. Multilingual and Cross-Lingual Complex Word Identification. In *Proceedings of RANLP*, pages 813–822, Varna, Bulgaria.

Xingxing Zhang and Mirella Lapata. 2017. Sentence Simplification with Deep Reinforcement Learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594.

Zhemin Zhu, 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 (Coling 2010)*, pages 1353–1361.