

Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia

Núria Gala, Johannes Ziegler

► **To cite this version:**

Núria Gala, Johannes Ziegler. Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. Computational Linguistics for Linguistic Complexity, workshop at COLING (Computational Linguistics conference), Dec 2016, Osaka, Japan. hal-01757941

HAL Id: hal-01757941

<https://hal-amu.archives-ouvertes.fr/hal-01757941>

Submitted on 4 Apr 2018

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.

Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia

Núria Gala

Aix Marseille Université
LIF-CNRS UMR 7279

163, Av. de Luminy 13288 Marseille
nuria.gala@univ-amu.fr

Johannes Ziegler

Aix Marseille Université
LPC-CNRS UMR 7290

3, place Victor Hugo 13331 Marseille
johannes.ziegler@univ-amu.fr

Abstract

Lexical complexity plays a central role in readability, particularly for dyslexic children and poor readers because of their slow and laborious decoding and word recognition skills. Although some features to aid readability may be common to many languages (e.g., the majority of 'easy' words are of low frequency), we believe that lexical complexity is mainly language-specific. In this paper, we define lexical complexity for French and we present a pilot study on the effects of text simplification in dyslexic children. The participants were asked to read out loud original and manually simplified versions of a standardized French text corpus and to answer comprehension questions after reading each text. The analysis of the results shows that the simplifications performed were beneficial in terms of reading speed and they reduced the number of reading errors (mainly lexical ones) without a loss in comprehension. Although the number of participants in this study was rather small (N=10), the results are promising and contribute to the development of applications in computational linguistics.

1 Introduction

It is a fact that lexical complexity must have an effect on the readability and understandability of text for people with dyslexia (Hyönä J., Olson R. K., 1995). Yet, many of the existing tools have only focused on the visual presentation of text, such as the use of specific dyslexia fonts or increased letter spacing (Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M. and Ziegler, J. C., 2012). Here, we investigate the use of text simplification as a tool for improving text readability and comprehension.

It should be noted that comprehension problems in dyslexic children are typically a consequence of their problems in basic decoding and word recognition skills. In other words, children with dyslexia have typically no comprehension problems in spoken language. However, when it comes to reading a text, their decoding is so slow and strenuous that it takes up all their cognitive resources. They rarely get to the end of a text in a given time, and therefore fail to understand what they read. Long, complex and irregular words are particularly difficult for them. For example, it has been shown that reading times of children with dyslexia grow linearly with each additional letter (Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M. and Zoccolotti, P., 2005) (Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D. and Schulte-Korne, G., 2003). Because children with dyslexia fail to establish the automatic procedures necessary for fluent reading, they tend to read less and less. Indeed, a dyslexic child reads in one year what a normal reader reads in two days (Cunningham, A. E. and Stanovich, K. E., 1998) - a vicious circle for a dyslexic child because becoming a fluent reader requires extensive training and exposure to written text (Ziegler, J. C., Perry, C. and Zorzi, M., 2014)

In this paper, we report an experiment comparing the reading performance of dyslexic children and poor readers on original and simplified corpora. To the best of our knowledge, this is the first time that such an experiment is undertaken for French readers. Our aim was to reduce the linguistic complexity of ten standardized texts that had been developed to measure reading speed. The idea was to identify the words and the structures that were likely to hamper readability in children with reading deficits. Our hypothesis was that simplified texts would not only improve reading speed but also text comprehension.

A lexical analysis of the reading errors enabled us to identify what kind of lexical complexity was particularly harmful for dyslexic readers and define what kind of features should be taken into account in order to facilitate readability.

2 Experimental Study

2.1 Procedure and participants

We tested the effects of text simplification by contrasting the reading performance of dyslexic children on original and manually simplified texts and their comprehension by using multiple choice questions at the end of each text. The children were recorded while reading aloud. They read ten texts, five original and five simplified in a counter-balanced order. Each text was read in a session with their speech therapists. The texts were presented on a A4 sheet printed in 14 pt Arial font. The experiment took place between december 2014 and march 2015.

After each text, each child had to answer the three multiple-choice comprehension questions without looking at the texts (the questions were the same for the original and the simplified versions of the text). Three possible answers were provided in a randomized order : the correct one, a plausible one taking into account the context, and a senseless one. Two trained speech therapists collected the reading times and comprehension scores, annotated the reading errors, and proposed a global analysis of the different errors (cf. 3.1) (Brunel, A. and Combes, M., 2015).

Ten children aged between 8 and 12 attending regular school took part in the present study (7 male, 3 female). The average age of the participants was 10 years and 4 months. The children had been formally diagnosed with dyslexia through a national reference center for the diagnosis of learning disabilities. Their reading age¹ corresponds to 7 years and 6 months, which meant that they had an average reading delay of 2 years and 8 months.

2.2 Data set

The corpora used to test text simplification is a collection of ten equivalent standardized texts (IReST, International Reading Speed Texts²). The samples were designed for different languages keeping the same difficulty and linguistic characteristics to assess reading performances in different situations (low vision patients, normal subjects under different conditions, developmental dyslexia, etc.). The French collection consists on nine descriptive texts and a short story (more narrative style).

The texts were analyzed using TreeTagger (Schmid, H., 1994), a morphological analyzer which performs lemmatization and part-of-speech tagging. The distribution in terms of part-of-speech categories is roughly the same in original and simplified texts, although simplified ones have more nouns and less verbs and adjectives. Table 1 shows the average number of tokens per text and per sentence, the average number of sentences per text, the distribution of main content words and the total number of lemmas :

	IReST originals	IReST simplified
Average number tokens/text	131.4	124.2
Average number tokens/sent	8,8	9,1
Average number sent/text	15,9	14,5
Average NOUNs	39.54%	39.70%
Average VERBs	31.07%	31.87%
Average ADJs	11.68%	9.48%
Total lemmas	779	728

TABLE 1 – IReST corpora features before and after manual simplifications.

2.3 Simplifications

Each corpus was manually simplified at three linguistic levels (lexical, syntactic, discursive). It is worth mentioning that, in previous work, text simplifications are commonly considered as lexical and

1. We used standardized reading tests to assess the reading level of each child, i.e. lAlouette (Lefavrais, 1967) and PM47 (Raven, 1976) and a small battery of tests to assess general cognitive abilities.

2. <http://www.vision-research.eu>

syntactic (Carroll, J. and Minnen, G. and Pearce, D. and Devlin, S. and Tait, J., 1999), little attention is generally paid to discourse simplification with a few exceptions. In this study, we decided to perform three kinds of linguistic transformations because we made the hypothesis that all of them would have an effect on the reading performance. However, at the time being, only the lexical simplifications have been analyzed in detail (cf. section 3.2).

The manual simplifications were made according to a set of criteria. Because of the absence of previous research on this topic, the criteria were defined by three annotators following the recommendations for readers with dyslexia (Ecalte and Magnan, 2006) for French and (Rello, L., 2014) for Spanish.

Lexical simplifications. At the lexical level, priority was given to high-frequency words, short words and regular words (high grapheme-phoneme consistency). Content words were replaced by a synonym³. The lexical difficulty of a word was determined on the basis of two available resources : Manulex (Lété et al., 2004)⁴, a grade-level lexical database from French elementary school readers, and FLELex (François et al., 2014)⁵, a graded lexicon for French as a foreign language reporting frequencies of words across different levels.

If the word in the original text had a simpler synonym (an equivalent in a lower level) the word was replaced. For instance, the word *consommer* ('to consume') has a frequency rate of 3.55 in Manulex, it was replaced by *manger* ('to eat') that has 30.13. In most of the cases, a word with a higher frequency is also a shorter word : *elle l'enveloppe dans ses fils collants pour le garder et le **consommer** plus tard > ... pour le garder et le **manger** plus tard* ('she wraps it in her sticky net to keep it and eat it later').

Adjectives or adverbs were deleted if there was an agreement among the three annotators, i.e. if it was considered that the information provided by the word was not relevant to the comprehension of the sentence. To give an example, *inoffensives* ('harmless') was removed in *Il y a des mouches inoffensives qui ne piquent pas* ('there are harmless flies that do not sting').

In French, lexical replacements often entail morphological or syntactic modifications of the sentence, in these cases the words or the phrases were also modified to keep the grammaticality of the sentence (e.g. determiner and noun agreement) and the same content (meaning). Example, respectively with number and gender agreement : *une partie des plantes meurt* and *quelques plantes meurent* ('some plants die'), or *la sécheresse* ('drought') and *au temps sec* ('dry weather').

Syntactic simplifications. Structural simplifications imply a modification on the order of the constituents or a modification of the sentence structure (grouping, deletion, splitting (Brouwers, L. and Bernhard, D. and Ligozat, A.-L. and François, T., 2014)). In French, the canonical order of a sentence is SVO, we thus changed the sentences where this order was not respected (for stylistic reasons) : *ensuite poussent des buissons* was transformed into *ensuite des buissons poussent* ('then the bushes grow'). The other syntactic reformulations undertaken on the IReST corpora are the following : passive voice to active voice, and present participle to present tense (new sentence through punctuation or coordinate conjunction).

Discursive simplifications. As for transformations dealing with the coherence and the cohesion of the text, given that the texts were short, we only took into account the phenomena of anaphora resolution, i.e. expliciting the antecedent of a pronoun (the entity which it refers to). Although a sentence where the pronouns have been replaced by the antecedents may be stylistically poorer, we made the hypothesis that it is easier to understand. For instance : *leurs traces de passage* ('their traces') was replaced by *les traces des souris* ('the mice traces').

The table 2 gives an idea of the transformations performed in terms of quantity. As clearly showed, the majority of simplifications were lexical :

3. The following reference resources were used : the database www.synonymes.com and the *Trésor de la Langue Française informatisé* (TLFi) <http://atilf.atilf.fr/tlf.htm>.

4. <http://www.manulex.com>

5. <http://cental.uclouvain.be/flelex/>

Lexical Simplifications		85.91%
Direct replacements	57.04%	
Removals	13.38%	
Replacements with morphological changes	4.93%	
Replacements with syntactical changes	10.56%	
Syntactic Simplifications		9.86%
Reformulations	7.75%	
Constituent order	2.11%	
Discursive Simplifications		4.23%
Total		100 %

TABLE 2 – Linguistic transformations on the IReST French corpora.

3 Results

Two different analyses were performed : one for quantitatively measuring the reading times, the number of errors and the comprehension scores. The second one took specifically into account the lexicon : the nature of the words incorrectly read.

3.1 Behavioral data analysis

Reading Times. The significance of the results was assessed with a pairwise t-test (Student)⁶. The results are shown on table 3 :

Variables	Original texts	Simplified texts	T value	Significance
Reading times (sec)	159.94	134.70	-3.528	0.006**
Reading speed (words per minute)	64.85	71.10	4.105	0.003**

TABLE 3 – Significance of the results obtained.

From this table it can be seen that the overall reading times of simplified texts were significantly shorter than the reading times of original texts. While this result can be attributed to the fact that simplified texts were slightly shorter than original texts, it should be emphasized that **reading speed (words per minute), which is independent of the length of a text, was significantly greater in simplified texts than in original texts.**

Number of errors. The total number of errors included :

- (A) the total number of skipped words, repeated words (words read twice), interchanged words, line breaks, repeated lines (line read twice)
- (B) the total number of words incorrectly read for lexical reasons (the word read is a pseudo-word or a different word)
- (C) the total number of words incorrectly read for grammatical reasons (the word read has the same grammatical category (part-of-speech) but varies on number, gender, tense, mode, person)

First of all, it should be noted that **participants made fewer errors in simplified texts than in original ones (5,5% vs 7,7%)**⁷. The table 4 shows the distribution of all the errors :

Type of error	Original texts		Simplified texts	
(A) reading	40	10.05%	36	13.64%
(B) lexical	182	45.73%	112	42.42%
(C) grammatical	176	44.22%	116	43.94%
	398	100,00%	264	100,00%

TABLE 4 – Distribution of the types of errors in original and simplified texts.

It can be noted that lexical and grammatical errors occurred equally often⁸.

Comprehension scores

6. ** significant results with $p < 0.01$

7. This difference was significant in a t-test ($t = 2,3$, $p < 0.05$)

8. A more detailed analysis of these errors is proposed on section 3.2.

The results of the comprehension questionnaire are better for simplified than for original texts (marginal gain⁹) as shown on table 5 :

Variable	Original texts	Simplified texts	T value	Significance
Comprehension score	2.08	2.30	1.819	0.10+

TABLE 5 – Significance of the results obtained.

These results entail that **dyslexic children read the simplified version of the corpus without a significant loss of comprehension. If anything, they showed a marginal increase in comprehension scores for simplified texts.**

3.2 Lexical analysis

As we were interested in the lexicon of the corpus, an analysis of the content words (i.e. nouns, verbs, adjectives, adverbs) incorrectly read was undertaken in order to better target the reading pitfalls. From our study, we identified 404 occurrences that were incorrectly read, corresponding to 213 different lemmas (to be precise, there were 235 tokens (22 were inflected variants), i.e. *arbre* and *arbres*, or *restaient*, *restent*, *rester*). 404 wrong read words corresponds to 26.81 % of the content words of the corpora, which means that more than **one word out of four is incorrectly read**.

It is worth mentioning that we did not count monosyllabic grammatical words as determiners, pronouns or prepositions, although an important number of errors occurred also on those tokens, i.e. *le* read *la* ('the'), *ces* read *des* ('these'), *pour* read *par* ('for'). We make the hypothesis that the readers concentrate their efforts on decoding content words, and not grammatical ones, because they are those that carry the semantic information and are thus important for text comprehension. Besides, as grammatical words are usually very short and frequent in French, they have a higher number of orthographic neighbours and people with dyslexia tend to confuse short similar words.

We distinguished the words that were replaced by a pseudo-word (29.46%) and those replaced by other existing words on French vocabulary (70.37%). These figures can be compared with those obtained by Rello and collaborators (Rello, L. and Baeza-Yates, R. and Saggion, H. and Pedler, J., 2012). Non-word errors are pronunciations that do not result in an existing word, real-word errors are pronunciations that result in an incorrect but existing word. Non-word errors appear to be higher in English (83%) and in Spanish (79%), but not in French where real-word errors were clearly a majority¹⁰ :

Category	English	Spanish	French
Real-word errors	17%	21%	70.54%
Non-word errors	83%	79%	29.46%

TABLE 6 – Error typology compared across languages.

The overall error typology that we propose is shown on table 7 :

Type of lexical replacement			Original word	English translations
Pseudo-word	119	29.46%	<i>grenouille</i> > * <i>greniole</i>	frog, *
Grammatical variant	135	33.42%	<i>oubliaient</i> > <i>oublent</i>	forgot, forget
Lexical replacement	84	20.79%	<i>attendent</i> > <i>attaquent</i>	wait, attack
Morphological variant	43	10.64%	<i>construction</i> > <i>construire</i>	build, to build
Orthographical neighbour	23	5.69%	<i>jaunes</i> > <i>jeunes</i>	yellow, young
Total	404	100%		

TABLE 7 – Error typology.

Grammatical variants concern variations on gender and number for nouns, and for person, tense and mode for verbs. Lexical replacements are words read as if they were other words with orthographical similarities (*lieu* > *île*, *en fait* > *enfin*, *commun* > *connu*, etc.). Morphological variants are words of

9. $p < 0.1$

10. This finding will deserve more attention in future work.

the same morphological family (*baisse* > *basse*, *malchanceux* > *malchance*). As for orthographical neighbours, we specifically distinguish word pairs where the difference is only of one letter (*raisins* > *raisons*, *bon* > *don*).

Concerning word length for all the mentioned features, 36.88% of the words read were replaced by words of strictly the same length (*forment* > *formant*, *catégorie* > **calégorie*), 14.11% were replaced by longer ones (*utile* > *utilisé*, *suffisant* > *suffisamment*), **49.01%** were replaced by shorter ones (*nourriture* > *nature*, *finie* > *fine*, *empilées* > *empli*). The average length of the 404 words incorrectly read is 7.65 characters (the shortest has three characters, *bon*, and the longest 16, *particulièrement*).

The average number of orthographical neighbours is 3.24, with eight tokens having more than ten neighbours : *bon*, *bois*, *basse*, *foule*, *fine*, *fil*, *garde*, *sont* ('good, wood, low, crowd, thin, thread, keeps, are').

As far as the grammatical categories are concerned, **the majority of the errors were on verbs**. They concerned grammatical variants of person, tense (past *imparfait* > present) and mode (present > present participle). The distribution on part-of-speech tags errors is shown on table 8 :

Part-of-speech tags of tokens incorrectly read		
VERB	196	48.51 %
NOUN	115	28.47%
ADJECTIVE	48	11.88%
ADVERB	25	6.19%
Other categories (determiners excluded)	20	4.95%

TABLE 8 – Part-of-speech distribution of the tokens in the corpora.

We analyzed the syllable structure of the 404 tokens. The average number of syllables is 2.09, the distribution is shown on table 9 :

Number of syllabs		
1 syllab	72	30,64%
2 syllabs	96	40,85%
3 syllabs	47	20,00%
4 syllabs	15	6,38%
5 syllabs	5	2,13%
	235	100,00%

TABLE 9 – Syllabs distribution of the tokens in the corpora.

In French, it is stated that the more frequent (and easier) structure is CV and V. In our results, 58,69% of the words contain this common structure, while **41,31%** present a more complex structure (CVC, CVCC, CYC¹¹, etc.), as shown on table 10 :

Syllable structure		
CV	230	47,03%
V	57	11,66%
CVC	107	21,88%
CVCC, CCVC, CYVC	47	9,61%
CYV, CCV, VCC, CVY	34	6,95%
VC, YV	10	2,04%
VCCC, CCYV, CCVCC	4	0,82%
	489	100,00%

TABLE 10 – Syllable structure.

We finally analyzed the consistency of grapheme-to-phoneme correspondences which is particularly irregular in French (silent letters, nasal vowels, etc.)¹². As mentioned above, the average length of the words incorrectly read is 7.65 and their average in number of phonemes is 4.95. This means that the

11. C is a consonant, V is a vowel, Y is a semi-vowel, i.e. [j] in *essayait* [e-se-je], [w] in *doivent* [dwav]

12. This is not the case for other languages, e.g. the Spanish writing system has consistent grapheme-to-phoneme correspondences.

average difference between the number of letters and the number of real phonemes is 2.71. Only four tokens were regular (same number of phonemes than letters : *existe, mortel, partir, plus* ('exists, mortal, leave, plus')). The highest difference is 6 in *apparaissent, épargnaient* ('appear, saved') with 12 letters and 6 phonemes each, and *mangeaient* ('ate') with 10 letters and 4 phonemes. **All the words incorrectly read were thus irregular as far as grapheme-to-phoneme consistency is concerned.**

4 Discussion : determining where complexity is

According to the literature, complexity for children with dyslexia should be found on long and less frequent words. More precisely, from the analysis of the reading errors obtained on our first pilot-study, the errors mainly occur on verbs and nouns with complex syllable structure, i.e. irregular grapheme-to-phoneme correspondences, words with many orthographic neighbours or many morphological family members which are more frequent. Visual similarity is a source of error, specially for the following pairs¹³ :

Letter alternation	Example	English translations
p/tt	<i>guêpe > guette</i>	wasp, watch
b/d	<i>bon > don, bien > dans</i>	good, gift / fine, in
d/q	<i>attendent > attaquent</i>	wait, attack
t/l	<i>ramifications > *ramificalons</i>	branching, *
q/g	<i>quand > grand</i>	when, big
g/j	<i>augmente > *anjmente</i>	increases, *
m/n	<i>commun > connu</i>	common, known
r/l	<i>grâce > glace</i>	grace, ice
l/i	<i>lieu > île, plus > puis</i>	place, island / plus, after
u/n	<i>déguiser > *dénise</i>	dress up, *
e/a	<i>vivent > vivant</i>	live, living
e/o	<i>veulent > *voulent</i>	want, *

TABLE 11 – Graphical alternations.

In all the replacements we can observe visual similarities. As shown in table 12, **the word that is actually read tends to be in most of the cases shorter and more frequent**¹⁴ than the original one :

Type of replacement	Example	English translations	Frequencies
Similar lexical items	<i>meurt > mur</i>	dies, wall	9.89 - 179.63
	<i>toiles > étoiles</i>	canvas/web, stars	16.77 - 121.99
	<i>poison > poisson</i>	poison, fish	16.64 - 230.20
Orthographical neighbours	<i>minuit > minute</i>	midnight, minute	35.57 - 57.70
	<i>branches > blanches</i>	branches, white	98.03 - 44.76
	<i>raisins > raisons</i>	grape, reason	9.86 - 22.87
Morphological variants	<i>banquets > banque</i>	banquets, bank	0.21 - 19.54
	<i>construction > construire</i>	build, to build	31.61 - 68.73
	<i>piqûres > piques</i>	bitting, endpin	5.03 - 0.93
Grammatical variants	<i>animaux > animal</i>	animals, animal	415.56 - 333.50
	<i>mangeaient > mangent</i>	ate, eat	9.66 - 31.69
	<i>permettent > permet</i>	they allow, he/se/it allows	29.90 - 89.65

TABLE 12 – Lexical replacements typology with frequencies of the tokens.

To sum up, lexical complexity for dyslexic readers in French is to be found on verbs and nouns longer than seven characters, presenting letters with similar equivalents, with complex syllables and irregular phoneme-to-grapheme consistency. Lexical replacements of words incorrectly read should consider shorter and more frequent words and words with higher grapheme-to-phoneme consistency.

5 Conclusion

In this paper we have presented the results of a first pilot-study aiming at testing the effects of text simplification on children with dyslexia. From our results, reading speed is increased without a loss of

13. Other possible similar pairs (not found in our corpora) : t/f, u/v, a/o

14. The frequencies have been extracted from the Manulex database (column including the five levels).

comprehension. It is worth mentioning that reading errors were lower on simplified texts (in this experiment, simplified texts contained a majority of lexical simplifications). The comprehensive analyses of reading errors allow us to propose a detailed description of lexical complexity for dyslexic children. The causes of lexical complexity were mainly related to word length (words longer than seven characters), irregular spelling-to-sound correspondences and infrequent syllable structures.

The insights obtained as a result of this first pilot-study are currently being integrated into a model aiming at providing better accessibility of texts for children with dyslexia. We are currently working in a new study with children in French schools to refine the features that are to be taken into account in our model. These results will be integrated into a tool that will automatically simplify texts by replacing complex lexical items with simpler ones.

Acknowledgements

We deeply thank the speech therapists Aurore Brunel and Mathilde Combes for collecting the reading data and providing a first analysis of the data. We also thank Luz Rello for her valuable insights on parts of the results.

References

- Brouwers, L. and Bernhard, D. and Ligozat, A.-L. and François, T. (2014). Syntactic French Simplification for French. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR) at EACL 2014*, page 4756, Gothenburg, Sweden.
- Brunel, A. and Combes, M. (2015). Simplification de textes pour faciliter leur lisibilité et leur compréhension. Mémoire de fin d'études en vue de l'obtention du certificat de capacité d'orthophonie. Master's thesis, Aix Marseille Univ.
- Carroll, J. and Minnen, G. and Pearce, D. and Devlin, S. and Tait, J. (1999). Simplifying Text for Language Impaired readers. In *Proceedings of European Association of Computational Linguistics*, pages 269–270.
- Cunningham, A. E. and Stanovich, K. E. (1998). What reading does for the mind. *Am Educator*, 22 :8–15.
- Ecalte, J. and Magnan, A. (2006). Des difficultés en lecture à la dyslexie : problèmes d'évaluation et de diagnostic. *Glossa*, 97 :4–19.
- François, T., Gala, N., Watrin, P., and Fairon, C. (2014). FLELex : a graded lexical resource for French foreign learners. In *Proceedings of International conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik.
- Hyönä J., Olson R. K. (1995). Eye fixation patterns among dyslexic and normal readers : effects of word length and word frequency. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 21(6) :1430–40.
- Lefavrais, P. (1967). Test de l'alouette.
- Lété, B., Sprenger-Charolles, L., and Colé, P. (2004). Manulex : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments and Computers*, 36 :156–166.
- Raven, J. C. (1976). Pm47 : Standard progressive matrices : Sets a, b, c, d and e.
- Rello, L. (2014). *DysWebxia. A Text Accessibility Model for People with Dyslexia*. PhD thesis, Universitat Pompeu Fabra, Barcelone.
- Rello, L. and Baeza-Yates, R. and Saggion, H. and Pedler, J. (2012). A First Approach to the Creation of a Spanish Corpus of Dyslexic Texts. In *LREC Workshop Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*, pages 22–26, Istanbul, Turkey.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *International Conference on new methods in language processing*, Manchester, UK.
- Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M. and Zoccolotti, P. (2005). Length effect in word naming in reading : role of reading experience and reading deficit in italian readers. *Developmental Neuropsychology*, 27(2) :217–235.
- Ziegler, J. C., Perry, C. and Zorzi, M. (2014). Modeling reading development through phonological decoding and self-teaching : Implications for dyslexia. *Philosophical Transactions of the Royal Society B*.
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D. and Schulte-Korne, G. (2003). Developmental dyslexia in different languages : Language-specific or universal ? *Journal of Experimental Child Psychology*, 86(3) :169–193.
- Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M. and Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences*, 109(28) :11455–11459.