# A survey of research on text simplification

Advaith Siddharthan
Department of Computing Science, University of Aberdeen

Text simplification, defined narrowly, is the process of reducing the linguistic complexity of a text, while still retaining the original information and meaning. More broadly, text simplification encompasses other operations; for example, conceptual simplification to simplify content as well as form, elaborative modification, where redundancy and explicitness are used to emphasise key points, and text summarisation to omit peripheral or inappropriate information. There is substantial evidence that manual text simplification is an effective intervention for many readers, but automatic simplification has only recently become an established research field. There have been several recent papers on the topic, however, which bring to the table a multitude of methodologies, each with their strengths and weaknesses. The goal of this paper is to summarise the large interdisciplinary body of work on text simplification and highlight the most promising research directions to move the field forward.

## Introduction

Text simplification, defined narrowly, is the process of reducing the linguistic complexity of a text, while still retaining the original information content and meaning. When simplifying text for certain categories of reader (e.g., children), text simplification can, and perhaps should, be defined more broadly to include operations such as conceptual simplification (where the content is simplified as well as form), elaborative modification (where redundancy and explicitness are used to emphasise key points), and text summarisation (to reduce text length by omitting peripheral or inappropriate information). Either way, the main goal of text simplification is to make information more accessible to the large numbers of people with reduced literacy. Adult literacy is a concern in developed and developing countries; for instance, one in six adults in the UK have poor literacy skills,[1] and only a quarter of Brazilians who have studied for 8 years can be considered fully

1. Source: The National Literacy Trust (http://www.literacytrust.org.uk).

*International Journal of Applied Linguistics* 165:2 (2014), **259–298**. DOI 10.1075/itl.165.2.06sid
ISSN 0019–0829 / E-ISSN 1783–1490 

literate (Aluísio et al., 2008). Other potential beneficiaries of text simplification include children, non-native speakers (including second language learners) and readers with reduced literacy arising from, for example, dyslexia, aphasia or deafness. There is a large body of evidence that manual text simplification is an effective intervention for many readers, but automatic simplification has only recently become an established research topic. There have been several recent papers on the topic, however, bringing to the table a multitude of methodologies, each with their strengths and weaknesses. This is, therefore, a good moment to take stock of the field. The goal of this paper is to summarise the large body of work on text simplification in different disciplines and highlight the most promising research directions to move the field forward.

Section 2 will summarise the motivation for automatic text simplification, including behavioural studies that evaluate the utility of simplifying text for different categories of readers. Section 3 will review the computational approaches that have been applied to text simplification, with an analysis of their strengths and weaknesses. Section 4 will further scrutinise gaps in our understanding of text simplification, and suggest some fruitful avenues for further research.

## The evidence for text simplification

While the main purpose of this review is to survey the methods used for automatic text simplification, it is worth reminding ourselves of why such systems are needed. Here, we summarise studies on how text difficulty impacts on comprehension, examine examples of simplified language from the real world and consider specific examples of target reader populations with reading difficulties. The goals are to understand what makes language difficult or simple, and to identify potentially useful text simplification operations based on the evidence in the literature.

### *Simplified texts and comprehension*

The most striking difference between highly skilled and poor readers is perhaps at the level of word processing. People for whom mapping words to meanings requires effort tend to be poor readers who rely overly on context and higher order mental processes and lack the efficient decoding skills of skilled readers (Anderson & Freebody, 1981). When working memory is devoted to basic word processing, higher level processing suffers (Anderson, 1981; Quigley & Paul, 1984). There are also differences in the way information is aggregated by poor and skilled readers. Skilled readers have a better ability to re-code concepts and relations into larger chunks, thus freeing up working memory for higher level processing (Daneman &

Carpenter, 1980). Mason & Kendall (1979) reported that manually splitting complex sentences into several shorter ones resulted in better comprehension for less skilled readers; they argued that reduction in the amount of information stored in working memory during syntactic processing frees up working memory for higher level semantic processing.

Several studies highlight the role of manual text simplification in comprehension. Students' reading comprehension has been shown to improve when texts have been manually modified to make the language more accessible (L'Allier, 1980), or to make the content more transparent through making discourse relations explicit (Beck et al., 1991). L'Allier (1980) found that text revision brought low ability readers above the performance level of middle ability readers on the original text and Linderholm et al. (2000) found that reformulating causal relations for relatively difficult texts had a significant facilitatory effect for poor readers. Similar results have been found for readers with low levels of domain expertise (Noordman & Vonk, 1992); Kamalski et al., 2008; McNamara et al., 1996). Further, specific information orderings were found to be facilitatory by Anderson & Davison (1988) and Irwin (1980) for readers with different reading ability. Connectives that permit pre-posed adverbial clauses have been found to be difficult for third to fifth grade readers, even when the order of mention coincides with the causal (and temporal) order (Anderson & Davison, 1988); this experimental result is consistent with the observed order of emergence of connectives in children's narratives (Levy, 2003). Thus the b) version of Example 1 below, from Anderson & Davison (1988), p. 35, should be preferred for children who can grasp causality, but who have not yet become comfortable with alternative clause orders:

(1) a. Because Mexico allowed slavery, many Americans and their slaves moved to Mexico during that time.
b. Many Americans and their slaves moved to Mexico during that time, because Mexico allowed slavery.

However, various other studies (Clark & Clark, 1968; Katz & Brent, 1968; Irwin, 1980) have suggested that for older school children, college students and adults, comprehension is better for the cause-effect presentation, both when the relation is implicit (no discourse marker) and explicit (with a discourse marker). Williams et al. (2003) suggested the use of version c) below, as it uses short sentences, a cause-effect information ordering and an easy to understand discourse marker "so":

(1) c. Mexico allowed slavery. So many Americans and their slaves moved to Mexico during that time.

Example (1) illustrates that information can often be presented in different ways, that decisions about information order can help determine lexis and syntax, and that discourse-level presentational choices can affect comprehension in different readers.

To summarise, there is evidence from studies using manually simplified texts that reading comprehension can be improved for reader with poor literacy by substituting difficult words, splitting long sentences, making discourse relations explicit, avoiding pre-posed adverbial clauses, and presenting information in cause-effect order. Such studies provided the early motivation for text simplification as a comprehension aid.

### *Examples of simplified language*

The behavioural studies above typically explore only a small set of simplification operations. It is worth looking at some specific contexts where simplified language is used in real life, to throw light on what other simplification operations might be desirable. The two examples in this section will serve to highlight some additional characteristics of simplified language, which can and should inform automatic text simplification systems.

#### *Motherese*

The most widely used simplified language, often referred to as *motherese* or *parentese*, is the language adults use to talk to children. Research on motherese has documented the grammatical adjustments made by adults when speaking to young children. Some of these adjustments have also been observed in other contexts; for example, bilingual accommodation (e.g., Giles et al., 1973). According to Hayes & Ahrens (1988), many of these adjustments are "systematic simplifications of the adult-to-adult speech standard". Among the most consistently noted of these simplifications are: reduction of pre-verb length and complexity; reduction in the number of verb inflections; replacement of first- and second-person pronouns by third-person pronouns and other salient nouns or names; reduction in the number of embedded clauses and conjunctions; shortening of utterance lengths; reduction in the number of disfluencies and fragments; and slowing of speech rate (Cross, 1977; Papoušek et al., 1987; Gleitman et al., 1984). Hayes & Ahrens (1988) further reported lexical simplification in motherese. While adults on average used 17 rare words per thousand tokens when speaking with other adults, they used just 9 with infants and preschool children and 12 with school children. Brodsky et al. (2007) demonstrated the relative simplicity of motherese by automatically learning a lexicon and grammar to parse a corpus of child-directed language.

While some of the simplification operations documented in studies on motherese are considered by the behavioural studies above, many are not (e.g., verb morphology, pronominal choice, pre-verb complexity and length). The observation about pre-verb length relates to the notion of *weight*, a phenomenon formulated by Behaghel (1930) as "Of two constituents of different size, the larger one follows the smaller one." This phenomenon has been studied extensively in the case of post-verb constituents (e.g., Wasow, 1997). Pre-verb length has a direct bearing on working memory load, as during incremental parsing, pre-verb material needs to be stored in a stack until the verb is encountered. There is then a notion of "end-weight", whereby syntactic complexity in the post-verb constituents is easier to process than in the pre-verb constituents.

*Controlled language*

While text simplification is a relatively recent topic of research in computational linguistics, there has been considerable interest in controlled generation, largely due to interest from industries in creating better (less ambiguous and easier to translate) user manuals (Wojcik et al., 1990; Wojcik & Hoard, 1996). *EasyEnglish*, part of IBMs internal editing environment, has been used as a pre-processing step for machine-translating IBM manuals (Bernth, 1998). *EasyEnglish* aimed to help authors remove ambiguity prior to translation; for example, given the sentence:

(2) a. A message is sent to the operator requesting the correct tape volume,

*EasyEnglish* suggests a choice of the following unambiguous alternatives to the author:

(2) b. A message that requests the correct tape volume is sent to the operator, OR
c. A message is sent to the operator that requests the correct tape volume

Systems like *EasyEnglish* are essentially authoring tools that detect ambiguity, ungrammaticality and complicated constructs and help an author revise a document. They do not revise or generate documents automatically and are controlled-generation *aids* rather than text simplification systems. Alternately, controlled language is used as a medium for ontology experts to create and edit ontologies through unambiguous natural language statements and queries (e.g., Power, 2012).

O'Brien (2003) provided a detailed comparison of 8 different controlled language rule sets used in industry, and distinguished between Human-Oriented Controlled Languages (HOCLs), the purpose of which is to improve readability, and Machine Oriented Controlled Languages (MOCLs), the purpose of which is to improve translatability. MOCLs are particularly important to industry as

multinational corporations need to maintain documentation and manuals in different languages. O'Brien classified rules for controlled languages into four types: lexical, syntactic, textual structure and pragmatic. In general, these rules are intended to standardise writing and eliminate ambiguity. Some examples are shown below (see O'Brien, 2003, for the full list).

1. *Lexical:* Rule out the use of particular acronyms, standardise spelling, rule out the use of synonyms, rule out specific pronouns and ambiguous anaphoric reference, rule out ambiguous conjunctions such as "as", rule out double negations, and restrict words to signal negation, insist on inclusion of relative pronoun, standardise format for numbers and dates, specify dictionary and rule out ambiguous words.
2. *Syntactic*: Specify rules for pre and post-modifier usage, rule out ellipsis, insist on the use of article or demonstrative, restrict size of noun cluster and rule out specific prepositions, such as "of", specify location of prepositions to reduce ambiguity, avoid use of present participle, rule out passive voice, insist on indicative mood, restrict apposition, rule out certain forms of conjunction, specify use of punctuation.
3. *Textual Structure*: Specify when lists or tables should be used, constrain maximum sentence and paragraph lengths, specify keywords to use for coherence, restrict use of parentheticals.
4. *Pragmatic:* Rule out the use of metaphor, slang or idiom, urge author to be as specific as possible.

Some of these rules are consistent with the simplified texts described in previous sections, including the need to restrict the use of vocabulary and syntax. However the guiding principle in controlled languages is not to make texts simpler necessarily, but to reduce the potential for misunderstanding by controlling ambiguity. This is an interesting principle, and perhaps one that text simplification systems need to pay more attention to. For instance, more frequent and shorter words are also more polysemous (Davies & Widdowson, 1974; Walker et al., 2011). Lexical simplification can therefore have the unintended effect of making text ambiguous or misleading, and potentially increase text difficulty.

### *Studies with specific target reader populations*

In Section 2.1, we looked at some user studies that demonstrated the potential of text simplification in schools. We now consider other target reader populations for whom text simplification has been proposed. We will first consider the literature on text simplification for second language learners, which bears some commonalities with the research summarised in Section 2.1. We will then consider three

specific conditions that result in language deficits (deafness, aphasia and dyslexia), as these have motivated systems for automatic text simplification as an assistive technology. There are of course other conditions which result in reading difficulties of various kinds; for example, dementia, or even normal ageing, when working memory limitations can impact on sentence processing skills.

*Second language learners*

There is a large body of literature investigating the role of simplified text in the second language (L2) acquisition process. The justification for this work can be found in Krashen (1985)'s influential input hypothesis that L2 learners acquire language when the input is comprehensible, but just a little beyond their current level of L2. An exhaustive survey of studies on modified input for L2 acquisition is beyond the scope of this article, but we will attempt to summarise the most salient points here.

Numerous studies show that reading comprehension improves for L2 learners when the input is simplified (e.g., Long & Ross, 1993; Yano et al., 1994; Tweissi, 1998; Gardner & Hansen, 2007), and indeed the majority of L2 learning materials at the beginning to intermediate levels still make use of simplified texts (Crossley et al., 2007). However, the community remains divided about the use of simplified texts by L2 learners.

A recurring concern relates to the potentially conflicting goals of improving reading comprehension on specific texts and improving L2 acquisition. Text simplification can deny learners the opportunity to learn the natural forms of language, or slow down language acquisition by removing linguistic items that the reader need to learn (Honeyfield, 1977; Long & Ross, 1993; Yano et al., 1994; Oh, 2001). To address this, some researchers have looked to adapt text in other ways. For instance, Long & Ross (1993) argued in favour of elaborating text rather than simplifying syntax and lexis. Reproducing an example from their paper, the sentence:

(3) a. Because he had to work at night to support his family, Paco often fell asleep in class.

can either be simplified as:

(3) b. Paco had to make money for his family. Paco worked at night. He often went to sleep in class.

or elaborated as:

(3) c. Paco had to work at night to earn money to support his family, so he often fell asleep in class next day during his teacher's lesson.

Long & Ross (1993) reported that the elaborated version improves comprehension without depriving learners of the raw data they need for language development. Green & Olsen (1986) made the additional point that text such as version 3 (b) is often written to make text conform to readability formulae, rather than serve an educational purpose. They concluded that "there is no educationally valid motive for continuing to adapt otherwise suitable texts to meet the demands of readability formulae". The hazards of using readability formulae as guides to writing simplified texts have been noted elsewhere, particularly with regard to the effect on intersentential text cohesion (e.g., Davison & Kantor, 1982).

### *Deafness*

Reading comprehension requires more than just linguistic knowledge. The reader also needs a cognitive base in order to construct a plausible meaning for a sentence. Deaf children face many reading difficulties due to experiential and linguistic deficits incurred in their early childhood (Quigley & Paul, 1984; Marschark & Spencer, 2010) and typically learn to read with inadequately developed cognitive and linguistic skills. As both syntactic analysis and semantic interpretation are constrained by the same working memory (Carpenter et al., 1994), the more the working memory that is required for parsing, the less is there is available for "processing" meaning. As a result, the deaf have trouble comprehending syntactically complex sentences.

Kelly (1996) reported studies that indicate that until deaf readers have achieved a reasonable level of syntactic competence it may be difficult for them to capitalise fully on their vocabulary knowledge. Quigley et al. (1977) reported that 10-year-old deaf children have difficulty with all complex constructs (coordination, subordination, pronominalisation, passive voice and relative clauses), and by the time they are 18, they are better able to comprehend coordination, pronominalisation and passive voice, but still have significant difficulty with relative and subordinate clauses. Robbins & Hatcher (1981) performed comprehension tests on deaf children aged 9–12 years, and reported that passive voice, relative clauses, conjunctions and pronouns affected comprehension the most. Interestingly, Robbins & Hatcher (1981) also found that controlling for word recognition did not improve comprehension on sentences containing these constructs. Various other studies further document the problems deaf readers have with complex syntax and vocabulary (e.g., Marschark & Harris, 1996; Lillo-Martin et al., 1991; Luckner & Handley, 2008).

### *Aphasia*

Aphasia is a language disorder resulting from physical brain damage, usually following a stroke or accident. While the precise reading comprehension issues associated with aphasia depend on the extent and location of brain damage and the level of pre-aphasia literacy, aphasics typically have trouble with long sentences, infrequent words and complicated grammatical constructs. They have themselves identified reading newspapers as a literary task that would help them keep in touch with the world (Parr, 1993).

Shewan & Canter (1971) investigated the relative effects of syntactic complexity, vocabulary and sentence length on auditory comprehension in aphasics. Length was increased by adding prepositional phrases and adjectives, lexical difficulty was measured by frequency of use in normal language and syntactic complexity was increased using passivisation and negations. They concluded that syntactic complexity provided the most difficulty for aphasics. Caplan (1992) reported experiments, involving 142 aphasic patients, that test comprehension on sentences containing different syntactic constructs (active voice, passive voice, relative clauses and coordination). The study showed a significant decrease in comprehension when sentences contained coordinated or relative clauses or passive voice. Comprehension was worst for relative clauses in the subject position, highlighting again the importance of keeping pre-verb length small.

### *Dyslexia*

Dyslexia is a neurological reading disability that is mostly characterised by difficulties with orthography, word recognition, spelling and decoding (Vellutino et al., 2004). While the precise nature of the cognitive deficit is still a matter of debate, there is a general consensus that a phonological deficit contributes to the condition (for an overview, see Ramus, 2003).

Dyslexics typically encounter problems when reading infrequent words and long words. Rello et al. (2013a) described an eye-tracking experiment to study whether dyslexics can benefit from lexical simplification. They found that using more frequent words caused the participants with dyslexia to read significantly faster, while the use of shorter words caused them to understand the text better. The role of syntactic processing in Dyslexia is less clear. There are some studies suggesting the existence of a syntactic processing weakness in readers with dyslexia; for example, in the identification of grammatical roles of verbs (Leikin, 2002), or through reduced hearing comprehension for syntactically complex sentences (Robertson & Joanisse, 2010; Vogel, 1974). However, the main difficulties appear to relate to word processing or orthography.

*Critical summary of manual text simplification*

To summarise, simplified language is used in a variety of contexts, and various target reader populations can benefit from text simplification. For instance, there is evidence that syntactic simplification facilitates comprehension for aphasics and the deaf, while dyslexics benefit from lexical simplification. This support for manual text simplification typically informs and motivates research into automatic text simplification.

There are also arguments against text simplification. As discussed in Section 2.3.1, a frequently expressed concern is that text simplification can impede language acquisition by denying learners the opportunity to learn the natural forms of language (Honeyfield, 1977; Long & Ross, 1993; Yano et al., 1994; Oh, 2001). Another related concern is *homogenisation*: The location of important information is often cued by the presence of unpredictable vocabulary. Simplification homogenises vocabulary across the text, and makes information harder to identify (Honeyfield, 1977). A third related concern, in the context of child language learning, is that children do not seem to find simplified texts interesting (Green & Olsen, 1986). Despite these criticisms, manual text simplification continues to be widespread in language teaching.

Much of the criticism concerns the implementation of text simplification, rather than the concept. As discussed in Section 2.3.1, some researchers argue in favour of elaborating text, rather than simplifying syntax and lexis, to facilitate language acquisition as well as comprehension. Various other studies highlight the hazards of using readability formulae as guides to writing simplified texts, and specifically the potential incoherence caused by removal of discourse connectives such as conjunctions (e.g., Davison & Kantor, 1982; Green & Olsen, 1986). Conversely, Beck et al. (1984) reported that when a text is intuitively simplified to improve coherence, this has the effect of reducing its readability according to common readability metrics as the text typically gets longer. There are numerous psycholinguistic studies that highlight the effect of inter-sentential cohesion on reading time and comprehension (e.g., Mason & Just, 2004; Myers et al., 1987; Keenan et al., 1984).

As we shall see in the next section, research on automatic text simplification has typically avoided issues to do with discourse structure or text cohesion (Siddharthan (2006) and Brouwers et al. (2014) are notable exceptions).

## Automatic text simplification

In recent years, the increasing application of machine translation approaches to text simplification, often referred to as "monolingual translation" and driven by the new availability of corpora of simplified texts, has suggested a dichotomy between manually designed systems with hand-written rules and approaches that learn from corpora using statistical models. This divide is rather artificial, and in reality, text simplification systems have from the beginning explored a variety of linguistic representations to encode simplification operations, and attempted to learn from data. Certain systems focus on syntactic simplification, involving specific constructs such as relative clauses, apposition, coordination, subordination and voice. When the problem is constrained in this manner, it is logical to hand-craft rules: These are likely to be finite (and small) in number, and the system developer is likely to have a good handle on the desired simplifications. On the other hand, most of the simplification operations observed in manually simplified texts involve more than this. When one attempts lexical simplification (word substitution) or syntactic simplifications that have a lexical component, the number of rules becomes too large to code by hand. It is in this context that data-driven approaches come into their own. However, even the earliest text simplification systems experimented with data-driven approaches.

This section will describe the key text simplification systems in roughly chronological order, highlighting their novelty and discussing how the field has evolved over time. It will also bring attention to their drawbacks before we discuss open issues in the field in Section 4.

### *Foundational systems*

Early studies of automatic text simplification covered a lot of ground, exploring hand-crafted systems, systems that learn simplification rules from text (and indeed, adopt ideas from machine translation), and analysing issues of lexical and syntactic simplification as well as text coherence. Some of the insights from these works have been rediscovered in recent years, while others have been forgotten. It is worth reminding ourselves of both kinds.

### *Chandrasekar's approach*

Chandrasekar et al.'s motivation for text simplification was initially to reduce sentence length as a pre-processing step for a parser. They treated text simplification as a two-stage process — *analysis* followed by *transformation*. Their research focus

was limited to syntactic simplification; specifically, dis-embedding relative clauses and appositives and separating out coordinated clauses.

Their first approach (Chandrasekar et al., 1996) was to hand-craft simplification rules, the example from their paper being:

V W:NP, X:REL_PRON Y, Z. → V W Z. W Y.

This can be read as "a string *V* followed by a noun phrase *W* and a relative pronoun *X* and sequence of words *Y* enclosed in commas, followed by a string *Z*, can be split into two sentences "V W Z" and "W Y", with the relative clause *Y* taking *W* as the subject". This rule can, for example, be used to simplify "*John, who was the CEO of a company, played golf.*" to "*John played golf. John was the CEO of a company.*"

In practice, linear pattern-matching rules like the hand-crafted one above do not work very well. For example, to simplify "*A friend from London, who was the CEO of a company, played golf, usually on Sundays.*", it is necessary to decide whether the relative clause attaches to "friend" or "London" and whether the clause ends at "company" or "golf". Text simplification can increase the throughput of a parser only if it reduces the syntactic ambiguity in the text. Hence, a text simplification system has to be able to make disambiguation decisions without a parser in order to be of use to parsing. This early work on syntactic simplification therefore raised more issues than it addressed.

Their second approach (Chandrasekar & Srinivas, 1997) was to have the program learn simplification rules from an aligned corpus of sentences and their hand-simplified forms. This was the earliest work on automatically acquiring text simplification operations, and along with work on automatic syntactic paraphrase by Dras (1999), discussed next, provided the basis of much contemporary work in the field. In Chandrasekar & Srinivas (1997), the original and simplified sentences were parsed using a Lightweight Dependency Analyser (LDA) (Srinivas, 1997) that acted on the output of a super-tagger.[2] These parses were chunked into phrases. Simplification rules were induced from a comparison of the structures of the chunked parses of the original and hand-simplified text. The learning algorithm worked by flattening sub-trees that were the same on both sides of the rule, replacing identical strings of words with variables and then computing tree-to-trees transformations to obtain rules in terms of these variables. Reproducing an example from their paper, Figure 1 shows the LDA parses of a complex sentence and its simplified version. Dis-embedding a relative clause in this manner required three changes:

**2.** A super-tagger localises the computation of linguistic structure by associating with lexical items rich descriptions that impose complex constraints in a local context. These are called super-tags (see Bangalore & Joshi, 1999, for details).

1. the subject relative Super-tag (Rel β) changes to the Transitive super-tag (Trans α).
2. the head of the relative clause is copied in place of the relative pronoun
3. Rel β and its dependants are separated out into a new sentence.

The rule was learnt as a set of tuples of parent-child relations between the super-tags assigned to the nodes in the parse tree. For those familiar with super-tags, the rule learnt from the sentence above is:

(B_COMPs B_PUs) (A_NXN B_NOnxOVnx1) (B_NOnxOVnx1 B_COMPs)
↓
(A_nxOVnx1 B_sPU) (A_nxOVnx1 A_NXN)

Note that the rule is represented entirely as syntax. It thus generalises to a variety of relative clauses with the same structure.

This early work did not progress further. The authors' motivation was to speed up parsing, and despite the theoretical interest in the automatic acquisition of rules, using a parser for text simplification negated that particular goal. Further, this approach required the manual simplification of a reasonable quantity of text. The authors justified this approach on the basis that hand-crafting rules is time consuming. However, the intuitions used to manually simplify sentences could perhaps have been used to directly write simplification rules, and it is unclear

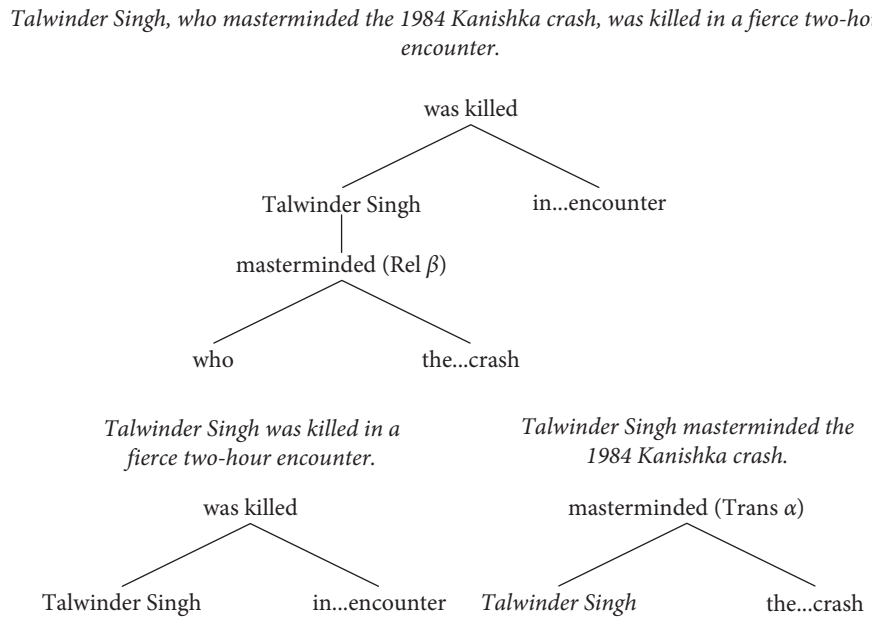


**Figure 1.** Chunked LDA representation of complex and simplified sentences

whether a system that learns a small number of rules from a corpus that has been simplified by hand will outperform a system in which the rules themselves have been hand-crafted. These arguments are worth revisiting when contemporary systems are discussed later on.

### *Dras's doctoral work*

The other key foundational work in text simplification is the PhD dissertation of Mark Dras (Dras, 1999). He refered to the problem of "reluctant paraphrase", where text is altered to fit externally specified constraints such as length, readability or in-house style guides. Like Chandrasekar & Srinivas (1997), Dras used the Tree Adjoining Grammar (TAG) formalism to represent a sentence. However, while Chandrasekar & Srinivas (1997) were only interested in sentence splitting operations, Dras considered a wider set of paraphrase operations. His approach was to map between two TAG grammars using the Synchronous TAG (S-TAG) formalism (more below), and to use Integer Programming to generate a text that satisfies the externally imposed constraints (such as length or readability) using minimal paraphrasing. The two key ideas here — synchronous grammars for monolingual paraphrase and constraint satisfaction using integer programming — have been rediscovered in recent work on text simplification (e.g., De Belder & Moens, 2010; Woodsend & Lapata, 2011; Siddharthan & Angrosh, 2014; Angrosh et al., 2014).

*Syntactic paraphrase:* The first, often overlooked, contribution of Dras is a comprehensive list of paraphrase operations for English that involve purely syntactic rewrites. See Dras (1999) for details, but some interesting examples include:

1. Light verb constructions:
   (a) Steven made an attempt to stop playing Hearts.
   (b) Steven attempted to stop playing Hearts.
2. Clausal Complements
   (a) His willingness to leave made Gillian upset.
   (b) He was willing to leave. This made Gillian upset.
3. Genitives
   (a) The arrival of the train
   (b) The train's arrival
4. Cleft constructions
   (a) It was his best suit that John wore to the ball.
   (b) John wore his best suit to the ball.

Dras (1999) enumerated several other syntactic transformations that could lead to simpler or shorter texts, including deletion operations and sentence splitting operations. Many of these have not been considered in more recent text simplification systems, and perhaps several of these constructs are worth revisiting.

*Synchronous grammars:* Unlike Chandrasekar & Srinivas (1997) who viewed syntactic simplification as a pre-processing tool for applications such as parsing or translation, Dras's motivation for exploring syntactic paraphrase was to study the properties of Synchronous TAGS (S-TAGS) and expand that formalism. The STAG formalism had already been used in applications such as machine translation and syntax-semantics mapping. However, these applications had exposed certain limitations of the formalism. Without getting into details, the standard definition of an S-TAG is a mapping between two parse derivations in different grammars, referred to here as *L* and *R*. An S-TAG derivation is then a pair of derivations $<D_L, D_R>$, where:

1. $D_L$ and $D_R$ are well-formed derivations with respect to their respective grammars
2. $D_L$ and $D_R$ are isomorphic; i.e., there is a one-to-one mapping between nodes in $D_L$ and $D_R$ that preserves dominance.

This second isomorphism property is common to many machine translation frameworks and is particularly problematic for syntactic paraphrase as it severely restricts the ability to reorder or delete sub-trees in the parse. Dras explored ways to generalise the S-TAG formalism to relax the isomorphism constraints. While Chandrasekar & Srinivas (1997) provided a practical system for sentence splitting in the TAG framework, their approach did not constrain the simplification in any manner, and could not provide any guarantees on the well-formedness of the simplified text. Dras explored a more formal approach that had the property of weak language preservation, and allowed for a formal characterisation of the properties of the output text. Specifically, he provided generalisations of the S-TAG formalism to allow for a range of operations including deletion, clause movement and promotion, and sentence splitting. The approach allowed S-TAGS to perform operations like promotion of arbitrarily deep relative clauses, and duplication of noun phrases (needed when making a stand-alone sentence from a relative clause). The generalisations worked with the same basic principles as S-TAG: There are two grammars operating synchronously in parallel, and the mapping continues to be restricted, in order for the weak language preservation property to hold.

*Constraint satisfaction:* The third contribution of Dras was a description of an optimisation framework for text simplification or reluctant paraphrase. The key idea was that a binary variable can be introduced for each paraphrase operation, representing whether the operation is performed or not:

$P_{ij}$: variable representing the *j*th potential paraphrase for sentence *i*.

The objective function to optimise for a whole text (not just a sentence) is then:

$z = \Sigma\ C_{ij}\ P_{ij}$, where $C_{ij}$ is the cost of applying the operation $P_{ij}$.

The cost $C_{ij}$ is the sum of the costs of all the different effects of the paraphrase. These include:

1. Length: Change in the total length of text
2. Readability: Change in the average sentence length
3. Lexical Density: Change in proportion of function words to total words
4. Sequential Variability: To ensure variety in sentence structure (as reflected in sentence length).

To evaluate these costs, Dras found it necessary to impose a restriction on one paraphrase operation per sentence, in order to avoid interactions between operations. This was rather restrictive, but it did mean that certain problems with syntactic simplification, such as the effect on coherence, could be largely avoided. Indeed most work on simplification has avoided consideration of discourse level issues. A key insight from Dras (1999) is the following: It is the properties of the text as a whole that need to be optimised, not the properties of sentences in isolation.

### *The PSET project*

The PSET (Practical Simplification of English Text) project was perhaps the first to apply natural language technologies to create reading aids for people with language difficulties (Devlin & Tait, 1998; Carroll et al., 1998). The emphasis was on automatically simplifying English texts to make them accessible to people with aphasia (see Section 2.3.3 for comprehension issues faced by aphasics).

PSET comprised three components: syntactic simplification, anaphora replacement, and lexical simplification.

*Syntactic simplification in PSET:* For syntactic simplification, the PSET project roughly followed the approach of Chandrasekar et al. PSET used a probabilistic LR parser (Briscoe & Carroll, 1995) for the analysis stage and unification-based pattern matching of hand-crafted rules over phrase-marker trees for the transformation stage.

An example of the kind of simplification rule used in the syntactic-simplification component of the PSET project is:

(S (? a) (S (? b) (S (? c) ) ) ) → (? a) (? c)

The left-hand side of this rule unifies with structures of the form shown in Figure 2 and the rule simply discards the conjunction *(? b)* and makes new sentences out of *(? a)* and *(? c)*. This rule could be used, for example, to simplify "*The proceedings are unfair and any punishment from the guild would be unjustified.*" to "*The proceedings are unfair. Any punishment from the guild would be unjustified.*".

The PSET project simplified two syntactic constructs: coordinated clauses and passive voice. The project reported an accuracy of 75% for simplifying subordination (Canning, 2002), but there were only 75 instances of coordination in the corpus of 100 news reports from the Sunderland Echo. The attempt at converting passive voice to active also had mixed success. Canning (2002) reports that only one out five passive constructs had an expressed surface agent. The rest were agentless as for example, in "*She was taken to Sunderland Royal Hospital*". Further, passive constructs were often deeply embedded within a sentence, making the agent difficult to recover. Canning (2002) reported that in her 100 news report corpus, there were only 33 agentive passive constructs. Out of these, her program converted only 55% correctly to active voice.

To summarise, the PSET project did not research syntactic simplification in depth, and attempted to simplify only two grammatical constructs. The PSET approach to syntactic simplification was ultimately too rudimentary to be useful. A major problem was that parser technology struggled with precisely those sentences that needed simplification. Another problem was that they considered only two syntactic constructs. This meant that there was typically only one simplification made per news report, which is unlikely to have made an impact on readability.

*Pronoun replacement and lexical simplification:* An important contribution of the PSET project was the application of a pronoun resolution algorithm to text simplification (Canning, 2002). The aim was to replace pronouns with their antecedent noun phrases, to help aphasics who might otherwise have difficulty in resolving them. Intra-sentential anaphora were not replaced, to avoid producing sentences like "*Mr Smith said Mr Smith was unhappy*". The anaphora resolution

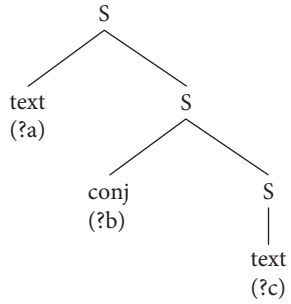


**Figure 2.** The structure matched by the pattern (S (? a) (S (? b) (S (? c) ) ) )

algorithm was based on CogNIAC (Baldwin, 1997), and Canning et al. (2000) reported a recall of 60% with precision of 84% on their newspaper text.

The PSET project (Devlin & Tait, 1998; Carroll et al., 1998) also implemented a synonym substitution system that aimed to replace difficult words (particularly nouns and adjectives) with simpler synonyms. They used *WordNet* (Miller et al., 1993) to identify synonyms, and obtained word frequency statistics from the Oxford Psycholinguistic Database (Quinlan, 1992) to determine the relative difficulty of words (Devlin & Tait, 1998).

### *Siddharthan's doctoral work*

Siddharthan's doctoral work on syntactic simplification aimed to address two specific shortcomings in the work described previously. First, the state of parsing technology and personal computing resources at the time meant that long complex sentences of the kind that would benefit from simplification often timed out, or returned fragments or errorful parses. One aspect of Siddharthan's work was to perform simplification without parsers, following in the spirit of Chandrasekar et al. (1996). This line of work resulted in machine-learning approaches for clause identification and attachment based on part of speech tags and shallow chunking, and was indeed able to demonstrate improvement in parser performance (Siddharthan, 2003a). While this was interesting at the time, perhaps advances in parsing technology and computational power have made these optimisations less important. Parsers do not time out anymore. However, some of the arguments from this thesis continue to hold; e.g., parsers are still suboptimal in analysing some constructs central to text simplification, such as relative clause attachment.

The second, more persistent contribution to the field, was a detailed analysis of the discourse and coherence implications of syntactic simplification. These effects become particularly important when, for example, Dras's "one paraphrase operation per sentence" constraint is relaxed. Recent systems have tended to ignore the discourse level implications of syntactic simplification, and as discussed later, have to some extent avoided the issue by evaluating systems on individual sentences. It is therefore worth summarising the key issues here.

*Syntactic simplification and text cohesion:* Siddharthan's PhD research involved a detailed study of syntactic simplification, with an emphasis on the preservation of text cohesion (Siddharthan, 2003b, 2006). He considered simplifying relative clauses, apposition, coordination and subordination, showing that these constructs can be reliably simplified in news reports using robust and shallow text analysis techniques, and that computational models of discourse structure (Rhetorical Structure Theory (Mann & Thompson, 1988), Centering (Grosz et al., 1995) and Salience (Lappin & Leass, 1994)) can be used to minimise the disruption

in discourse structure caused by syntactic rewriting. For example, his approach (Siddharthan, 2003a) ensured that the sentence 4(a) was simplifed as 4(b) and not the misleading 4(c):

(4) a. Mr. Anthony, who runs an employment agency, decries program trading, but he isn't sure it should be strictly regulated.
b. Mr. Anthony runs an employment agency. Mr. Anthony decries program trading. But he isn't sure it should be strictly regulated.
c. Mr. Anthony decries program trading. Mr. Anthony runs an employment agency. But he isn't sure it should be strictly regulated.

It is possible that alterations at the level of syntax can degrade the text at the level of discourse in other ways; in particular, the referents of pronouns can become ambiguous or hard to resolve. Siddharthan (2003c) also identified and fixed such cases. Consider the two sentences in Example 5 (a):

(5) a. Dr. Knudson found that some children with the eye cancer had inherited a damaged copy of chromosome No. 13 from a parent, who had necessarily had the disease. Under a microscope he could actually see that a bit of chromosome 13 was missing.

When the first sentence is simplified, the resultant text is:

(5) b. Dr. Knudson found that some children with the eye cancer had inherited a damaged copy of chromosome No. 13 from a parent. This parent had necessarily had the disease. Under a microscope **he** could actually see that a bit of chromosome 13 was missing.

Siddharthan (2003c) used a computational model of the reader's attentional state to detect that the pronoun "he" in the succeeding sentence is difficult to resolve correctly, and replaced it with its antecedent, "Dr Knudson".

Siddharthan motivated the need for a *regeneration* component in text simplification systems by showing how naive syntactic restructuring of text could significantly disturb its discourse structure. He formalised the interactions between syntax and discourse during the text simplification process and showed that in order to preserve conjunctive cohesion (how text fragments are connected with discourse cues) and anaphoric coherence (how easy it is for the reader to resolve anaphora), it was necessary to model both intentional structure and attentional state.

### *Contemporary systems*

Text simplification can be viewed as an example of a monolingual translation task, where the source language needs to be translated into a simplified version of the

same language. Many of the foundational systems described above already take inspiration from this translation analogy, and this trend has accelerated in recent years due to the availability of the Simple English Wikipedia (simple.wikipedia.org) as a corpus of simplified English.

A parallel corpus of aligned source and target (simplified) sentences can be created by (a) using Wikipedia revision histories to identify revisions that have simplified the sentence, and (b) aligning sentences in Simple English Wikipedia with sentences from the original English Wikipedia articles. Many contemporary systems apply machine translation approaches to learn text simplification from this parallel Wikipedia corpus. Some are extensions of models previously used for a related monolingual translation task called Sentence Compression, where the focus is on reducing the length of a sentence by deleting words and phrases. Others directly use phrase or syntax based machine translation. There is also ongoing research into syntactic simplification that continues to use hand-written transfer rules. This section discusses the strengths and weaknesses of all these approaches.

*Extensions of sentence compression approaches*

Some contemporary work in text simplification has evolved from research in sentence compression, a related research area that aims to shorten sentences for the purpose of summarising the main content. Sentence compression has historically been addressed in a generative framework, where transformation rules are learnt from parsed corpora of sentences aligned with manually compressed versions, using ideas adapted from statistical machine translation. The compression rules learnt are typically syntactic tree-to-tree transformations (Knight & Marcu, 2000; Galley & McKeown, 2007; Riezler et al., 2003) of some variety. These approaches focused on *deletion* operations, mostly performed low down in the parse tree to remove modifiers and adjuncts. They also made assumptions about isomorphism between the aligned trees, which meant they could not be readily applied to more complex reformulation operations such as *insertion* and *reordering*, and sentence-splitting operations are particularly troublesome to implement. Cohn & Lapata (2009) provided an approach to sentence compression based on Synchronous Tree Substitution Grammar (STSG) that in principle could handle the range of reformulation operations. There were various parallels to the Synchronous Tree Adjoining Grammar approach of Dras (1999) developed for reluctant paraphrase. However, given Cohn & Lapata (2009)'s focus on sentence compression, they did not demonstrate the expressivity of their framework for sentence simplification.

Woodsend & Lapata (2011) further developed this line of research. Their model was based on quasi-synchronous grammar (Smith & Eisner, 2006) and integer linear programming. Quasi-synchronous grammars, like the Generalised

Synchronous TAGs of Dras (1999), aim to relax the isomorphism constraints of synchronous grammars, in this case by generating a loose alignment between parse trees. Woodsend & Lapata (2011) use quasi-synchronous grammars to generate all possible rewrite operations for a source tree, and then integer linear programming to select the most appropriate simplification. They used very similar constraints to those of Dras for linear programming, but differed in an important way. Unlike Dras, optimisation was done at the sentence level; i.e., it was not the characteristics of the whole text that was optimised, but individual sentences. This meant that many aspects of good writing were harder to model, such as variation in sentence length, and indeed complexity. Recently, Siddharthan & Angrosh (2014) described a synchronous dependency grammar for text simplification, that combines a manually constructed grammar for syntactic rules and an automatically acquired grammar for lexical rules and paraphrase. Extending this, Angrosh et al. (2014) presented a system that closely followed Dras by optimising the characteristics of a text, rather than individual sentences, using a linear programming framework.

*Text simplification as monolingual machine translation*

Specia (2010), Wubben et al. (2012) and Coster & Kauchak (2011) applied Phrase Based Machine Translation (PBMT) to the task of text simplification, making use of the Moses open source toolkit for statistical machine translation (Koehn et al., 2007). PMBT is a two stage process. The first stage is alignment. Usually, the input is a pair of source-target sentences. The first step is word alignment; these word alignments are then extended to created phrase alignments, where a phrase is just a sequence of words. PMBT does not make use of syntax. The output of the alignment stage is a phrase table containing aligned sequences of words in source and target language, each with a probability indicating the likelihood of the phrase translation. The second step in PMBT is referred to as decoding. This process uses the phrase table and a language model of the target language to find the best translation of a source sentence to the target sentence. PMBT is well suited to the monolingual task, as phrases that are not in the phrase table can be copied over directly. Indeed, the original sentence would be a valid translation of itself; the challenge is to design a decoder that facilitates simplification.

While Specia (2010) used the Moses toolkit off the shelf, Wubben et al. (2012) and Coster & Kauchak (2011) approached the decoding stage differently. Coster & Kauchak (2011) focused on deletion operations, by extending the alignment stage to allow alignments between a source phrase and an empty target phrase. In contrast, Wubben et al. (2012) extended the decoder to re-rank translations based on dissimilarity. The aim was to find phrase alignments where the simple phrase

is as different as possible to the original phrase, the intuition being that such paraphrases are most likely to simplify the text. Note that PMBT can only perform a small set of simplification operations, such as lexical substitution, deletion and simple paraphrase. They are not well suited for reordering or splitting operations.

In contrast, Zhu et al. (2010) presented an approach based on syntax-based SMT (Yamada & Knight, 2001), consisting of a translation model, a language model and a decoder. The translation model encoded probabilities for four specific rewrite operations on the parse trees of the input sentences: substitution, reordering, splitting, and deletion. Splitting was encoded as two probabilities. A segmentation table stored probabilities of sentence splitting at particular words (e.g., *which*). A completion table stored probabilities of the splitting word to be deleted from the translation, and for the governing phrase to be inserted to complete the sentence. This allowed the translation model to handle constructs such as relative clauses and apposition. Other translation tables held probabilities for substitutions, reorderings and deletions. Their decoder combined a trigram language model with an optimisation of node probabilities in the generated tree.

### *Hand-crafted systems*

Hand-crafted systems for text simplification typically make use of transfer rules that operate on the output of a parser. Various systems use phrasal parse trees as the representation, following the approach of the PSET project (Canning, 2002). Candido Jr et al. (2009) presented a rule-based system to automatically simplify Brazilian Portuguese text for people with low literacy skills. They proposed a set of operations to simplify 22 syntactic constructs, as identified in Aluísio et al. (2008) through manual analysis of simplified texts. The operations, not all of which are implemented in their system, are to (a) split the sentence, (b) replace a discourse marker with a simpler and/or more frequent one, (c) change passive to active voice, (d) invert the order of the clauses, (e) convert to subject-verb-object ordering, and (f) change topicalization and detopicalization of adverbial phrases. Brouwers et al. (2014) followed a similar approach for syntactic simplification of French, using handcrafted rules based on a typology of simplification rules extracted manually from a corpus of simplified French.

De Belder & Moens (2010) used a rule-based system to simplify the same constructs as Siddharthan (2006): apposition, relative clauses, subordination and coordination. Their representation was phrasal parse trees, as produced by the Stanford Parser (Klein & Manning, 2003). They followed Dras (1999) in deciding which sentences to simplify through constraint satisfaction at the level of the entire document, instead of on a per sentence basis.

Other systems use dependency parses as the representation to write transformation rules. Bott et al. (2012) described a text simplification for Spanish that can simplify relative clauses, coordination and participle constructions. Additionally they also performed "quotation inversion", where they replace constructs such as "*Quoted speech, said X*" with "*X said: Quoted speech*". Siddharthan (2010) described a framework that could handle a much wider range of lexico-syntactic simplification operations using transformation rules over type dependency structures. The approach was demonstrated using rules to reformulate sentences expressing causality (e.g., "*The cause of the explosion was an incendiary device*" to "*The explosion occurred because of an incendiary device*"). Siddharthan (2011) built on that work, implementing rules for simplifying relative clauses, apposition, voice conversion, coordination and quotation inversion. Siddharthan & Angrosh (2014) used aligned corpora to acquire lexicalised transfer rules within this framework.

*Comparison of contemporary text simplification approaches*

The systems described above differ primarily in the level of linguistic knowledge they encode. PBMT systems use the least knowledge, and as such are ill-equipped to handle simplifications that require morphological changes, syntactic reordering or insertions. While syntax-based approaches use syntactic knowledge, they need not offer a treatment of morphology. Both Zhu et al. (2010) and Woodsend & Lapata (2011) used the Stanford Parser (Klein & Manning, 2003) for syntactic structure, which does not provide morphological information. This means that while some syntactic reordering operations can be performed well, others requiring morphological changes cannot. Consider converting passive to active voice (e.g., from "*trains are liked by John*" to "*John likes trains*"). Besides deleting auxiliaries and reordering the arguments of the verb, there is also a requirement to modify the verb to make it agree in number with the new subject "John", and take the tense of the auxiliary verb "are".

Hand-crafted systems such as Siddharthan (2011) use transformation rules that encode morphological changes as well as deletions, re-orderings, substitutions and sentence splitting, and can handle voice change correctly. However, hand-crafted systems are limited in scope to syntactic simplification. While purely syntactic rules can be written by hand, there are too many lexico-syntactic and lexical simplifications to enumerate manually. This is where the statistical systems gain an advantage. There have been recent attempts at combining approaches to create hybrid systems; for example, combining hand written rules for syntax with automatically acquired rules for lexicalised constructs (Siddharthan & Angrosh, 2014), and using PBMT for lexicalised paraphrase and deeper semantics for syntactic simplification (Narayan & Gardent, 2014).

The above discussion raises the question of how different systems using different levels of linguistic knowledge compare in practice. Unfortunately, this question is difficult to answer. While there have been some evaluations comparing phrase-based and syntax-based MT frameworks, these have not been compared to hand-crafted text simplification systems. Further, all the evaluations of the individual systems discussed above are on a small scale, performed on sentences in isolation, and performed using either automatic metrics or using ratings by fluent readers. As such, none of these evaluations can help us answer the basic question: Would a poor reader benefit from reading a document simplified by computer? We will further discuss the evaluation of text simplification systems in Section 3.3.

*Lexical simplification*

Research on lexical simplification has tended to fall into one of two categories, expanding difficult words with dictionary definitions (e.g., Kaji et al., 2002) or other explanations, and lexical substitutions — replacing difficult words with easier synonyms.

The former is particularly important when simplifying technical documents for lay readers, since technical terms might not have easy synonyms. Elhadad (2006) used corpus frequencies from the Reuters Health E-line news-feed (www.reutershealth.com), a resource in which Reuters journalists summarise technical publications such as clinical trials for lay readers, to determine how difficult medical terms are for lay readers. They reported a 70% recall and 90% precision in identifying difficult terms, as judged by readers. Zeng-Treitler et al. (2008) further explored a contextual network algorithm to estimate consumer familiarity with health terms. They created a graph of term co-occurrence; i.e., each node in the graph represented a medical term, and edges in the graph represented co-occurrences between the terms. The graph was initialised with pre-existing knowledge, so that a set of nodes was assigned a familiarity score of 0, and another set was assigned a familiarity score of 1. They presented an algorithm that infered familiarity of the other nodes based on connections to known nodes, and reported that their contextual model outperformed frequency-based models, with a significantly higher correlation to human familiarity judgements.

Elhadad (2006) used the Google "define:" functionality to retrieve definitions of terms, reporting a comprehensibility rating of 3.7 for sentences with the definitions added, compared to 2.2 without definitions and 4.3 for ideal definitions, as provided by medical experts. Zeng-Treitler et al. (2007) and Kandula et al. (2010) went one step further. They identified difficult terms in the text and simplified them either by replacing them with easier synonyms or by explaining them using simpler terms that were related, using a short phrase to describe the relationship

between the difficult term and the selected related term; e.g., the technical term "*Pulmonary atresia*" was simplified as "*Pulmonary atresia (a type of birth defect)*".

Several groups have studied lexical substitution. The PSET project (Devlin & Tait, 1998) implemented a synonym substitution system that aimed to replace difficult words (particularly nouns and adjectives) with simpler synonyms. They used *WordNet* (Miller et al., 1993) to identify synonyms, and obtained word frequency statistics from the Oxford Psycholinguistic Database (Quinlan, 1992) to determine the relative difficulty of words. De Belder & Moens (2010) extended this approach to use limited word sense disambiguation using a latent words language model which used Bayesian Networks to represent word sequences and contextual meanings of words. Walker et al. (2011) highlighted ambiguity as another factor to take into account. They noted that there was a correlation between the corpus frequency of words and the number of WordNet senses they have. They reported that readers expressed a slight preference towards unambiguous but less common words over more common but ambiguous words.

More recently, there has been interest in using the Simple English Wikipedia. Biran et al. (2011) defined the corpus complexity of a word as the ratio of its frequencies in English Wikipedia and Simple English Wikipedia. They multipled this value with the length of the word to estimate its difficulty. They used distributional similarity to identify synonym pairs, using a 10 word window around each occurrence of a word in English Wikipedia to create its context vectors, and using the cosine metric to calculate similarity. They showed that this method performed better than a baseline approach of replacing words with their most frequent synonym from WordNet, for grammaticality of output, meaning preservation, as well as simplicity.

Yatskar et al. (2010) more directly used Simple English Wikipedia edit histories to mine lexical simplifications. There are different types of edit operations (e.g., correct, simplify or spam) that Wikipedia editors can perform. The goal of the paper was to identify the simplify operations. They identified "trusted" simplification operations in two ways, (1) by searching the metadata added by editors to revisions for the expression "*simp*" (which would match morphological variants of "simplify"), and (2) by a probabilistic model that captured this mixture of different edit operations. However, they only aimed to acquire simplification pairs devoid of context; they did not evaluate a lexical simplification component through user evaluation at the sentence level.

Deléger et al. (2013) described the construction of a parallel corpus in French of technical and lay writing in the medical domain. They used this to extract paraphrases that represent simplification of technical to lay writing, like equivalences between nominal and verbal constructions, or Greek/Latin derived and modern

language terminology (Deléger & Zweigenbaum, 2009), and the adaptation of the techniques to English (Deléger & Zweigenbaum, 2010).

Other related tasks have been suggested. Specia et al. (2012) described the results of a shared task for lexical simplification. The task required that annotators and systems rank a number of alternative substitutes all deemed adequate for a target word in context, according to how "simple" these substitutes were. The task only required ranking of words by simplicity. Most participants relied on corpus frequencies to rank words, differing mainly in the choice of corpus. The best performing system (Jauhar & Specia, 2012) additionally used psycholinguistic features associated with words, such as how concrete, visual, familiar they are, and their typical age of acquisition.

### *Text simplification in different languages*

Today, there is active research in text simplification systems for a variety of languages, including Basque (Aranzabe et al., 2012), Bulgarian (Lozanova et al., 2013), Danish (Klerke & Søgaard, 2013), Dutch (Daelemans et al., 2004), English (e.g., De Belder & Moens, 2010; Zhu et al., 2010; Coster & Kauchak, 2011; Siddharthan, 2011; Woodsend & Lapata, 2011; Wubben et al., 2012; Siddharthan & Angrosh, 2014; Narayan & Gardent, 2014), French (Seretan, 2012; Brouwers et al., 2014), Korean (Chung et al., 2013), Italian (Barlacchi & Tonelli, 2013), Japanese (Inui et al., 2003), Portuguese (Aluísio et al., 2008; Watanabe et al., 2009), Spanish (Bott et al., 2012) and Swedish (Smith & Jönsson, 2011; Abrahamsson et al., 2014). To a large extent, the principles of text simplification remain the same; multi-clause sentences are split, and difficult vocabulary is replaced. However, different languages introduce different challenges. For instance, Japanese uses three character sets: Kanji are ideographic and generally considered difficult to learn; however, readers familiar with Chinese might find this easier than the two phonetic scripts Hiragana and Katakana. To provide another example, research on English assesses word familiarity or difficulty through corpus frequencies. However, this method needs to be adapted for compounding languages such as Swedish (e.g., Abrahamsson et al., 2014). A challenge facing most languages is the lack of large corpora of simplified text such as the Simple English Wikipedia. This means that systems are typically constructed using hand-crafted rules. Increasingly, this process is informed by a manual or semi-automated analysis of a limited sample of simplified language (e.g., Aluísio et al., 2008; Brouwers et al., 2014). Specia (2010) is an exception; she manually created a corpus of 4483 simplified sentences aligned with the originals and used phrase-based machine translation).

*Evaluating Text Simplification Systems*

As mentioned in Section 3.2.4, there is no consensus on how text simplification systems should be evaluated. Evaluations in the literature have tended to be on a small scale, at the level of a sentence, and evaluated either using automatic metrics or manually by fluent readers. Such evaluations do not really help us understand the utility of text simplification systems for target populations. It is therefore worth considering the evaluation of text simplification systems in more detail.

Recent approaches to evaluating computer-generated text tend to consider either "naturalness" or "usefulness". Following evaluation methodologies commonly used for machine translation and summarisation, there have been attempts to measure naturalness by comparison to human-generated gold standards. An early example is Langkilde-Geary (2002), who generate sentences from a parsed analysis of an existing sentence, and evaluate by comparison to the original. More recently, several text simplification papers have used this approach (Zhu et al., 2010; Coster & Kauchak, 2011; Woodsend & Lapata, 2011; Wubben et al., 2012). Zhu et al. (2010) evaluated their system on a set of 100 sentences from English Wikipedia, aligned with 131 sentences from Simple English Wikipedia. They used the BLEU and NIST scores (Machine Translation metrics that measure word and word sequence overlap between the system output and manual translations) and also reported various readability scores that only take into account the output sentence, such as the Flesch Reading Ease test and n-gram language model perplexity. Coster & Kauchak (2011) evaluated their system against two sentence compression systems (Knight & Marcu, 2000; Cohn & Lapata, 2009), showing improvement in BLEU scores against these systems. Woodsend & Lapata (2011) and Wubben et al. (2012) also reported BLEU scores.

However, this approach has been criticised at many levels (see for example, Sripada et al. (2003)); for instance, because there are many good ways to realise a sentence, and fluency judgements in the monolingual case are more subtle than for machine translation. There is thus considerable scepticism about the use for automatic metrics such as BLEU or ROUGE for evaluating text simplification systems.

Readability metrics, by comparison, do not rely on reference texts, and try to model the linguistic quality of a text based on features derived from the text. This body of work ranges from the Flesch Metric (Flesch, 1951), which is based on average word and sentence length, to more systematic evaluations of various lexical, syntactic and discourse characteristics of a text (e.g., Pitler & Nenkova, 2008, who assess readability of textual summaries). Developing automatic metrics to better evaluate text quality is still an active research area (e.g., Pitler & Nenkova, 2008; Louis & Nenkova, 2013). Even with recent advances though, readability metrics

only provide indirect assessments of grammaticality or comprehensibility. As Bott et al. (2012) observe, it is necessary to distinguish readability from *comprehensibility*, particularly in the context of assistive technologies.

More widely accepted evaluation methods typically involve the solicitation of human judgements. Some researchers have suggested measuring edit distance by using a human to revise a system generated text and quantifying the revisions made (Sripada et al., 2003). Several text simplification applications are evaluated by asking fluent readers which version they prefer (Siddharthan et al., 2011), or through the use of Likert scales (Likert, 1932) for measuring fluency or grammaticality (e.g., Siddharthan, 2006; Woodsend & Lapata, 2011; Wubben et al., 2012). However, such approaches that use judgements by fluent readers tell us very little about the comprehensibility of a text for target reader populations.

In the psycholinguistic literature, various offline and online techniques have been used to investigate sentence processing by readers. Online techniques (eye-tracking (Duchowski, 2007), neurophysiological (Friederici, 1995), etc.) offer many advantages in studying how readers process a sentence. Though these are difficult to set up and also resource-intensive, eye-tracking in particular is beginning to be used in text simplification evaluations (e.g., Bott et al., 2012): Fixation time is particularly relevant to evaluating lexical simplification. There are also a few instances of offline techniques being used to test comprehension in the context of text simplification research. For example, Jonnalagadda et al. (2009) used Cloze tests (Taylor, 1953), and Siddharthan & Katsos (2012) explored magnitude estimation (Bard et al., 1996) and sentence recall (Lombardi & Potter, 1992; Potter & Lombardi, 1990). Canning (2002) reported reading times and scores on question answering tests. Still, there are very few instances of text simplification that have been evaluated with target reader populations. Typical evaluations have involved collecting ratings for fluency and/or correctness from fluent readers, but this is unsatisfactory. There are a range of methods described in behavioural studies of sentence processing, but these have yet to gain acceptance as an evaluation method for text simplification systems.

### *Applications of Automatic Text Simplification*

Text simplification systems have been used in different ways. Early work looked at using syntactic simplification as a pre-processor to improve parser performance (Chandrasekar et al., 1996; Siddharthan, 2003a), and Dras (1999) used text simplification as an application to study the formal properties of generalised synchronous grammars. More recently, Heilman & Smith (2010) used sentence simplification as a first step towards generating factual questions from texts. However, most applications of text simplification still fall under the following categories.

*Assistive Technologies*

Assistive technologies remain the main motivation for many groups researching text simplification. Candido Jr et al. (2009) described a web authoring tool to help writers create simplified text. Watanabe et al. (2009) described a reading assistance tool for Portuguese to facilitate low literacy readers through text simplification. Petersen (2007) applied automatic text simplification to preparing texts for teaching English as a foreign language. De Belder & Moens (2010) built a text simplification system aimed at children. Daelemans et al. (2004) applied automatic sentence simplification to TV programme subtitles to help deaf viewers. Lozanova et al. (2013) targeted Bulgarian sign language users, and Chung et al., (2013) simplified web documents for readers. Automatic text simplification has also been applied to Aphasia (Devlin & Tait, 1998; Carroll et al., 1998), Dyslexia (Rello et al., 2013b) and Autism (Evans et al., 2014). Several researchers have investigated the use of text simplification for facilitating access to medical texts by simplifying terminology (Elhadad, 2006; Zeng-Treitler et al., 2007; 2008; Kandula et al., 2010).

*Summarisation*

Text Simplification has been applied to multi-document summarisation tasks, where short summaries (often 100 words) need to be generated from a set of related news stories. Various other summarisers use syntactic simplification, often as a means for sentence shortening, prior or post sentence extraction (Conroy & Schlesinger, 2004; Vanderwende et al., 2007; Siddharthan et al., 2011). Most systems use simplification to remove peripheral information from the summary. Siddharthan et al. (2011) in contrast argued that inclusion of information present in relative clauses, apposition and copula should be made on the basis of how familiar and salient people and organisations are. They used syntactic simplification to collect such parenthetical information, and select facts to include through a referring expression generation component.

*Information Extraction*

Klebanov et al. (2004) introduced the notion of an Easy Access Sentence (EAS), which they defined in the context of a text *T* as a grammatical sentence with only one finite verb, which does not make any claims not present in *T*. They further suggested that an EAS is to be preferred if it contains more named entities (and therefore less pronominal or other referring expressions). Thus an EAS is in effect a linguistic realisation of a single fact, aimed at making information extraction easier. In the authors' words (page 744), "to mediate between the information-rich

natural language data and applications that are designed to ensure the effective use of canonically structured and organized information…". They generated EASs from a dependency parse and reported a precision and recall of 50% and 30% respectively of their system compared to manually identified EASs.

Various groups have used text simplification components as a pre-processing tool for information extraction and text-mining application in the bio-medical domain (Jonnalagadda et al., 2009; Ong et al., 2007; Miwa et al., 2010; Peng et al., 2012). Such applications rely on identifying lexico-syntactic patterns in text that express the semantic information to be mined. By simplifying complex sentences first, and then matching patterns in the simplified sentences, certain problems with data sparsity during pattern acquisition can be overcome. In effect, text simplification is used to create a form of controlled language to assist information extraction. For example, Peng et al. (2012) reported that syntactic simplification results in a 20% improvement in recall and 10% improvement in accuracy in identifying sentences pertaining to biological events.

## General conclusions and discussion

Today, the field of automatic text simplification enjoys a high profile, and there are numerous international workshops organised at major computational linguistics conferences. PITR (Predicting and Improving Text Readability for target reader populations) aims at bringing together researchers interested in research issues around readability and text simplification, with a focus on work involving target user groups. NLP4ITA (Natural Language Processing for Improving Textual Accessibility) is focused on tools and resources aimed at target populations. Both workshops are supported by the newly formed SIGSLPAT (Special Interest Group on Speech and Language Processing for Assistive Technologies) within ACL (Association for Computational Linguistics). SLPAT also organises its own annual workshop, but with a wider focus that includes speech technologies and user interfaces. ATS-MA (Automatic Text Simplification — Methods and Applications in a Multilingual Society) is a new workshop with an emphasis on multilingual issues. Through these initiatives, a community is beginning to emerge, which bodes well. Given the applications to assistive technologies, and the need for more rigorous evaluations, it is important to bring a more interdisciplinary approach to the field. Specific questions that need addressing are:

1. *How good does automatic text simplification need to be?* The output of machine translation systems tend to be read by fluent readers of the target language. Thus, even errorful and ungrammatical translations can be understood, and

found useful, by readers who do not know the source language. This is not the case with text simplification. The typical target reader of a text simplification system has poor reading skills. Thus, bad system output might be unusable, even when it could be understood by a fluent reader. Automatic evaluations of text simplification systems, or even evaluations of fluency and correctness using fluent readers, are difficult to interpret until this basic question is answered.

2. *How good are simplified language resources?* Drawing parallels with translation studies, text is usually translated by someone who is a native speaker of the target language and has a good understanding of the source language. But who should (and indeed, does) simplify text? Assuming that there is no such thing as a native "Simplified English" speaker, we need to better understand the quality of resources such as Simple English Wikipedia, which have been used to train many SMT based simplification systems.
3. *How should text simplification systems be evaluated?* As discussed earlier, there have been few user studies to date that evaluate text simplification systems with real users. Evaluations of fluency and correctness have been on a small scale (as few as 20 sentences for recent papers). There has been no evaluation of hand-crafted systems versus machine translation inspired systems that analyses the strengths or weaknesses of either. It is therefore still not clear how good text simplification systems really are, or need to be to be useful.
4. *How easy are text simplification systems to adapt for particular users?* Different target populations have different simplification needs (see Section 2.3). Many systems have been developed with particular target populations in mind (see Section 3.4.1). However, the systems described in Section 3.2 are intended as general purpose systems. It is unclear whether these systems can be adapted for particular users, and what costs would be involved.

To summarise, following early work on text simplification that focused on the use of a small set of syntactic rules to simplify English, research has diversified in many directions. There are now groups working on a variety of languages, using a variety of frameworks, and expanding their coverage and accuracy of simplification operations. However, recent research has focused on sentence-level simplification, and tended to ignore issues raised in early work about the effect of text simplification on discourse structure and text coherence. More systematic evaluations are needed that measure text comprehension by end users. Specifically, the field needs to better understand how good text simplification needs to get before it is can be considered useful. For all the progress that has been made, it is clear that there is some distance to travel yet.

## References


Abrahamsson, E., Forni, T., Skeppstedt, M., & Kvist, M.. (2014). Medical text simplification using synonym replacement: Adapting assessment of word difficulty to a compounding language. *Proceedings of the Third Workshop on Predicting and Improving Text Readability for target reader populations*.

Aluísio, S. M., Specia, L., Pardo, T. A., Maziero, E. G., & Fortes, R. P. (2008). Towards Brazilian Portuguese automatic text simplification systems. *Proceedings of the Eighth ACM Symposium on Document Engineering*.

Angrosh, M., Nomoto, T., & Siddharthan, A. (2014). Lexico-syntactic text simplification and compression with typed dependencies. *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*.

Anderson, R. (1981). A proposal to continue a center for the study of reading. Technical Report 487, University of Illinois, Center for the Study of Reading, Urbana-Champaign.

Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In John Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anderson, R. C., & Davison, A. (1988). Conceptual and empirical bases of readability formulas. In Alice Davison & G. M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Aranzabe, M. J., Daz de Ilarraza, A., &Gonzalez-Dios, I. (2012). First approach to automatic text simplification in Basque. *Proceedings of the Natural Language Processing for Improving Textual Accessibility (NLP4ITA) Workshop (LREC 2012)*, Istanbul, Turkey.

Baldwin, B. (1997). CogNIAC: High precision coreference with limited knowledge and linguistic resources. *Proceedings of the ACL Workshop on Operational Factors in Practical, Robust Anaphora resolution for Unrestricted Texts*.

Bangalore, S., & Joshi, A. K. (1999). Supertagging: An approach to almost parsing. *Computational Linguistics*, 25(2), 237–265.

Bard, E., Robertson, D., & Sorace, A. (1996). Magnitude estimation for linguistic acceptability. *Language*, 72(1), 32–68. DOI: 10.2307/416793

Barlacchi, G., & Tonelli, S. (2013). ERNESTA: A sentence simplification tool for children's stories in Italian. *Computational Linguistics and Intelligent Text Processing*, (pp. 476–487). Springer. DOI: 10.1007/978-3-642-37256-8_39

Beck, I. L., McKeown, M. G., Omanson, R. C., & Pople, M. T. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly*, 19(3), 263–277. DOI: 10.2307/747821

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly*, 251–276. DOI: 10.2307/747763

Behaghel, O. (1930). Von deutscher Wortstellung [On German word order]. *Zeitschrift für Deutschkunde, Jargang 44 der Zeitschrift für deutschen Unterricht*, 81(9).

Bernth, A. (1998). EasyEnglish: Preprocessing for MT. *Proceedings of the Second International Workshop on Controlled Language Applications*.

Biran, O., Brody, S., & Elhadad, N. (2011). Putting it simply: A context-aware approach to lexical simplification. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

Bott, S., Saggion, H., & Mille, S. (2012). Text simplification tools for Spanish. *LREC*.

Briscoe, T., & Carroll, J. (1995). Developing and evaluating a probabilistic LR parser of part-of-speech and punctuation labels. *Proceedings of the ACL/SIGPARSE 4th International Workshop on Parsing Technologies*.

Brodsky, P., Waterfall, H., & Edelman, S. (2007). Characterizing motherese: On the computational structure of child-directed language. *Proceedings of the 29th Cognitive Science Society Conference*, ed. DS McNamara & JG Trafton.

Brouwers, L., Bernhard, D., Ligozat, A.-L., & François, T. (2014). Syntactic sentence simplification for French. *Proceedings of the Third Workshop on Predicting and Improving Text Readability for Target Reader Populations*.

Candido Jr, A., Maziero, E., Gasperin, C., Pardo, T. A., Specia, L., & Aluísio, S. M. (2009). Supporting the adaptation of texts for poor literacy readers: A text simplification editor for Brazilian Portuguese. *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*.

Canning, Y (2002). Syntactic simplification of Text. Ph. D. thesis, University of Sunderland, UK.

Canning, Y., Tait, J., Archibald, J., & Crawley, R. (2000). Replacing Anaphora for Readers with Acquired Dyslexia. *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC'00)*.

Caplan, D. (1992). *Language: Structure, processing, and disorders*. Cambridge, Massachusetts: MIT Press.

Carpenter, P., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In Morton Ann Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1075–1122). New York: Academic Press.

Carroll, J., Minnen, G., Canning, Y., Devlin, S., & Tait, J. (1998). Practical simplification of English newspaper text to assist aphasic readers. *Proceedings of AAAI98 Workshop on Integrating Artificial Intelligence and Assistive Technology*.

Chandrasekar, R., Doran, C., & Srinivas, B. (1996). Motivations and methods for text simplification. *Proceedings of the 16th International Conference on Computational Linguistics (COLING '96)*.

Chandrasekar, R., & Srinivas, B. (1997). Automatic induction of rules for text simplification. *Knowledge-Based Systems*, 10, 183–190. DOI: 10.1016/S0950-7051(97)00029-4

Chung, J.-W., Min, H.-J., Kim, J., & Park, J. C. (2013). Enhancing readability of web documents by text augmentation for deaf people. *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics*.

Clark, H., & Clark, E. (1968). Semantic distinctions and memory for complex sentences. *The Quarterly Journal of Experimental Psychology*, 20(2), 129–138. DOI: 10.1080/14640746808400141

Cohn, T., & Lapata, M. (2009). Sentence compression as tree transduction. *Journal of Artificial Intelligence Research*, 34(1), 637–674.

Conroy, J. M., & Schlesinger, J. D. (2004). Left-brain/Right-brain multi-document summarization. *Proceedings of the 4th Document Understanding Conference (DUC 2004) at HLT/NAACL 2004*, Boston, MA.

Coster, W., & Kauchak, D. (2011). Learning to simplify sentences using wikipedia. *Proceedings of the Workshop on Monolingual Text-To-Text Generation*.

Cross, T. G. (1977). *Mothers' speech adjustments: The contribution of selected child listener variables*. Talking to children: Language input and acquisition, 151–188.

Crossley, S. A., Louwerse, M. M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, 91(1), 15–30. DOI: 10.1111/j.1540-4781.2007.00507.x

Daelemans, W., Höthker, A., & Sang, E. F. T. K. (2004). Automatic sentence simplification for subtitling in Dutch and English. *LREC*.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466. DOI: 10.1016/S0022-5371(80)90312-6

Davies, A., & Widdowson, H. (1974). The teaching of reading and writing. *Techniques in Applied Linguistics*, 3.

Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly*.

De Belder, J., & Moens, M.-F. (2010). Text simplification for children. *Proceedings of the SIGIR Workshop on Accessible Search Systems*.

Deléger, L., Cartoni, B., & Zweigenbaum, P. (2013). Paraphrase detection in monolingual specialized/lay comparable corpora. *Building and Using Comparable Corpora*, (pp. 223–241). Springer. DOI: 10.1007/978-3-642-20128-8_12

Deléger, L., & Zweigenbaum, P. (2009). Extracting lay paraphrases of specialized expressions from monolingual comparable medical corpora. *Proceedings of the 2nd Workshop on Building and Using Comparable Corpora: From Parallel to Non-parallel Corpora*.

Deléger, L., & Zweigenbaum, P. (2010). Identifying paraphrases between technical and lay Corpora. *LREC*.

Devlin, S., & Tait, J. (1998). The use of a psycholinguistic database in the simplification of text for aphasic readers. In J. Nerbonne (Ed.), *Linguistic Databases* (pp. 161–173). Stanford, California: CSLI Publications.

Dras, M. (1999). Tree adjoining grammar and the reluctant paraphrasing of text. Ph. D. thesis, Macquarie University NSW 2109 Australia.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*, Vol. 373. Springer.

Elhadad, N. (2006). Comprehending technical texts: Predicting and defining unfamiliar terms. *Proceedings AMIA Annual Symposium*, 2006.

Evans, R., Orasan, C., & Dornescu, I. (2014). An evaluation of syntactic simplification rules for people with autism. *Proceedings of the Third Workshop on Predicting and Improving Text Readability for Target Reader Populations*.

Flesch, R. (1951). *How to test readability*. New York: Harper and Brothers.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, 50(3), 259–281. DOI: 10.1006/brln.1995.1048

Galley, M., & McKeown, K. (2007). Lexicalized Markov Grammars for Sentence Compression. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*.

Gardner, D., & Hansen, E. C. (2007). Effects of lexical simplification during unaided reading of English informational texts. *TESL Reporter*, 40(2), 27–59.

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, 2(2), 177–192. DOI: 10.1017/S0047404500000701

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language*, 11(01), 43–79. DOI: 10.1017/S0305000900005584

Green, G. M., & Olsen, M. S. (1986). *Preferences for and comprehension of original and readability-adapted materials*. Technical report, Champaign, Ill.: University of Illinois at Urbana-Champaign, Center for the Study of Reading.

Grosz, B., Joshi, A., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2), 203–226.

Hayes, D. P., & Ahrens, M. (1988). Vocabulary simplification for children: A special case of motherese. *Journal of Child Language*, 15(2), 395–410. DOI: 10.1017/S0305000900012411

Heilman, M., & Smith, N. A. (2010). Extracting simplified statements for factual question generation. *Proceedings of QG2010: The Third Workshop on Question Generation*.

Honeyfield, J. (1977). Simplification. *TESOL Quarterly*, 431–440. DOI: 10.2307/3585739

Inui, K., Fujita, A., Takahashi, T., Iida, R., & Iwakura, T. (2003). Text simplification for reading assistance: A project note. *Proceedings of the Second International Workshop on Paraphrasing-Volume 16*.

Irwin, J. (1980). The effects of explicitness and clause order on the comprehension of reversible causal relationships. *Reading Research Quarterly*, 15(4), 477–488. DOI: 10.2307/747275

Jauhar, S. K., & Specia, L. (2012). UOW-SHEF: SimpLex–lexical simplicity ranking based on contextual and psycholinguistic features. *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*.

Jonnalagadda, S., Tari, L., Hakenberg, J., Baral, C., & Gonzalez, G. (2009). Towards effective sentence simplification for automatic processing of biomedical text. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*.

Kaji, N., Kawahara, D., Kurohash, S., & Sato, S. (2002). Verb paraphrase based on case frame alignment. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*.

Kamalski, J., Sanders, T., & Lentz, L. (2008). Coherence marking, prior knowledge, and comprehension of informative and persuasive texts: Sorting things out. *Discourse Processes*, 45(4), 323–345. DOI: 10.1080/01638530802145486

Kandula, S., Curtis, D., & Zeng-Treitler, Q. (2010). A semantic and syntactic text simplification tool for health content. *Proceedings of AMIA Annual Symposium*, 2010.

Katz, E., & Brent, S. (1968). Understanding connectives. *Journal of Verbal Learning & Verbal Behavior*.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 115–126. DOI: 10.1016/S0022-5371(84)90082-3

Kelly, L. (1996). The interaction of syntactic competence and vocabulary during reading by deaf students. *Journal of Deaf Studies and Deaf Education*, 1(1), 75–90. DOI: 10.1093/oxfordjournals.deafed.a014283

Klebanov, B. B., Knight, K., & Marcu, D. (2004). Text simplification for information-seeking applications. *On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE*, (pp. 735–747). Springer. DOI: 10.1007/978-3-540-30468-5_47

Klein, D., & Manning, C. (2003). Accurate unlexicalized parsing. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*.

Klerke, S., & Søgaard, A. (2013). Simple, readable sub-sentences. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop*.

Knight, K., & Marcu, D. (2000). Statistics-based summarization — step one: Sentence compression. *Proceedings of The American Association for Artificial Intelligence Conference (AAAI-2000)*.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., et al. (2007). Moses: Open source toolkit for statistical machine translation. *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*, Vol. 1. London: Longman.

L'Allier, J. (1980). An evaluation study of a computer-based lesson that adjusts reading level by monitoring on task reader characteristics. Ph. D. thesis, University of Minnesota, Minneapolis, MN.

Langkilde-Geary, I. (2002). An empirical verification of coverage and correctness for a general-purpose sentence generator. *Proceedings of the 12th International Natural Language Generation Workshop*.

Lappin, S., & Leass, H. (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4), 535–561.

Leikin, M. (2002). Processing syntactic functions of words in normal and dyslexic readers. *Journal of Psycholinguistic Research*, 31(2), 145–163. DOI: 10.1023/A:1014926900931

Levy, E. T. (2003). The roots of coherence in discourse. *Human Development*, 169–88. DOI: 10.1159/000070367

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.

Lillo-Martin, D., Hanson, V., & Smith, S. (1991). Deaf readers' comprehension of complex syntactic structure. *Advances in Cognition, Education, and Deafness*, 146–151.

Linderholm, T., Everson, M., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more-and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction*, 18(4), 525–556. DOI: 10.1207/S1532690XCI1804_4

Lombardi, L., & Potter, M. (1992). The regeneration of syntax in short term memory* 1. *Journal of Memory and Language*, 31(6), 713–733. DOI: 10.1016/0749-596X(92)90036-W

Long, M. H., & Ross, S. (1993). *Modifications that preserve language and content*. Technical report, ERIC.

Louis, A., & Nenkova, A. (2013). What makes writing great? First experiments on article quality prediction in the Science Journalism Domain. *Transactions of the Association for Computational Linguistics*, 1, 341–352.

Lozanova, S., Stoyanova, I., Leseva, S., Koeva, S., & Savtchev, B. (2013). Text modification for Bulgarian sign language users. *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*.

Luckner, J. L., & Handley, C. M. (2008). A summary of the reading comprehension research undertaken with students who are deaf or hard of hearing. *American Annals of the Deaf*, 153(1), 6–36. DOI: 10.1353/aad.0.0006

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Towards a functional theory of text organization. *Text*, 8(3), 243–281.

Marschark, M., & Harris, M. (1996). Success and failure in learning to read: The special case of deaf children. *Reading comprehension difficulties: Processes and intervention*, 279–300.

Marschark, M., & Spencer, P. E. (2010). *The Oxford handbook of deaf studies, language, and education*, Vol. 2. Oxford University Press.

Mason, J., & Kendall, J. (1979). Facilitating reading comprehension through text structure manipulation. *Alberta Journal of Medical Psychology*, 24, 68–76.

Mason, R. A., & Just, M. A. (2004). How the brain processes causal inferences in text: A theoretical account of generation and integration component processes utilizing both cerebral hemispheres. *Psychological Science*, 15(1), 1–7. DOI: 10.1111/j.0963-7214.2004.01501001.x

McNamara, D., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43. DOI: 10.1207/s1532690xci1401_1

Miller, G. A., Beckwith, R., Fellbaum, C. D., Gross, D., & Miller, K. (1993). *Five papers on wordnet*. Technical report. Princeton, N. J.: Princeton University.

Miwa, M., Saetre, R., Miyao, Y., & Tsujii, J. (2010). Entity-focused sentence simplification for relation extraction. *Proceedings of the 23rd International Conference on Computational Linguistics*.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, 26(4), 453–465. DOI: 10.1016/0749-596X(87)90101-X

Narayan, S., & Gardent, C. (2014). Hybrid simplification using deep semantics and machine translation. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL'14)*.

Noordman, L. G. M., & Vonk, W. (1992). Reader's knowledge and the control of inferences in reading. *Language and Cognitive Processes*, 7, 373–391. DOI: 10.1080/01690969208409392

O'Brien, S. (2003). Controlling controlled English. An analysis of several controlled language rule sets. *Proceedings of EAMT-CLAW*, 3, 105–114.

Oh, S.-Y. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly*, 35(1), 69–96. DOI: 10.2307/3587860

Ong, E., Damay, J., Lojico, G., Lu, K., & Tarantan, D. (2007). Simplifying text in medical literature. *Journal of Research in Science Computing and English*, 4(1), 37–47.

Papoušek, M., Papoušek, H., & Haekel, M. (1987). Didactic adjustments in fathers' and mothers' speech to their 3-month-old infants. *Journal of Psycholinguistic Research*, 16(5), 491–516. DOI: 10.1007/BF01073274

Parr, S. (1993). Aphasia and literacy. Ph. D. thesis, University of Central England.

Peng, Y., Tudor, C. O., Torii, M., Wu, C. H., & Vijay-Shanker, K. (2012). iSimp: A sentence simplification system for biomedical text. *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on*.

Petersen, S. (2007). Natural language processing tools for reading level assessment and text simplification for bilingual education. Ph. D. thesis, University of Washington, Seattle, WA.

Pitler, E., & Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Potter, M., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences* 1. *Journal of Memory and Language*, 29(6), 633–654. DOI: 10.1016/0749-596X(90)90042-X

Power, R. (2012). OWL simplified English: A finite-state language for ontology editing. *Proceedings of the Third International Workshop on Controlled Natural Language*.

Quigley, S. P., & Paul, P. V. (1984). *Language and deafness*. San Diego, California: College-Hill Press.

Quigley, S. P., Power, D., & Steinkamp, M. (1977). The language structure of deaf children. *The Volta Review*, 79(2), 73–84.

Quinlan, P. (1992). *The Oxford psycholinguistic database*. U. K: Oxford University Press.

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, 13(2), 212–218. DOI: 10.1016/S0959-4388(03)00035-7

Rello, L., Baeza-Yates, R., Dempere, L., & Saggion, H. (2013). Frequent words improve readability and short words improve understandability for people with dyslexia. *Proceedings of INTERACT*, Vol. 13.

Rello, L., Bayarri, C., Gòrriz, A., Baeza-Yates, R., Gupta, S., Kanvinde, G., Saggion, H., Bott, S., Carlini, R., & Topac, V. (2013b). DysWebxia 2.0!: More accessible text for people with dyslexia. *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*.

Riezler, S., King, T. H., Crouch, R., & Zaenen, A. (2003). Statistical sentence condensation using ambiguity packing and stochastic disambiguation methods for lexical-functional grammar. *Proceedings of the Human Language Technology Conference and the 3rd Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL'03)*.

Robbins, N. L., & Hatcher, C. (1981). The effects of syntax on the reading comprehension of hearing-impaired children. *The Volta Review : journal of the Alexander Graham Bell Association for the Deaf*, 83, 105–115.

Robertson, E. K., & Joanisse, M. F. (2010). Spoken sentence comprehension in children with dyslexia and language impairment: The roles of syntax and working memory. *Applied Psycholinguistics*, 31(1), 141–165. DOI: 10.1017/S0142716409990208

Seretan, V. (2012). Acquisition of syntactic simplification rules for French. *LREC*.

Shewan, C., & Canter, G. (1971). Effects of vocabulary, syntax and sentence length on auditory comprehension in aphasic patients. *Cortex*, 7, 209–226. DOI: 10.1016/S0010-9452(71)80001-1

Siddharthan, A. (2003a). Syntactic simplification and text cohesion. Ph. D. thesis, University of Cambridge, UK.

Siddharthan, A. (2010). Complex lexico-syntactic reformulation of sentences using typed dependency representations. *Proceedings of the 6th International Natural Language Generation Conference (INLG 2010)*.

Siddharthan, A. (2003b). Preserving discourse structure when simplifying text. *Proceedings of the European Natural Language Generation Workshop (ENLG), 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'03)*.

Siddharthan, A. (2003c). Resolving pronouns robustly: Plumbing the depths of shallowness. *Proceedings of the Workshop on Computational Treatments of Anaphora, 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'03)*.

Siddharthan, A. (2006). Syntactic simplification and text cohesion. *Research on Language and Computation*, 4(1), 77–109. DOI: 10.1007/s11168-006-9011-1

Siddharthan, A. (2011). Text simplification using typed dependencies: A comparison of the robustness of different generation strategies. *Proceedings of the 13th European Workshop on Natural Language Generation*.

Siddharthan, A., & Angrosh, M. (2014). Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL'14)*.

Siddharthan, A., & Katsos, N. (2012). Offline sentence processing measures for testing readability with users. *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations*.

Siddharthan, A., Nenkova, A., & McKeown, K. (2011). Information status distinctions and referring expressions: An empirical study of references to people in news summaries. *Computational Linguistics*, 37(4), 811–842. DOI: 10.1162/COLI_a_00077

Smith, C., & Jönsson, A. (2011). Automatic summarization as means of simplifying texts, an evaluation for swedish. *Proceedings of the 18th Nordic Conference of Computational Linguistics (NoDaLiDa-2010)*, Riga, Latvia.

Smith, D. A., & Eisner, J. (2006). Quasi-synchronous grammars: Alignment by soft projection of syntactic dependencies. *Proceedings of the Workshop on Statistical Machine Translation*.

Specia, L. (2010). Translating from complex to simplified sentences. *Proceedings of the Conference on Computational Processing of the Portuguese Language*.

Specia, L., Jauhar, S. K., & Mihalcea, R. (2012). Semeval-2012 task 1: English lexical simplification. *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*.

Srinivas, B. (1997). Complexity of lexical descriptions and its relevance to partial parsing. Ph. D. thesis, University of Pennsylvania, Philadelphia, PA.

Sripada, S., Reiter, E., & Davy, I. (2003). SumTime-Mousam: Configurable marine weather forecast generator. *Expert Update*, 6(3), 4–10.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*.

Tweissi, A. I. (1998). The effects of the amount and type of simplification on foreign language reading comprehension. *Reading in a Foreign Language*, 11(2), 191–204.

Vanderwende, L., Suzuki, H., Brockett, C., & Nenkova, A. (2007). Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion. *Information Processing & Management*, 43(6), 1606–1618. DOI: 10.1016/j.ipm.2007.01.023

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & 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. DOI: 10.1046/j.0021-9630.2003.00305.x

Vogel, S. A. (1974). Syntactic abilities in normal and dyslexic children. *Journal of Learning Disabilities*, 7(2), 103–109. DOI: 10.1177/002221947400700211

Walker, A., Siddharthan, A., & Starkey, A. (2011). Investigation into human preference between common and unambiguous lexical substitutions. *Proceedings of the 13th European Workshop on Natural Language Generation*.

Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, 9(1), 81–105. DOI: 10.1017/S0954394500001800

Watanabe, W. M., Junior, A. C., Uzêda, V. R., Fortes, R. P. d. M, Pardo, T. A. S., & Aluísio, S. M. (2009). Facilita: Reading assistance for low-literacy readers. *Proceedings of the 27th ACM international conference on Design of communication*.

Williams, S., Reiter, E., & Osman, L. (2003). Experiments with discourse-level choices and readability. *Proceedings of the European Natural Language Generation Workshop (ENLG), 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'03)*.

Wojcik, R., & Hoard, J. (1996). Controlled languages in industry. In Ronald Cole (Ed.), *Survey of the state of the art in Human Language Technology* (pp. 274–276). http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html.

Wojcik, R., Hoard, J., & Holzhauser, K. (1990). The boeing simplified English checker. *Proceedings of the International Conference in Human Machine Interaction and Artificial Intelligence in Aeronautics and Space*.

Woodsend, K., & Lapata, M. (2011). Learning to simplify sentences with quasi-synchronous grammar and integer programming. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Wubben, S., vanden Bosch, A., & Krahmer, E. (2012). Sentence simplification by monolingual machine translation. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*.

Yamada, K., & Knight, K. (2001). A syntax-based statistical translation model. *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*.

Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, 44(2), 189–219. DOI: 10.1111/j.1467-1770.1994.tb01100.x

Yatskar, M., Pang, B., Danescu-Niculescu-Mizil, C., & Lee, L. (2010). For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Zeng-Treitler, Q., Goryachev, S., Kim, H., Keselman, A., & Rosendale, D. (2007). Making texts in electronic health records comprehensible to consumers: A prototype translator. *Proceedings of AMIA Annual Symposium*, Vol. 2007.

Zeng-Treitler, Q., Goryachev, S., Tse, T., Keselman, A., & Boxwala, A. (2008). Estimating consumer familiarity with health terminology: A context-based approach. *Journal of the American Medical Informatics Association*, 15(3), 349–356. DOI: 10.1197/jamia.M2592

Zhu, Z., Bernhard, D., & Gurevych, I. (2010). A monolingual tree-based translation model for sentence simplification. *Proceedings of the 23rd International Conference on Computational Linguistics*.

## *Author's address*

Advaith Siddharthan
Department of Computing Science
University of Aberdeen
King's College
Aberdeen AB24 3UE
United Kingdom

advaith@abdn.ac.uk