



**HAL**  
open science

## Assisted Lexical Simplification for French Native Children with Reading Difficulties

Firas Hmida, Mokhtar Boumedylen Billami, Thomas François, Nuria Gala

► **To cite this version:**

Firas Hmida, Mokhtar Boumedylen Billami, Thomas François, Nuria Gala. Assisted Lexical Simplification for French Native Children with Reading Difficulties. The Workshop of Automatic Text Adaptation, 11th International Conference on Natural Language Generation, 2018, Tilbourg, Netherlands. hal-02077299

**HAL Id: hal-02077299**

**<https://hal.science/hal-02077299>**

Submitted on 22 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Assisted Lexical Simplification for French Native Children with Reading Difficulties

**Firas Hmida**<sup>1</sup> Aix-Marseille Univ.  
LPL UMR 7309  
Aix-en-Pce, France

**Mokhtar B. Billami**<sup>1</sup> Aix-Marseille Univ.  
LIS UMR 7020  
Marseille, France

**Thomas François**<sup>2</sup> UCLouvain  
CENTAL/IL&C  
Louvain-la-Neuve, Belgium

**Núria Gala**<sup>1</sup> Aix-Marseille Univ.  
LPL UMR 7309  
Aix-en-Pce, France

firstname.lastname@{<sup>1</sup>univ-amu.fr,<sup>2</sup>uclouvain.be}

## Abstract

For poor-readers and dyslexic children, reading is often a pitfall to social integration and academic progress. The school support of these children usually requires adapted texts, specialised glossaries and dedicated management tools. In this paper, we propose a method which exploits French lexical resources to automatically simplify words in order to provide adapted texts. Despite the difficulty of the task, the conducted evaluations show that the proposed methodology yields better results than the state of the art word2vec techniques for lexical simplification.

## 1 Introduction

Learning to read is a complex and lengthy process leading to a fundamental skill which is crucial for academic, professional and personal success. Yet, according to the Progress in International Reading Literacy (PIRLS<sup>1</sup>) 2001, the overall performances of French young readers is gradually decreasing from evaluation to evaluation: 39% of the students are in difficulty at the end of primary school according to the study carried by the Cycle of Disciplinary Evaluations Performed on Samples<sup>2</sup>. Statistically, every year, 2 to 5 children in a classroom present a specific language impairment (from poor-reading to dyslexia, with a large variability). Ziegler et al. (2003) show that the problems of comprehension among children with reading difficulties are mostly due to the difficulties in decoding words in order to recognise them. In other words, these children do not suffer from

oral comprehension problems. However, when it comes to reading a text, it turns out that all their efforts are so focused on decoding that they exhaust their cognitive capacity. Out of hand, they get tired, give up reading and lose the meaning of what they have already read.

In this context, scholars have found it valuable to control the reading difficulty of pedagogical materials using readability formulae (DuBay, 2004). Text readability can be defined as the ease with which a reader can read and understand a text. Readability assessment techniques enable a better association between texts and readers, which tends to increase the benefits of reading practices. However, even if readability formulae are useful to find appropriate texts for a given level of reading proficiency, they do not allow to adapt a given text to a specific reader, as is generally needed for poor readers or readers with dyslexia.

More recently, Natural Language Processing (NLP) techniques have allowed the development of more efficient tools to support reading. Among them are advanced readability models (Collins-Thompson, 2014) that automatically assess the readability of a text from a larger number of text characteristics. Another promising area is automated text simplification (ATS), which aims to automatically substitute complex linguistic phenomena in texts by simpler equivalents while keeping the meaning preserved as much as possible. ATS is generally described as involving two sub-tasks (Saggion, 2017): syntactic simplification and lexical simplification (LS). In this paper, we will be concerned with the second one, because, as far as poor and dyslexic readers are concerned, automatic lexical simplification is a first and crucial step in order to simplify a text for this population.

As it has been highlighted in the literature, long and less frequent words are especially difficult for poor readers (Ziegler et al., 2003; Spinelli et al.,

<sup>1</sup>[https://timssandpirls.bc.edu/pirls2001i/pdf/pl\\_IR\\_book.pdf](https://timssandpirls.bc.edu/pirls2001i/pdf/pl_IR_book.pdf)

<sup>2</sup>Cycle des Évaluations Disciplinaires Réalisées sur Échantillons (CEDRE).

2005). Gala and Ziegler (2016) also identified that, for French children with dyslexia, inconsistent words as far as the grapheme-phoneme relation is concerned (different length of the number of letters and phonemes in a word) contribute to the difficulty in reading.

In this paper, we address the challenging task of LS, which has not yet been systematically investigated for French. We compare two approaches: the first one is based on the exploitation of a lexical resource, ReSyf<sup>3</sup> (Billami et al., 2018), which contains disambiguated ranked synonyms in French; the second one is based on word embedding and draws from Glavaš and Štajner (2015). Although previous studies have prioritised statistical methods over the use of resources to acquire synonyms, we are not aware of a previous study having compared statistical models with a **disambiguated** synonym resource. We believe that this property could significantly enhance the selection of relevant candidates for substitution in a given context. Another property of ReSyf is that synonyms have already been ranked by reading difficulty using Billami et al. (2018) method.

The paper is organised as follows: Section 2 presents existing methods for LS. Section 3 describes our method and Section 4 discusses on the results. Some concluding remarks are to be found at the end, along with future work.

## 2 State of the Art

Text simplification refers to the process of transforming a text into an equivalent which is easier to read and to understand than the original, while preserving, in essence, the original content (Bott et al., 2012). Lexical simplification (LS) is dedicated to the substitution of complex words by simpler synonyms. Complex words are here considered as mono-lexical units which are difficult to read (i.e. decode), especially for poor and dyslexic readers. LS aims to provide, for a complex word in a text, a simpler substitute making this text more accessible to the reader (the meaning and the syntactic structure of the text is as far as possible preserved).

Previous works have shown the contribution of LS to make texts more accessible to different audiences: people with dyslexia (Rello et al., 2013a,c,b), with aphasia (Carroll et al., 1998; Devlin, 1999), illiterate and poor readers (Aluísio

et al., 2010) to name a few. Text simplification systems exist for various languages, for example: English (Carroll et al., 1998; Horn et al., 2014; Glavaš and Štajner, 2015; Paetzold and Specia, 2017), Spanish (Bott et al., 2012; Rello et al., 2013a), Swedish (Keskišärkkä, 2012), and Portuguese (Aluísio and Gasperin, 2010). However, to our knowledge, there is no full-fledged ATS system for French available, although some authors have investigated related aspects (i.e. simplified writing for language-impaired readers (Max, 2006), French readability for French as a Foreign Language (FFL) (François and Fairon, 2012), syntactic simplification (Seretan, 2012; Brouwers et al., 2014), and lexical simplification for improving the understanding of medical terms (Grabar et al., 2018)).

While first LS systems (Carroll et al., 1998; Devlin, 1999) used to combine WordNet (Miller, 1998) and frequency information from words, more recent ones are more sophisticated and rely on supervised machine learning methods. Their architecture can be represented with four steps as follows (Shardlow, 2014):

1. Complex Word Identification (CWI): aims to identify target words that need simplification. In CWI, the methods based on large corpora and thesaurus dominate the top 10 in SemEval 2016 (Paetzold and Specia, 2016). In most recent CWI shared task took place in June 2018, word length and word frequency based features lead to the best results.
2. Substitution Generation (SG), to provide candidate substitution for each complex identified word. In SG, Horn et al. (2014) obtain candidates from a parallel corpus contains Wikipedia and simplified version of Wikipedia yielding major step against earlier approaches (Devlin, 1999). Glavaš and Štajner (2015) use word embeddings models to generate candidate substitution leading to even better results.
3. Substitution Selection (SS), to filter out context candidates. Generally, SS first requires all the disambiguation of the candidates provided by the previous step. Only candidates matching the Part-Of-Speech (POS) of the target complex word are retained.
4. Substitute ranking (SR), to sort the retained

<sup>3</sup><https://cental.uclouvain.be/resyf/>

candidates according to their complexity. In SR, the performance of the state of the art is achieved by the supervised methods: SVM-Rank (Horn et al., 2014) and Boundary Ranking (Paetzold and Specia, 2016). Supervised methods have the caveat of requiring annotated data, nonetheless as consequence they can be adapted according to the target audience.

In practice, this process is not literally respected in LS methods. For some approaches (Biran et al., 2011; Bott et al., 2012; Glavaš and Štajner, 2015), all words are potentially complex and need simplification. Each word is replaced only if it has a simpler synonym. Some other methods merge the SS into the SR step.

Here, we consider the Glavaš and Štajner (2015) method as a baseline of the state of the art. This method is based on the exploitation of general resources in a general context. The baseline relies on word embeddings to generate substitute candidates. Glavaš and Štajner (2015) only replace a target word if it has a lower frequency than the selected candidate substitution. However, we propose a LS method that exploits a new lexical resource, ReSyf (cf. subsection 3.2) which aims to provide simpler substitutes to identified complex words in French, the overall idea being to adapt texts for children who have difficulties with basic reading and comprehension skills in early grades, and who have French as a mother tongue.

### 3 Lexical Simplification: our method

We present here our methodology to build the LS system. Figure 1 illustrates the architecture of the system that we detail in the next subsections. In brief, we start from a sentence in which complex words have been identified. We then use ReSyf to get candidates for substitution (section 3.2). If the complex word has several meanings, we use automatic word sense disambiguation to select the best set of candidates (section 3.3). The last step consists in selecting the simplest candidate to be used in the simplified sentence (section 3.4).

#### 3.1 Complex Word Identification (CWI)

In this work, we focus our interest on the generation of candidate substitutes in order to improve the quality of the simplification. We use a list of complex reference words that have beforehand been identified as complex by human experts. For

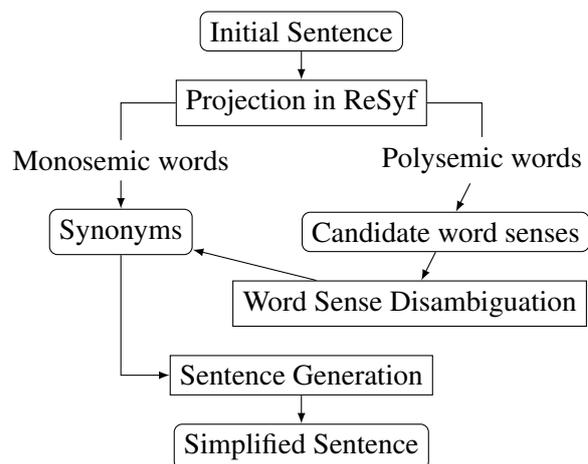


Figure 1: Architecture of the LS system

example, in the sentence *Le castor est un excellent nageur*<sup>4</sup>, the word *excellent* has been tagged as complex.

#### 3.2 Using ReSyf for Lexical Simplification (LS)

ReSyf<sup>5</sup> (Billami et al., 2018) is a lexical resource which includes a disambiguated set of synonyms that are ranked by difficulty. The ranking (order of appearance and weight in the vector) is calculated taking into account intra-lexical features (i.e. length of words, syllabic structure, morphological structure, number of orthographical neighbours, etc.), morpho-semantic features (i.e. number of morphemes, frequency of morphemes, polysemy, etc.) and psycholinguistic features (i.e. frequency index, etc.) (François et al., 2016). ReSyf contains more than 57 000 disambiguated lemmas initially extracted from JeuxDeMots<sup>6</sup>, a freely available lexical network containing fine-grained semantic information (Lafourcade, 2007).

In order to obtain our synonyms, for each input sentence, complex words are projected in ReSyf. The candidate substitutes are provided by JeuxDeMots and are classified according to the meanings of the complex words. For instance, table 1 shows the candidate substitutes of *excellent* that can be found in ReSyf.

For monosemic words, a list of candidate substitutes, ranked according to their complexity, is directly obtained from ReSyf. For polysemic words like *excellent*, a further disambiguation step

<sup>4</sup>The beaver is an excellent swimmer.

<sup>5</sup><https://cental.uclouvain.be/resyf/>

<sup>6</sup><http://www.jeuxdemots.org>

Sense	Substitutes
<i>Excellent (délicieux)</i>	<i>fin, fameux</i>
<i>Excellent (formidable)</i>	<i>bon, fort</i>

Table 1: Senses and substitutes in ReSyf for ‘excellent (delicious)’ (‘fine, well’) and ‘excellent (great)’ (‘good, strong’)

is applied to choose the most appropriate meaning matching the context of the sentence. ReSyf integrates the generation of candidate substitutes as well as their ranking according to the complexity measure by Billami et al. (2018).

### 3.3 Word Sense Disambiguation (WSD)

The simplification of *excellent* requires a disambiguation in order to choose a sense among *délicieux* and *formidable* (cf. table 1). To this aim, we apply an algorithm that uses semantic representations for words and word senses (called *semantic signatures*). These signatures have been created and validated by Billami and Gala (2017) and use the structural properties of JeuxDeMots. Furthermore, the ReSyf senses are the same as the senses from JeuxDeMots, encoded with the *semantic refinement* relation.

The semantic signatures that we have used integrate different relationships. Some of these relationships correspond to lexical functions related to the vocabulary itself (such as *associated idea* and *synonymy*) or to hierarchical semantic relations (such as *hypernymy* and *hyponymy*). The algorithm that we propose for complex word sense disambiguation is described below.

First, we initialize the score of each candidate word sense (lines 1–2). Second, we compare each candidate word sense with each word belonging to the context of the complex word to disambiguate by using semantic signatures (lines 3–5). Third, for this comparison of each pair (sense, word), we use among others another associative relation defined in JeuxDeMots, named *inhibition*. This relation allows to return, for a given target, the terms that tend to be excluded by this target (line 6). For example, the sense ‘excellent (great)’ excludes the words *delicious* and *tasty*. The score of the semantic similarity between a candidate word sense and a context word is computed only and only if there is no inhibition relation between them. This way of proceeding to the selection of the words of the context gives the advantage to the senses which

---

**Algorithm 1:** Disambiguate a target word by using semantic signatures of words and word senses

---

**Input:**

$t_w$ : target word to treat

$sem\_ref(t_w)$ : senses set for  $t_w$

$CXT(t_w)$ : words of the context of  $t_w$ , except this latter

**Result:**

$\hat{Sense}_{t_w}$ : sense of  $t_w$  with the highest score

**Data:**

$Rel_{Inhib}$ : set of pairs of terms whose elements of each pair are linked by the *inhibition relation* with a nonzero positive weight

$S(t)$ : a signature for a given term  $t$

**1 Initialization:**

2  $Score_{refs\_C} = \emptyset$  /\* associate each sense of  $t_w$  with its score. \*/

3 **for**  $sense_i \in sem\_ref(t_w)$  **do**

4      $Score(sense_i) = 0$ ;

5     **for**  $word_j \in CXT(t_w)$ , with

6          $j \in \{1, \dots, |CXT(t_w)|\}$  **do**

7         **if**  $(sense_i, word_j) \notin Rel_{Inhib}$

8             **then**

9              $Score(sense_i) = Score(sense_i) + Sim(S(sense_i), S(word_j))$ ;

10      $Score_{refs\_C} \leftarrow$

11      $Score_{refs\_C} \cup (sense_i, Score(sense_i))$ ;

12 **if**  $(|Best(Score_{refs\_C})| \geq 2)$  **then**

13      $\hat{Sense}_{t_w} \leftarrow$

14      $FIRST_S(t_w, Best(Score_{refs\_C}))$

15 **else**

16      $\hat{Sense}_{t_w} \leftarrow Best(Score_{refs\_C})$

---

exclude less words to have a more important score.

The semantic similarity (*i.e.*  $Sim$  in algorithm 1) that we use takes into account the relation between two lexical units to compare (Billami and Gala, 2017). If the relation exists between the two elements to compare, we have a perfect similarity else the cosine similarity is estimated by using their semantic signatures (line 7). Fourth, we calculate the score of similarity for each candidate word sense (line 8).

Besides, if there are at least two senses with the highest score, the sense returned by the algorithm is the one with the highest weight in ReSyf (*i.e.*  $FIRST_S$  in algorithm 1, lines 9–10). Otherwise,

the one and only best sense is returned (lines 11 – 12). In the previous example, the retained sense for *excellent* is *formidable*.

### 3.4 Sentence Generation

Once the complex word is disambiguated, the first three ranked synonyms in ReSyf, corresponding to that word-sense, are retained to generate simplified versions of the initial sentence. For instance, the previous sentence can be simplified *Le castor est un bon nageur*<sup>7</sup>.

## 4 Evaluation

Our aim is to assess the use of lexical resource such as ReSyf for the task of lexical simplification. We hypothesise that having already a disambiguated set of synonyms reduces the amount of noisy candidates created by statistical algorithms. To check this hypothesis, we have evaluated the quality of the substitutions produced by the two methods: the baseline based on Glavaš and Štajner (2015) and our method based on ReSyf, relying on the evaluation made by two experts.

### 4.1 Corpus of Evaluation

The corpus used for evaluation contains literary and scientific texts usually read in classrooms at primary levels (children aged 7 to 9 years old) in France. Within the 187 sentences of the corpus, experts have identified 190 complex words that have to be simplified (thus, we have an average of 1 complex word per sentence).

### 4.2 Word Embedding for SG

The generation of candidate for substitution by the baseline is based on a Word Embedding model<sup>8</sup> (Fauconnier, 2015) trained on the FrWaC corpus, a “1.6 billion word corpus constructed from the Web limiting the crawl to the .fr domain”<sup>9</sup> and using medium-frequency words as seeds. The corpus was POS-tagged and lemmatized with the Tree-Tagger<sup>10</sup>.

### 4.3 Human evaluation

In this paper, the simplifications have been evaluated according to their complexity and the context

where they appear. Three substitutes are proposed for every complex word. The provided simplifications were assessed by two evaluators following these instructions:

- The substitute must be simpler than the complex word
- The substitute must fit the context of the sentence
- If the complex word appears as a substitute, it is invalidated

The table 2 shows examples of the evaluation. *Beau* and *fort* do not match the context. We computed the inter-rater reliability of the human annotation. Even though selecting good candidates for simplification has been regarded as a complex task for human, we obtained a  $\kappa$  of 0.625 for the baseline model and a  $\kappa$  of 0.656 for the annotation of ReSyf’s results.

Based on this human annotation, we have computed two evaluation metrics. *Precision1* is a global precision. Every simplification is considered as an independent sentence. This measure aims to calculate the number of valid simplifications among all the provided ones (*i.e* simplified sentences). *Precision2* allows us to verify, for an initial sentence if, at least, one valid simplification appears among the three proposed ones. For each original sentence, only one valid simplification is counted, even if there are two or three valid ones. If none of the three simplifications are correct, then the count is 0. This measure counts the number of initial sentences that have at least one valid simplification.

$$Precision1 = \frac{\#valid\_simp}{\#all\_simplifications}$$

$$Precision2 = \frac{\#at\_least\_one\_valid\_simp}{\#all\_initial\_sentences}$$

By analysing the simplifications produced by the baseline and our method, table 3 shows that ReSyf provides better results than Word2Vec techniques. In table 3, Precision1 and Precision2 count valid simplifications only if both of the annotators agree on the proposed substitute. This table also shows that our method produces more suitable simplifications (16.3% and 51.9%) than the baseline (15.7% and 49.4%).

<sup>7</sup>The beaver is a good swimmer.

<sup>8</sup><http://fauconnier.github.io/#data>

<sup>9</sup><http://wacky.sslmit.unibo.it/doku.php?id=corpora>

<sup>10</sup>[www.cis.uni-muenchen.de/~schmid/tools/TreeTagger](http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger)

Initial Sentence	Simplified Sentence	Evaluation
Le castor est un <b>excellent</b> nageur	Le castor est un <b>beau</b> nageur	0
Le castor est un <b>excellent</b> nageur	Le castor est un <b>fort</b> nageur	0
Le castor est un <b>excellent</b> nageur	Le castor est un <b>bon</b> nageur	1

Table 2: Evaluation example for simplified versions of the sentence *Le castor est un excellent nageur*.

Methods	Precision1 (%)	Precision2 (%)
Baseline	15.7	49.4
LS_ReSyf	<b>16.3</b>	<b>51.9</b>

Table 3: LS evaluation result of annotators 1 and 2

#### 4.4 Discussion

The annotators have noticed that our method provides more linguistically-motivated substitutes than the word2vec method for LS. Indeed, SG from word2vec relies on the distance that separates word vectors. This distance is referred to the context and the general semantic distance that could yield not only appropriate synonyms, but also noise from other less suitable relations like antonymy. However, ReSyf is built from particular semantic relations from JeuxDeMots, especially synonymy and hypernymy which are more suitable for LS. Table 4 shows examples of obtained substitutes from ReSyf and word2vec methods for *marchandise* ('good') and *garçonnet* ('young child').

Complex Word	Word2Vec Substitute	ReSyf Substitute
Marchandise	transport, douane, commissionnaire	article, produit, marchandise
Garçonnet	filles, père, mère	enfant, garçon, petit

Table 4: Examples of provided substitutes by word2vec and ReSyf methods

Taking into account the difficulty of the task, the evaluation not surprisingly shows that the results are slightly better when the lexical resource ReSyf is used. Word2vec being based on the presence of a lexical unit in a same context, antonyms, wrong hyponyms or wrong senses are to be found among the wrong candidates for lexical substitutions. For instance, the antonym *fin* ('end') is proposed as a substitute for *début* ('beginning') or the sense 'lawyer' is proposed instead of 'bar' for the

polysemic French word *barreau*.

Wrong candidates found using ReSyf largely concern cases where multi-word expressions (MWE) are present in the text. As the MWE is currently not detected, we replaced one token of the expression by a simpler synonym that does not necessarily fits the context of the whole expression. For instance, in the expression *se prendre de sympathie* ('sympathize with'), we replaced *sympathie* by simpler candidates such as *intérêt* or *accord*. The second candidate clearly does not fit in the global expression, that would ideally need to be substitute as a whole, for example, by *s'occuper de*. ReSyf also proposes "complex words" such as: *long*, *animal* and *présent*. These words are not particularly considered as complex (as regards to their length, frequency or syllable structure). There is currently work in progress to better identify complex words as regards to the difficulties that poor-readers and dyslexic children may have.

## 5 Conclusion and perspectives

In this paper we have presented a method for lexical substitution that uses a lexical resource where French synonyms are disambiguated and ranked according to the difficulty to be read and understood. The results obtained after the evaluation show that using a lexical resource improves the results (51.9%) as regards to a state-of-the-art system based on word2vec (49.4%). The lexical resource that we have used provides already ranked synonyms, and once the system identifies the complex word to replace, our method is able to provide better candidates for the simplifications.

In future work, we plan (1) to evaluate our method with a greater amount of data and (2) to extend our work to automatically identify complex words in context (CWI). Our final aim is to

propose a whole methodology for a lexical simplification system that will adapt texts for French scholars facing difficulties learning to read.

## Acknowledgments

This work has been funded by the French Agence Nationale pour la Recherche, through the ALECTOR project (ANR-16-CE28-0005).

## References

- Sandra Maria Aluísio and Caroline Gasperin. 2010. Fostering digital inclusion and accessibility: the porsimples project for simplification of portuguese texts. In *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, pages 46–53. Association for Computational Linguistics.
- Sandra Maria Aluísio, Lucia Specia, Caroline Gasperin, and Carolina Scarton. 2010. Readability assessment for text simplification. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–9. Association for Computational Linguistics.
- Mokhtar B. Billami, Thomas François, and Núria Gala. 2018. Resyf: a French lexicon with ranked synonyms. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, pages 2570–2581.
- Mokhtar B. Billami and Núria Gala. 2017. Creating and validating semantic signatures : application for measuring semantic similarity and lexical substitution. In *The 24th edition of the French NLP conference*, Orleans, France.
- Or Biran, Samuel Brody, and Noémie Elhadad. 2011. Putting it simply: a context-aware approach to lexical simplification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 496–501. Association for Computational Linguistics.
- Stefan Bott, Luz Rello, Biljana Drndarevic, and Horacio Saggion. 2012. Can Spanish be simpler? lexis: Lexical simplification for Spanish. *Proceedings of COLING 2012*, pages 357–374.
- Laetitia Brouwers, Delphine Bernhard, Anne-Laure Ligozat, and Thomas François. 2014. Syntactic sentence simplification for French. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR) at EACL 2014*, page 47–56.
- John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10.
- Kevyn Collins-Thompson. 2014. Computational assessment of text readability: A survey of current and future research. *International Journal of Applied Linguistics*, 165(2):97–135.
- Siobhan Lucy Devlin. 1999. *Simplifying natural language for aphasic readers*. Ph.D. thesis, University of Sunderland.
- William H. DuBay. 2004. *The principles of readability*. Impact Information. Available on <http://www.nald.ca/library/research/readab/readab.pdf>.
- Jean-Philippe Fauconnier. 2015. French word embeddings. <http://fauconnier.github.io>.
- Thomas François, Mokhtar B. Billami, Núria Gala, and Delphine Bernhard. 2016. Bleu, contusion, ecchymose : tri automatique de synonymes en fonction de leur difficulté. In *Actes de la conférence Traitement Automatique des Langues Naturelles*, pages 15–28.
- Thomas François and Cédric Fairon. 2012. An “AI readability” formula for french as a foreign language. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, page 466–477.
- Núria Gala and Johannes Ziegler. 2016. Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC), COLING 2014*, pages 59–66.
- Goran Glavaš and Sanja Štajner. 2015. Simplifying lexical simplification: do we need simplified corpora? In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, volume 2, pages 63–68.
- Natalia Grabar, Emmanuel Farce, and Laurent Sparrow. 2018. Étude de la lisibilité des documents de santé avec des méthodes d’oculométrie. In *Proceedings of the Conférence Traitement Automatique des Langues Naturelles*.
- Colby Horn, Cathryn Manduca, and David Kauchak. 2014. Learning a lexical simplifier using wikipedia. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 458–463.
- Robin Keskisärkkä. 2012. Automatic text simplification via synonym replacement. Master’s thesis, Dept of Computer and Information Science at Linköping University, Sweden.

- Mathieu Lafourcade. 2007. Making people play for lexical acquisition. In *In Proc. SNLP 2007, 7th Symposium on Natural Language Processing*.
- Aurélien Max. 2006. Writing for language-impaired readers. In *Proc. of Computational Linguistics and Intelligent Text Processing: 7th International Conference, CICLing 2006*, page 567–570, Mexico City, Mexico.
- George Miller. 1998. *WordNet: An electronic lexical database*. MIT press.
- Gustavo Paetzold and Lucia Specia. 2016. Semeval 2016 task 11: Complex word identification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 560–569.
- Gustavo Paetzold and Lucia Specia. 2017. A survey on lexical simplification. *Journal of Artificial Intelligence Research*, 60:549–593.
- Luz Rello, Ricardo Baeza-Yates, Stefan Bott, and Horacio Saggion. 2013a. Simplify or help?: Text simplification strategies for people with dyslexia. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, page 15. ACM.
- Luz Rello, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013b. Frequent words improve readability and short words improve understandability for people with dyslexia. In *IFIP Conference on Human-Computer Interaction*, pages 203–219. Springer.
- Luz Rello, Clara Bayarri, Azuki Górriz, Ricardo Baeza-Yates, Saurabh Gupta, Gaurang Kanvinde, Horacio Saggion, Stefan Bott, Roberto Carlini, and Vasile Topac. 2013c. Dyswebxia 2.0!: more accessible text for people with dyslexia. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, page 25. ACM.
- Horacio Saggion. 2017. Automatic text simplification. *Synthesis Lectures on Human Language Technologies*, 10(1):1–137.
- V. Seretan. 2012. Acquisition of syntactic simplification rules for french. In *Proceedings of the eight international conference on language resources and evaluation (LREC'12)*, pages 4019–4026.
- Matthew Shardlow. 2014. A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications*, 4(1):58–70.
- Donatella Spinelli, Maria De Luca, Gloria Di Filippo, Monica Mancini, Marialuisa Martelli, and Pierluigi Zoccolotti. 2005. Length effect in word naming in reading: Role of reading experience and reading deficit in italian readers. *Developmental neuropsychology*, 27(2):217–235.
- Johannes C. Ziegler, Conrad Perry, Anna Ma-Wyatt, Diana Ladner, and Gerd Schulte-Körne. 2003. Developmental dyslexia in different languages: Language-specific or universal? *Journal of experimental child psychology*, 86(3):169–193.