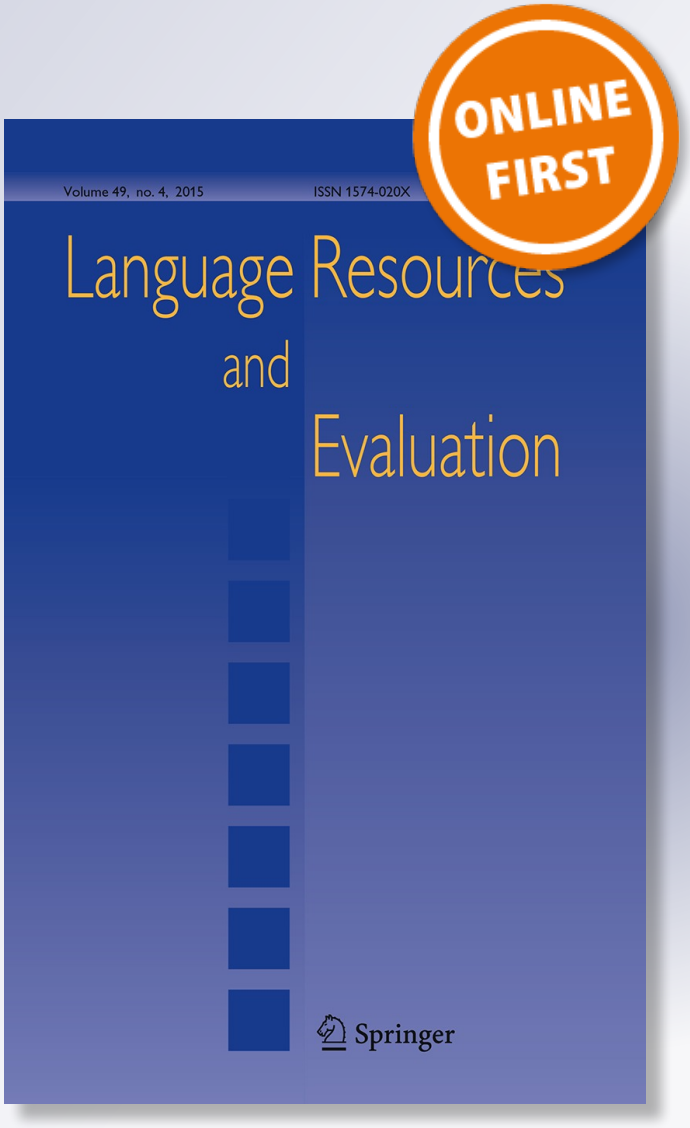
A resource of errors written in Spanish by people with dyslexia and its linguistic, phonetic and visual analysis

Luz Rello, Ricardo Baeza-Yates & Joaquim Llisterri

Language Resources and Evaluation

ISSN 1574-020X

Lang Resources & Evaluation
DOI 10.1007/s10579-015-9329-0



Springer

Lang Resources & Evaluation
DOI 10.1007/s10579-015-9329-0

ORIGINAL PAPER

A resource of errors written in Spanish by people with dyslexia and its linguistic, phonetic and visual analysis

Luz Rello[1] · **Ricardo Baeza-Yates**[2,3] · **Joaquim Llisterri**[4]

Abstract In this work we introduce the analysis of *DysList*, a language resource for Spanish composed of a list of unique spelling errors extracted from a collection of texts written by people with dyslexia. Each of the errors was annotated with a set of characteristics as well as with visual and phonetic features. To the best of our knowledge, this is the largest resource of this kind in Spanish. We also analyzed all the features of Spanish errors and our main finding is that dyslexic errors are phonetically and visually motivated.

Keywords Errors · Dyslexia · Visual · Phonetics · Resource · Spanish

1 Introduction

Dyslexia is a reading and spelling disorder of neurological origin (American Psychiatric Association 2000; World Health Organization 1993).[1] It is characterized by difficulties with accurate and fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the

[1] In some literature, dyslexia is referred to as a specific reading disability only (Vellutino et al. 2004) and dysgraphia as its written manifestation (Romani et al. 1999).

✉ Luz Rello
luzrello@gmail.com

[1] HCI Institute, School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

[2] Web Research Group, Univ. Pompeu Fabra, Barcelona, Spain

[3] Yahoo Labs, Sunnyvale, CA, USA

[4] Departament de Filologia Espanyola, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

Published online: 03 February 2016

phonological component of language that is often unexpected in comparison to other cognitive abilities (Lyon et al. 2003). Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge (Orton Dyslexia Society Research Committee 1994). Although dyslexia is universal, its prevalence varies depending on the language, ranging from 10 to 17.5 % of the population of the USA (Interagency Commission on Learning Disabilities 1987) to 7.5–11 % of the Spanish speaking population (Carrillo et al. 2011).[2] An estimation of the presence of dyslexic texts in the Web shows that at least 0.67 % of the spelling errors found in the English Web (Baeza-Yates and Rello 2011) and 0.43 % in the Spanish Web (Rello and Baeza-Yates 2012), originate from people with dyslexia.

The spelling errors that people with dyslexia write are very valuable and have been used for various purposes, ranging from diagnosing dyslexia to software applications targeted to people with dyslexia. However, resources such as corpora or lists of dyslexic errors are scarce. In this paper we describe the creation of a corpus-based list of errors written by people with dyslexia in Spanish. We compare the different types of dyslexic errors in Spanish and English and analyze the general characteristics as well as the phonetic and visual features of the errors in Spanish. The main contributions of this paper are a resource and its analysis:[3]

- The resource *DysList*, composed of a list of unique errors extracted from a set of texts written by people with dyslexia. The errors are annotated with general characteristics and with phonetic, and visual information. To the best of our knowledge, this is the largest resource of this kind in Spanish, and it is freely available online.[4]
- In Spanish, substitution errors are the most frequent ones (58.84 %), while transpositions are the less frequent type (1.45 %). If we consider classes of letters, the most frequent errors are: inserting or deleting a consonant (37.9 %), deleting or inserting a vowel (37.5 %), and substituting two letters which are used to represent a similar sound or deleting/inserting an ⟨h⟩ (15.4 %).
- Dyslexic errors are visually motivated. More concretely, taking into account handwriting we found that:
  - 67.11 % of the error letters are fuzzy letters.
  - 46.91 % of the error letters had a mirror feature.
  - 38.23 % of the error letters had a rotation feature.
- Dyslexic errors are phonetically motivated. More concretely:

[2] This estimation was carried out taking into account primary schools in the region of Murcia, Spain.

[3] The preliminary resource was described in the *Proceedings of the LREC 2102 Workshop on Natural Language Processing for Improving Textual Accessibility* (NLP4ITA), 27 May, Istanbul, Turkey, pp. 22–26 (Rello et al. 2012). This work presents an enlarged collection of texts with new annotations. The preliminary analysis appeared in the *Proceedings of the 9th International Conference on Language Resources and Evaluation* (LREC 2014), 26–31 May, Reykjavik, Iceland, pp. 1289–1296 (Rello et al. 2014). In comparison with the LREC paper, which is focused on the language resource only, this paper extends previous work by presenting the analyses of the errors using linguistic, phonetic and visual features.

[4] *Available at:* grupoweb.upf.es/WRG/resources/DysWebxia/DysList_resource.csv.gz.

- The most frequent errors involve letters in which the one-to-one correspondence between graphemes and phones is not maintained (⟨b, v⟩, ⟨g, j⟩, ⟨c, z⟩, ⟨c, s⟩, ⟨r⟩) and the letter ⟨h⟩, which, in most cases, does not have a phonetic realization in Spanish.
- Vowels substitutions take place between phones sharing one or two phonetic features, lip rounding being the most frequently involved in errors; only 15.87 % of the vowel substitution errors correspond to phones that do not have any feature in common. Diphthongs account just for the 0.94 % of errors in the resource.
- Consonant substitutions occur more frequently in single consonants (46.37 %) than in consonant clusters (0.60 %). Most of the substitutions errors between consonants take place within consonants sharing their three phonetic features (48.43 %), while confusions between consonants sharing one (19.52 %) or two (26.15 %) features are less commonly found; only 5.52 % of the substitution errors concern consonants which do not share any phonetic feature.

The rest of this paper is organized as follows. In the next section we cover related work and show how dyslexic errors have been used for different purposes. In Sect. 3 we present the first step for the development of *DysList*, the creation of a collection of texts written by children with dyslexia, the extraction of the errors, and its classification. In Sect. 4 we describe the resource *DysList*, a list of dyslexic errors, explaining the annotation criteria. In Sect. 5 we present the analysis of the dyslexic errors focusing in general characteristics, the visual feature analysis and the phonetic analysis. We discuss the results and draw conclusions in Sect. 6, as well as a comparison with a similar resource in English in an "Appendix".

2 Related work

In this section we summarize previous approaches to the use of dyslexic errors as a source of knowledge and present corpora or lists of errors written by people with dyslexia.

2.1 The knowledge of dyslexic errors

In general terms, errors are a source of knowledge. For instance, the presence of errors in the textual Web has been used for detecting spam (Piskorski et al. 2008) or measuring the quality of web content (Gelman 2008).

Since the kinds of errors that people with dyslexia make are related to the types of difficulties that they have (Sterling et al. 1998), their written errors have been used for various purposes such as studying or diagnosing dyslexia, as well as accessibility.

First, the analyses of writing errors made by people with dyslexia were used in previous literature to study different aspects of dyslexia (Connelly et al. 2006; Aragón and Silva 2000). For example, the specific types of dyslexic errors highlight

different aspects of dyslexia (Treiman 1997), such as the phonological processing deficit (Lindgrén and Laine 2011; Moats 1996). Also dyslexic error rates vary depending on the language writing system (Lindgrén and Laine 2011). However, compared to non-dyslexics, people with dyslexia present more errors attributable to phonological impairment, spelling knowledge, and lexical mistakes (Sterling et al. 1998).

Second, since people with dyslexia exhibit higher spelling error rates than non-dyslexic people (Coleman et al. 2009), there are diagnoses of dyslexia based on the spelling score (Schulte-Körne et al. 1996; Toro and Cervera 1984). Also, the spelling error rate is being used as a diagnosing factor in the current official Catalonian protocol (Col·legi de Logopedes de Catalunya 2011).

Third, the exploration of corpora of dyslexic errors (Pedler 2007; Rello et al. 2012) was used for various accessibility related purposes such as the development of specialized tools like spellcheckers (Korhonen 2008; Li et al. 2013; Pedler 2007), text prediction software,[5] games (Rello et al. 2012, 2014), or word processors which perform text customization taking into account frequent writing errors (Gregor et al. 2003).

2.2 Language resources related to dyslexia

Corpora of Dyslexic Texts To the best of our knowledge, there is only one corpus of dyslexic texts, the one used by Pedler (2007) for the creation of a spellchecker of real-word errors (see definition in Sect. 3.2) made by people with dyslexia. This corpus in English has 3134 words and 363 errors (Pedler 2007). It is composed of: word-processed homework (saved before it was spellchecked) produced by a third year secondary school student; two error samples used for a comparative test of spellcheckers (Mitton 1996); and short passages of creative writing produced by secondary school children of low academic ability in the 1960s (Holbrook 1964). To develop a program designed to correct errors made by people with dyslexia, that initial corpus was enlarged to 21,524 words containing 2654 errors, with over 800 real-word errors. The additional sources for that corpus were: texts from a student with dyslexia, texts from an online typing experiment (Spooner 1998), samples from dyslexic bulletin boards and mailing lists, and stories written by children with dyslexia.

Lists of Dyslexic Errors Regarding lists of dyslexic errors, the only similar resource is the list of English confusion sets compiled by Pedler (2007),[6] extracted from the corpus of texts written by people with dyslexia mentioned before. This list is composed of 833 confusion sets. A confusion set is a small group of words that are likely to be confused with one another, such as *weather* and *whether*.

To the best of our knowledge, with the exception of our study, there are no publicly available Spanish language resources—corpora and lists—composed of texts and errors, written by people with dyslexia.

[5] *Penfriend XL* (http://www.penfriend.biz/).

[6] http://www.dcs.bbk.ac.uk/jenny/resources.html.

3 A collection of dyslexic texts in Spanish

In this section we explain how we collected texts written by children with dyslexia, extracted the errors, and classified them.

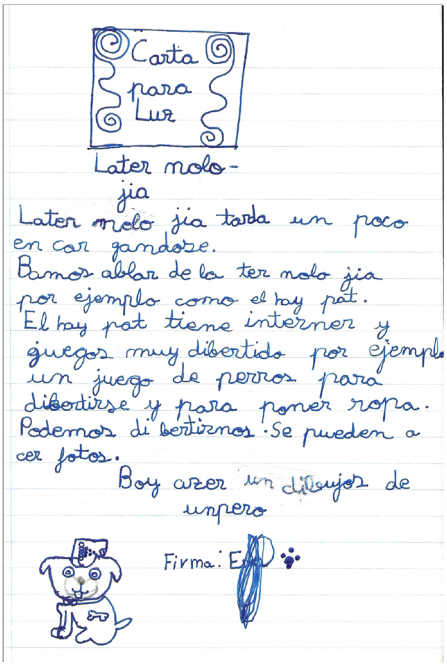
3.1 Extracting errors from dyslexic texts

Manifestations of dyslexia vary among languages (Goulandris 2003) but also among subjects and ages (Vellutino et al. 2004). For instance, misspelling rate in dyslexic children is higher than in adults (Sterling et al. 1998). However, experiments evidence that adults with dyslexia have a continuing problem in the lexical domain, manifested in a poor spelling ability (Sterling et al. 1998). Therefore, we collected texts written by a similar population in terms of age, education and native language (Spanish); all the authors of the texts have been diagnosed with dyslexia. These texts were all handwritten and we transcribed them manually. The words that we were not able to transcribe due to the illegibility of the handwriting were marked.

We used a total of 83 texts composed of 54 school essays and homework exercises provided by teachers from children and teenagers with dyslexia between 6 and 15 years old, and 29 texts provided by parents of children with dyslexia. Some of these texts were school essays and others were written specifically for this study (Fig. 1). All the texts belong to children with diagnosed dyslexia (one text per participant). However, in most of the Spanish diagnoses of dyslexia, degrees and different types of dyslexia are not taken into consideration—contrary to the diagnoses in English-speaking countries. Therefore we could not include in the annotation different severity levels of dyslexia.

Most of the texts came from schools in Catalonia, in which Catalan is taught alongside with Spanish. Since Spanish spelling might be influenced by the acquisition of Catalan orthographic rules, we included the feature "Language transfer" in the annotation of the errors. Positive transfer was difficult to identify, since this would correspond to words which have the same spelling in Catalan and Spanish, so only instances of negative transfer (i.e., a spelling error in Spanish which can be explained by the influence of Catalan, such as writing the palatal nasal [ɲ] with the digraph ⟨ny⟩ as in Catalan instead of using ⟨ñ⟩ as expected in Spanish) were considered. These type of errors have been documented for secondary school students without dyslexia by Pujol (2004) and are also treated as cases of transfer at the graphematic level by Hernández García (1998), among others.

From our text collection we manually extracted the misspelled words, without taking into account illegible handwritten words. We did not consider capitalization or accentuation errors since most children at that age are still learning how to capitalize and accentuate in Spanish. From this set of words we extracted 894 different correct-misspelled pairs with a total of 1171 errors. For instance, the words *accesibilidad ('accessibility')* and *sigilosamente ('stealthily')* are the ones that have more different misspelled variants (12).



Carta para Luz *'Letter for Luz'*

***Later nolojia**
La tecnología
'The technology'

***Later *nolo *jia tarda un poco en *car *gandose.**
La tecnología tarda un poco en cargarse.
'The technology takes a little while to load'

***Bamos *ablar de la *ter *molo *jia por ejemplo como el *hay *pot.**
Vamos a hablar de la tecnología, por ejemplo como el iPod.
'Let's talk about technology such as the iPod'

El *hay *pot tiene *interner y *guejos muy *dibertido
El iPod tiene Internet y juegos muy divertidos
'The iPod has internet and very fun games'

por ejemplo un juego de *perror para *dibertirse y para poner ropa.
por ejemplo un juego de perros para divertirse y para poner ropa.
'for example a game of dogs for fun and to put clothes.'

Podemos *di *bertirnos. Se pueden *a *cer fotos.
Podemos divertirnos. Se pueden hacer fotos.
We can have fun. You can take pictures.

***Boy *azer un *dibujos de un *pero**
Voy a hacer un dibujo de un perro.
'I'll draw a picture of a dog.'

Firma: Eva
'Signature: Eva'

Fig. 1 Example of a handwritten text of a 9 year old girl with dyslexia (*left*) and its transcription in Spanish and English (*right*)

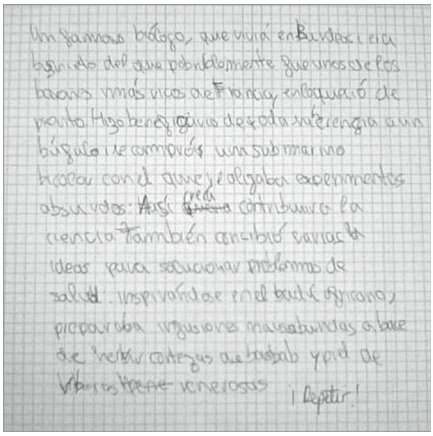
3.2 Types of dyslexic errors

The type of errors we found are consistent with previous studies in Spanish (Aragón and Silva 2000) and English (Pedler 2007).[7] We classify errors as follows:

(a) Errors based on the degree of difference with the intended or target word:
Simple errors They differ from the intended word by only a single letter or two adjacent letters. They can be due to (i) substitution, **bonde (donde, 'where')*, (ii) insertion, **cerreza (cereza, 'cherry')*, (iii) omission, **mometo (momento, 'moment')* or (iv) transposition, **porceso (proceso, 'process')*.
Multi-errors They differ in more than one letter from the target word such as **pallazo (payaso, 'clown')*.
Word boundary errors They are run-ons and split words. A run-on is the result of omitting a space, such as **talvez (tal vez, ' maybe')*. A split word occurs when a space is inserted in the middle of a word, such as **a drede (adrede, 'intentionally')*.

(b) Errors based on their correspondence with existing words:

[7] The examples report the correct word—the first word between parentheses—the related English translation—the second word in quotes. The erroneous word are those preceded by an asterisk '*'. We use the standard linguistic conventions: '⟨⟩' for graphemes, '/ /' for phonemes and '[ ]' for phones.



Un famoso biólogo, que *viviá en Burdeos, *i era biznieto de que *pobrblemente fue *unos de los barones más ricos de Francia, enloqueció de pronto. Hizo *beneficirio de toda su herencia a un búfalo y se *comprós un submarino bicolor con el que realizaba *expermentos absurdos. Así creía contribuir a la ciencia. También concibió varias ideas para solucionar problemas de salud inspirándose en el *budú africano, preparaba infusiones *nausabundas a base de hervir cortezas de *baubab y piel de víboras *venerosas.

Fig. 2 Story written in Spanish by a 14 years old boy with dyslexia

Real-word errors These are misspellings that result in another valid word. For instance, **pala (‘shovel’)* being the intended word *palabra (‘word’)*.
Non-word errors Misspellings that do not result in another correct word.

(c) Errors based on their position:
*First letter errors *ace (hace, ‘does’)*.
*Middle letter errors *cárzel (cárcel, ‘prison’)*.
*Last letter errors *interios (interior, ‘interior’)*.

One example of a fragment of our texts is given in Fig. 2. An approximated literal translation for this example is:

> A famous biologist, who lived in Bordeaux and was great-grandson of who probably was one of the wealthiest barons in France, suddenly went mad. He chose a buffalo as the beneficiary of his inheritance and bought a bicolored submarine with which he made absurd experiments. He believed that with this he contributed to science. He also conceived various ideas to solve health problems inspired by African voodoo, preparing nauseating infusions based on boiled baobab barks and the skin of poisonous snakes.

Here we have the following simple errors: (1) substitution: **i (y, ‘and’), *budú (vudú, ‘voodoo’), *venerosas (venenosas, ‘poisonous’)*, and **baubab (baobab, ‘baobab’)*; (2) insertion: **comprós (compró, ‘bought’)*; and (3) omission: **expermentos (experimentos, ‘experiments’), *unos (uno, ‘some’), *beneficirio (beneficiario, ‘beneficiary’)*, and **nausabundas (nauseabundas, ‘nauseating’)*. There is also one multi-error word with one omission and one substitution, **pobrblemente (probablemente, ‘probably’)*. All of them are non-word errors with the exception of **unos*, which seems to be a concordance error. Most errors are in the middle (7) while we have three errors in the last position and two in the first position of the word. Notice that **i*, probably an error due to transfer from Catalan (see Sect. 5.3.1), counts for both positions, the first and the last one.

We manually annotated the type(s) of error for each of the errors found in the texts and compared them with Pedler's corpus in English. In the case that there were two kinds of errors we annotated them as a multi-error; for instance, in **devidreo (de vidrio, 'of glass')* a boundary error is combined with a simple substitution error.

4 DysList

In this section we explain how we create a new resource out of our collection of texts: *DysList*, an annotated resource of dyslexic errors. First, we explain the annotation of the dyslexic errors with general features and with specific phonetic and visual information, and then we present the results of the analysis of the errors.

4.1 Annotation of dyslexic errors

With the exception of the visual features, we manually annotated each of the word-error pairs to create *DysList* with the following information:

- *Target word* the intended word the person aimed to write.
- *Misspelled word* the wrongly written word.
- *Damerau–Levenshtein distance* the minimum number of edits (insertion, deletion, substitution, transposition) required to change the misspelled error into the (target) correct word (Damerau 1964; Levenshtein 1965).[8]
- *Target and misspelled word frequencies* defined as the number of hit counts in a major search engine for web pages written in Spanish.
- *Target and misspelled length* number of characters.
- *Error position* the position in the target word where the error occurs.
- *Target word syllables* number of syllables of the target word.
- *Target syllable* the structure of the syllable where the error occurs, such as C (onsonant)V(owel), CVC, or CCV, among others.
- *Type of error*
  - S *Substitution* change one letter for another, for example **reelly (really)*.
  - I *Insertion* insert one letter, like in **situartion (situation)*. A word that has been split in two different tokens is counted as an insertion, like in **sub marine (submarine)*.
  - D *Deletion* omit one letter, as in **approch (approach)*. Run-on word boundary errors, like in **alot (a lot)*, are counted as one deletion.[9]
  - T *Transposition* reversing the order of two adjacent letters, for example **artcile (article)*.
- *Real-word* this Boolean attribute records if the error produced another real-word. For instance, *witch* being *which* the intended word.

[8] The edit or Levenshtein distance (Levenshtein 1965) is the minimum number of substitutions, insertions and deletions to transform one string into another. The Damerau version (Damerau 1964) counts a transposition as a single error instead of two errors. Notice that there might be more than one solution for the transformation associated to the edit distance.

[9] Notice that a deletion in the target word is an insertion in the misspelled word and *vice versa*.

Table 1 Visual features of the annotated target and error letters

Visual features	Values	Letter(s)
Mirror (type)	**V** = vertical, **H** = horizontal, **B** = both, **N** = none	**H** = ⟨n, u⟩, **B** = ⟨b, d, p, q⟩
Mirror (hand)	**V** = vertical, **H** = horizontal, **B** = both, **N** = none	**V** = ⟨g, h, m, n, u, v, w, y⟩, **B** = ⟨b, d, p, q⟩
Height (type)	**A** = ascender, **D** = descenders, **B** = both, **N** = none	**A** = ⟨b, d, f, h, k, l, t⟩, **D** = ⟨g, j, p, q, y⟩
Height (hand)	**A** = ascender, **D** = descenders, **B** = both, **N** = none	**A** = ⟨b, d, h, k, l, t⟩, **D** = ⟨g, j, p, q, y, z⟩, **B** = ⟨f⟩
Line (type)	**V** = vertical, **H** = horizontal, **B** = both, **N** = none	**H** = ⟨a, e, f, s⟩, **V** = ⟨m, w⟩, **B** = ⟨k⟩
Line (hand)	**V** = vertical, **H** = horizontal, **B** = both, **N** = none	**H** = ⟨k, z⟩, **V** = ⟨m, w⟩
Rotation (type)	**Y** = yes, **N** = no **Y** = yes, **N** = no	**Y** = ⟨a, e, d, b, p, q, n, u⟩
Rotation (hand)	**Y** = yes, **N** = no **Y** = yes, **N** = no	**Y** = ⟨a, b, d, e, h, m, n, p, q, u, w, y⟩
Fuzzy (type)	**Y** = yes, **N** = no **Y** = yes, **N** = no	**Y** = ⟨b, c, d, f, g, i, j, l, n, ñ, o, p, q, t, u, v⟩
Fuzzy (hand)	**Y** = yes, **N** = no **Y** = yes, **N** = no	**Y** = ⟨b, d, g, h, m, n, ñ, p, q, s, r, u, v, w, y, z⟩

- *Visual information* for each of the target and the error graphemes we annotate the letters involved in the error with the following visual information, considering both handwritten and typewritten (*sans serif*) text. See Table 1.
  Mirror letter (handwriting/typewriting) such as ⟨d⟩ and ⟨b⟩ or ⟨m⟩ and ⟨w⟩, with four possible values: vertical, horizontal, both, or none.
  Height (handwriting/typewriting) letters with descenders (e.g. ⟨p, q⟩, or ⟨g⟩), letters with ascenders (e.g. ⟨t⟩, or ⟨b⟩), both (e.g. ⟨f⟩), or none (e.g. ⟨n, m⟩, or ⟨s⟩).
  Line (handwriting/typewriting) vertical (e.g. ⟨m⟩), horizontal (e.g. ⟨e⟩), or none (e.g. ⟨o⟩).
  Rotation (handwriting/typewriting) Boolean attribute that indicates if the rotation of a letter produces another letter, such as ⟨d⟩ and ⟨p⟩.
  Fuzzy letters (handwriting/typewriting) Boolean attribute that indicates if the letter has similar visual letters (not due to a rotation or mirroring) such as ⟨s⟩ and ⟨z⟩.
- *Phonetic information* each of the target and the error phones associated to the graphemes in the text are annotated using traditional articulatory phonetic features (International Phonetic Association 1999):
  Phone type vowel (e.g. [a]) or consonant (e.g. [p]); combinations of vowels forming a diphthong (e.g. [i̯a]) and consonant clusters in syllabic onsets (e.g. [pl]) have also been annotated as specific phone types.

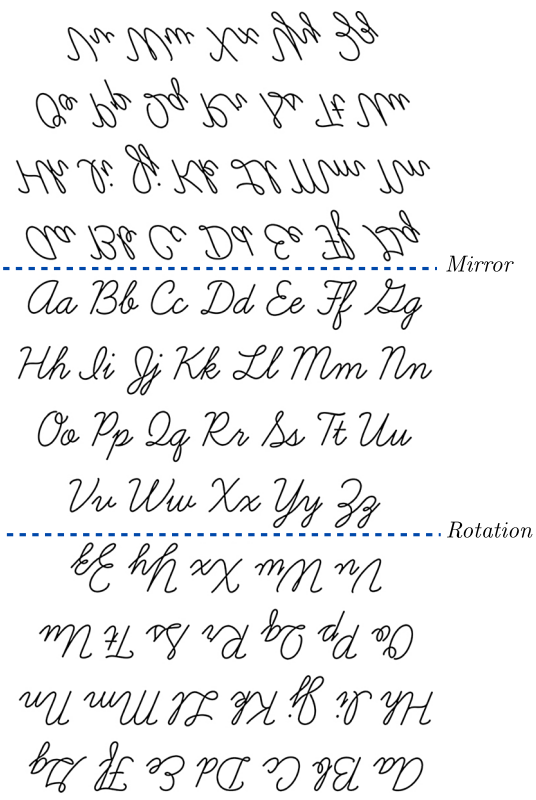


Fig. 3 Handwritten cursive letters with visual transformations

For consonants:

Voicing voiced (e.g. [b]) or voiceless (e.g. [p]).

Manner of articulation plosive (e.g. [p]), nasal (e.g. [m]), trill (e.g. [r]), tap or flap (e.g. [ɾ]), fricative (e.g. [f]), lateral (e.g. [l]), approximant (e.g. [β̞]), or affricate (e.g. [tʃ]).

Place of articulation bilabial (e.g. [p]), labiodental (e.g. [f]), interdental (e.g. [θ̪]), dental (e.g. [d̪]), alveolar (e.g. [s]), palatal (e.g. [tʃ]), or velar (e.g. [k]).

For vowels:

Height open (e.g. [a]), mid (e.g. [e]), or close (e.g. [i]).

Place of articulation front (e.g. [i]), central (e.g. [a]), or back (e.g. [u]).

Lip rounding rounded (e.g. [u]) or unrounded (e.g. [i]).

- *Language transfer* some of the errors in the list were due to transfer from Catalan to Spanish. Hence, we tagged the errors that could be explained by the influence of Catalan spelling. For instance, **accessiblidad (accesibilidad, 'accessibility')* may be due to the existence of the word *accessibilitat* in Catalan.

Table 2 Speech Assessment Methods Phonetic Alphabet (SAMPA) and International Phonetic Alphabet (IPA) symbols and phonetic features

SAMPA	IPA	Type	Voicing (cons.) Rounding (vowel)	Manner (cons.) Height (vowel)	Place
p	p	Consonant	Voiceless	Plosive	Bilabial
b	b	Consonant	Voiced	Plosive	Bilabial
t	t̪	Consonant	Voiceless	Plosive	Dental
d	d̪	Consonant	Voiced	Plosive	Dental
k	k	Consonant	Voiceless	Plosive	Velar
g	g	Consonant	Voiced	Plosive	Velar
m	m	Consonant	Voiced	Nasal	Bilabial
n	n	Consonant	Voiced	Nasal	Alveolar
J	ɲ	Consonant	Voiced	Nasal	Palatal
N	ŋ	Consonant	Voiced	Nasal	Velar
tS	tʃ	Consonant	Voiceless	Affricate	Palatal
dZ	dʒ	Consonant	Voiced	Affricate	Palatal
B	β̞	Consonant	Voiced	Approximant	Bilabial
f	f	Consonant	Voiceless	Fricative	Labiodental
T	θ̪	Consonant	Voiceless	Fricative	Interdental
D	ð̪	Consonant	Voiced	Approximant	Dental
s	s	Consonant	Voiceless	Fricative	Alveolar
z	z	Consonant	Voiced	Fricative	Alveolar
jj	ʝ	Consonant	Voiced	Fricative	Palatal
j	j	Consonant	Voiced	Approximant	Palatal
x	x	Consonant	Voiceless	Fricative	Velar
G	ɣ	Consonant	Voiced	Approximant	Velar
l	l	Consonant	Voiced	Lateral	Alveolar
L	ʎ	Consonant	Voiced	Lateral	Palatal
rr	r	Consonant	Voiced	Trill	Alveolar
r	ɾ	Consonant	Voiced	Tap	Alveolar
i	i	Vowel	Unrounded	Close	Front
i_^	i̯	Non-syllabic vowel	Unrounded	Close	Front
e	e	vowel	Unrounded	Mid	Front
a	a	Vowel	Unrounded	Open	Central
o	o	Vowel	Rounded	Mid	Back
u	u	Vowel	Rounded	Close	Back
u_^	u̯	Non-syllabic vowel	Rounded	Close	Back

4.2 Criteria for the visual characteristics

Since there are many handwriting alphabets we took into consideration a cursive alphabet frequently used to teach in Spanish schools (see Fig. 3). In any case, visual features do not change much with a different cursive alphabet.

Table 3 Grapheme to phone correspondences for Spanish

Grapheme	Context	Transcription
⟨a⟩	All contexts	a
⟨b⟩	In word-initial position; after ⟨m⟩ or ⟨n⟩	b
	All other cases	B
⟨c⟩	Followed by ⟨e⟩ or ⟨i⟩	T
	Followed by ⟨b, d, g⟩ (preceding ⟨a, o, u⟩), ⟨m, n, ñ⟩ or ⟨v⟩	G
	All other cases	k
⟨ch⟩	All contexts	tS
⟨d⟩	In word-initial position; after ⟨l⟩, ⟨m⟩ or ⟨n⟩	d
	All other cases	D
⟨e⟩	All contexts	e
⟨f⟩	All contexts	f
⟨g⟩	In word-initial position followed by ⟨r⟩, ⟨l⟩, ⟨a⟩, ⟨o⟩ or ⟨u⟩	g
	After ⟨m⟩ or ⟨n⟩ followed by ⟨a⟩, ⟨o⟩ or ⟨u⟩	
	Followed by ⟨i⟩ or ⟨e⟩	x
	All other cases	G
⟨h⟩	In word-initial position followed by ⟨ie⟩	jj
	All other cases	No sound
⟨i⟩	In nuclear position in the syllable	i
	In non nuclear position in the syllable	i_^
⟨j⟩	All contexts	x
⟨k⟩	All contexts	k
⟨l⟩	All contexts	l
⟨ll⟩	All contexts	L
⟨m⟩	All contexts	m
⟨n⟩	Followed by ⟨p⟩, ⟨b⟩, ⟨v⟩, ⟨m⟩ or ⟨f⟩	m
	Followed by ⟨c⟩, ⟨q⟩ and ⟨a⟩, ⟨o⟩ or ⟨u⟩	N
	All other cases	n
⟨ñ⟩	All contexts	J
⟨o⟩	All contexts	o
⟨p⟩	All contexts	p
⟨qu⟩	All contexts	k
⟨r⟩	In word-initial position; preceded by ⟨l⟩, ⟨n⟩ or ⟨s⟩	rr
	All other cases	r
⟨rr⟩	All contexts	rr
⟨s⟩	Followed by ⟨b⟩, ⟨d⟩, ⟨g⟩, ⟨l⟩, ⟨m⟩, ⟨n⟩ or ⟨r⟩	z
	All other cases	s
⟨t⟩	In syllable-final position	D
	All other cases	t
⟨u⟩	Without diaeresis preceded by ⟨g⟩ or ⟨q⟩	No sound
	In nuclear position in the syllable	u
	In non nuclear position in the syllable	u_^

Table 3 continued

Grapheme	Context	Transcription
⟨v⟩	In word-initial position; after ⟨m⟩ or ⟨n⟩	b
	All other cases	B
⟨w⟩	Followed by a vowel in words of English origin	gu_^
	Followed by a consonant in words of English origin	u_^
	In initial-word position in words of German origin	b
⟨x⟩	All contexts	ks
	In the words *México* and *Texas*	x
⟨y⟩	In word-initial position, after ⟨n⟩ or ⟨l⟩	jj
	After a consonant different from ⟨n⟩ or ⟨l⟩	j
	Preceded and followed by a vowel after a syllable boundary	j
	Preceded or followed by a vowel within the same syllable	i_^
⟨z⟩	All contexts	T

4.3 Criteria for the phonetic transcription

4.3.1 SAMPA symbols and phonetic features

The set of phones considered for the transcription of *DysList* (Table 2) is based on the inventory presented in Llisterri and Mariño (1993). In this proposal, the phones required for the transcription of Spanish were selected after a study of the frequency of occurrence of more than 100,000 segments in a corpus of phonetic transcriptions of semi-spontaneous interviews. The final inventory was established by eliminating all the phones with a frequency of occurrence below 0.1 % in the corpus.

SAMPA (Speech Assessment Methods Phonetic Alphabet) (Wells 2005) symbols were used for the transcription of error and target phones in *DysList*; their correspondence with the symbols from the International Phonetic Alphabet International Phonetic Association (1999) used in this paper is shown in Table 2.

4.3.2 Grapheme to phone correspondences for Spanish

The grapheme to phone correspondences for Spanish (Table 3) followed in the transcription of *DysList* are based on those presented in Llisterri and Mariño (1993), with some modifications. These changes are introduced to take into account more recent works in Spanish phonetics (Gil 2007; Machuca 2000) and the criteria proposed by the Royal Spanish Academy (*Real Academia Española*) (2005).

- instead of following the traditional distinction between semi-vowels and semi-consonants, the realizations of /i/ and /u/ in diphthongs are considered non-syllabic vowels irrespective of their position in the sequence of vowels (Gil 2007). They are represented as [i_^] and [u_^] using the X-SAMPA conventions (Wells 2000) shown in Table 2.

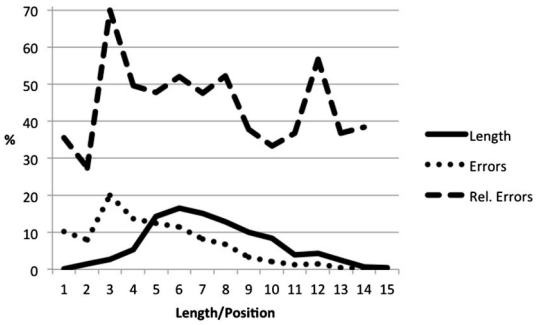


Fig. 4 Percentage distribution of *DysList* word length/position, error positions and relative percentage of errors in each position

- the criteria for the transcription of ⟨w⟩ follow the recommendations found in the *Diccionario panhispánico de dudas* (*Pan-Hispanic Dictionary of Doubts*) Real Academia Española (2005);
- the criteria for the transcription of ⟨x⟩ are based in those presented in Machuca (2000); although the realization of ⟨x⟩ as [s] would be possible when ⟨x⟩ is followed by another consonant such as in *éxtasis ('ecstasy')*, in a formal speaking style ⟨x⟩ would be realized as [ks] in this position. *México* and *Texas*, always pronounced with [x], are treated as exceptions (Machuca 2000);
- the criteria for the transcription of ⟨y⟩ are adapted from those suggested by Gil (Gil 2007), although the plosive realization [ɟ] is not considered. Moreover, the potential alternation between affricate ([dʒ]) and fricative ([j]) realizations of ⟨y⟩ in words such as *yo ('I')* and *cónyuge ('partner')* or of ⟨hi⟩ in words such as *hierba ('grass')* has not been taken into account, since it depends on the speaking style; in these cases, the fricative realization is proposed;
- the digraphs ⟨ch⟩, ⟨ll⟩, ⟨rr⟩, and ⟨qu⟩ are considered as a single unit, since these combination of letters represent a single phone in Spanish: [tʃ], [ʎ], [r] and [k]. The combination ⟨gu⟩ is also treated as a single unit when it is used to represent the sounds [g] or [ɣ̟] (as in *guerra ('war')* [gera] or in *seguir ('follow')* [seɣ̟iɾ]).

The transcriptions correspond to the standard pronunciation of Peninsular Spanish (since this is the variety used by the authors of the texts from which errors were extracted) as described in the classical work by Navarro Tomás (1980) and in more recent contributions such as those of Gil (2007) and Machuca (2000).

It should be reminded that the transcription does not take into account phonetic phenomena that might take place between words; thus, *bata ('dressing gown')* is transcribed as [bata], although [b] would be realized as [β̟] in a sequence such as *mi bata ('my dressing gown')*.

Table 4 Syllable types (left) and number of syllables (right) in words where errors occur

Syllable type	Percentage	No. of syllables	Percentage
CV	37.40	3	33.30
CVC	21.35	2	26.30
None	13.15	4	17.68
CCV	8.20	1	11.87
CVV	7.77	5	7.51
CVVC	6.06	6	3.25
VC	3.67	7	0.09
CCVC	1.54		
V	0.60		
VV	0.09		
CCVV	0.09		
CCVCC	0.09		

5 Analysis of dyslexic errors

In this section we present the results of the analyses of the errors taking into consideration their general linguistic characteristics (Sect. 5.1), as well as their visual and phonetic features (Sects. 5.2 and 5.3, respectively).

5.1 General characteristics

Frequency in the Spanish Web: The target word web frequencies[10] ranged from 190 for *arbolazo ('big tree')*, to 1,389,717,667 for *en ('in')*. The error word web frequencies ranged from 0 for *aczecibilidad (accesibilidad, 'accessibility')*, to 1,178,165,310 for the real-word error **ha (a, 'to')*. On average, correct words were 4.63 times more frequent than words with errors, with the exception of real word errors.

Length and error position: The lengths of the target words range from 1 to 20 characters, with the mode at length 6 and an average length of 7.47 letters. Figure 4 gives the percentage distribution of target word lengths, the percentage distribution of the word positions where the errors appear, and the relative percentage of errors in each position (that is, the percentage of errors in that position with respect to all words that have at least that length). The third position has the highest percentage of errors (70 %), independently of the length of the word.

Syllables: The number of syllables in the words containing spelling errors ranges from one to seven (Table 4, right). In these cases, we observed eleven types of syllables, with the distributions shown in Table 4 (left). 'None' refers to the boundary errors such as **a drede (adrede, 'in purpose')*.

[10] Here we refer to all web pages written in Spanish, not the web pages from Spain. For determining whether a web page was written in Spanish, we used Google Advanced Search settings http://www.google.com/advanced_search.

Table 5 Percentages of error types (left) and frequent specific dyslexic errors (right), within the category *simple errors*

Error type	Percentage	Error type	Letter(s)	Percentage
Substitution	58.84	Substitution	⟨b, v⟩	11.36
Insertion	13.40	Deletion	*space*	6.75
Deletion	26.30	Substitution	⟨g, j⟩	5.46
Transposition	1.45	Deletion	⟨h⟩	4.53
		Insertion	*space*	3.07
		Substitution	⟨c, z⟩	2.82
		Substitution	⟨c, s⟩	2.22
		Deletion	⟨r⟩	2.22
		Insertion	⟨r⟩	2.13

Damerau-Levenshtein distance: In most cases the distance between the misspelled word and the target word is just 1 (73.3 %), with 21.6 % of the cases at distance 2 and only 5.1 % at distance 3 or greater.

Type of error: In Table 5 (left) we give the percentages of every error type. As we can notice, substitution errors are the most frequent ones (near 60 %), while Ramírez and López (2006) state that simple omissions (deletions) are the most frequent kind of error for Spanish. Although dyslexia is popularly known for the transposition errors, less than 1 % of the errors where of this type. This is consistent with Meng et al. (2005) which states that only 30 % of people with dyslexia have trouble with reversing letters and numbers. In Damerau's corpus (1964), 80 % of the misspellings were simple errors.

In our analysis we consider some specific phonetic errors coming from digraphs that represent a single sound in Spanish (such as ⟨ll⟩ and ⟨rr⟩). We found 229 different errors and the most frequent errors (down to 2 %) are shown in Table 5 (right). From this table we can notice that nine different cases of the four error types represent more than 40 % of all errors found, showing the extreme bias of them (*i.e.* less than 4 % of the unique errors cover more than 40 % of the cases). The most frequent case produces more than 11 % of the errors and involve two graphemes that in Spanish have the same phonetic realization, ⟨b⟩ and ⟨v⟩ (which is not the case in English). Analyzing this and other frequent cases, we found three large groups of errors:

- Inserting or deleting a consonant represent 37.9 % of the errors, excluding ⟨h⟩ and ⟨y⟩, which are included in the next cases.
- Deleting or inserting a vowel, including ⟨y⟩ that can have the same phonetic values as ⟨i⟩ in certain contexts, represent 37.5 % of the errors.
- Substituting two letters that might have the same phonetic realization (e.g. ⟨g⟩ and ⟨j⟩) or deleting/inserting an ⟨h⟩, a letter that in Spanish most of the time does not correspond to any sound, represent 15.4 % of the errors.

Table 6 Visual features of the annotated target and error letters (see Table 1)

Visual features	Letters	Percentage
Mirror (type)	Correct	**None** = 26.81, **N** = 57.90, **H** = 3.93, **B** = 11.36
	Error	**None** = 33.39, **N** = 54.74, **H** = 4.01, **B** = 7.86
Mirror (hand)	Correct	**None** = 26.81, **N** = 47.65, **H** = 14.18, **B** = 11.36
	Error	**None** = 33.39, **N** = 39.28, **H** = 19.47, **B** = 7.86
Height (type)	Correct	**None** = 26.81, **A** = 19.04, **N** = 43.81, **D** = 10.33
	Error	**None** = 33.39, **A** = 11.44, **N** = 44.41, **D** = 10.76
Height (hand)	Correct	**None** = 26.81, **A** = 18.53, **N** = 42.70, **D** = 11.44, **B** = 0.51
	Error	**None** = 33.39, **A** = 11.44, **N** = 41.33, **D** = 13.83
Line (type)	Correct	**None** = 33.39, **V** = 0.85, **N** = 58.67, **H** = 13.66
	Error	**None** = 26.81, **V** = 1.11, **N** = 54.48, **H** = 11.02
Line (hand)	Correct	**None** = 26.81, **V** = 0.85, **N** = 71.22, **H** = 1.11
	Error	**None** = 33.39, **V** = 1.11, **N** = 62.43, **H** = 3.07
Rotation (type)	Correct	**None** = 26.81, **Y** = 22.63, **N** = 50.56
	Error	**None** = 33.39, **Y** = 18.19, **N** = 48.42
Rotation (hand)	Correct	**None** = 26.81, **Y** = 30.57, **N** = 42.61
	Error	**None** = 33.39, **Y** = 20.58, **N** = 46.03
Fuzzy (type)	Correct	**None** = 26.81, **Y** = 44.41, **N** = 28.78
	Error	**None** = 33.39, **Y** = 43.47, **N** = 23.14
Fuzzy (hand)	Correct	**None** = 26.81, **Y** = 44.66, **N** = 28.52
	Error	**None** = 33.39, **Y** = 41.59, **N** = 25.02

Notice that these three groups cover more than 90 % of the errors.

We also studied the position of the errors without finding any important preference, although most errors occur within the target word. The four most frequent cases were inserting an ⟨h⟩ at the beginning of the word (3.7 %), substituting ⟨b⟩ by ⟨v⟩ at the first (2.8 %) or third (2.1 %) positions, and inserting an ⟨e⟩ in the second position (2.8 %). Finally, only 8.97 % of the errors were real-word errors.

5.2 Visual features analysis

To analyze the visual features we used Chi-Square goodness of a fit to establish whether or not an observed frequency distribution (in the error letters) statistically differs from a theoretical distribution (the correct letters). In our case, the distribution of error letters differ from the correct letters for typewritten visual features ($\chi^2(9) = 97.67, p < 0.001$) as well as for handwriting visual features ($\chi^2(9) = 377.59, p < 0.001$). See Table 6 for the distribution of the visual features among the error and correct letters.

The distributions of the percentages of correct letters and errors letters in relationship with their mirror visual characteristics were significantly different for the typewriting case ($\chi^2(4) = 55.58, p < 0.001$) as well as for the handwriting case

Table 7 Percentages of errors with mirror visual features

Feature	Mirror (type)				Mirror (hand)			
	Error			Total (%)	Error			Total (%)
	B	H	N		B	H	N	
Correct								
Both	2.17	0	17.70	19.87	2.17	15.53	2.17	19.87
Horizontal	0	0	3.67	3.67	7.51	4.17	11.35	23.03
None	12.35	3.84	60.27	76.46	4.84	12.69	39.57	57.10
Total (%)	14.52	3.84	81.64		14.52	32.39	53.09	

Table 8 Percentages of errors with height visual features

Feature	Height (type)				Height (hand)			
	Error			Total (%)	Error			Total (%)
	D	A	N		D	A	N	
Correct								
Both	0	0	0	0	0	0	1.00	1.00
Descender	11.35	3.51	3.67	18.53	11.52	3.84	5.18	20.54
Ascender	1.00	3.67	18.86	23.53	1.00	3.67	17.86	22.53
None	6.68	10.85	40.40	57.93	12.35	10.52	33.06	55.93
Total (%)	19.03	18.03	62.93		24.87	18.03	57.10	

($\chi^2(5) = 137.83, p < 0.001$). See the contingency table (Table 7) for the percentages. For handwriting almost half of the error letters (46.91 %) had at least one mirror feature.

The percentages of error letters differ from the correct letters by height visual features ($\chi^2(9) = 324.56, p < 0.001$) as well as handwriting visual features ($\chi^2(12) = 244.13, p < 0.001$). See the contingency table (Table 8) for the percentages. Most of the errors occurred in letters with no ascenders neither descenders, 62.93 % for typewriting and 57.10 % for handwriting.

The percentages of error letters differ from the correct letters by line visual features ($\chi^2(9) = 73.29, p < 0.001$) as well as handwriting visual features ($\chi^2(9) = 34.21, p < 0.001$). See the contingency table (Table 9) for the percentages.

The percentages of error letters differ from the correct letters by rotation visual features ($\chi^2(4) = 23.13, p < 0.001$) as well as handwriting visual features ($\chi^2(4) = 32.59, p < 0.001$). See the contingency table (Table 10) for the percentages. If the target letter had a rotation feature this lead to 31.55 % of errors for typewriting and 38.23 % for handwriting.

The percentages of correct fuzzy letters differ from the percentage of error fuzzy letters taking into account both typewriting ($\chi^2(4) = 76.36, p < 0.001$) and handwriting typographies ($\chi^2(4) = 41.10, p < 0.001$). See the contingency

Table 9 Percentages of errors with line visual features

Feature	Line (type)				Line (hand)			
	Error			Total (%)	Error			Total (%)
	H	V	N		H	V	N	
Correct								
Horizontal	5.84	0.17	11.52	17.53	0	0	2.00	2.00
Vertical	0	0	1.67	1.67	0	0	1.67	1.67
None	7.85	1.67	71.29	80.81	5.84	1.84	88.65	96.33
Total (%)	13.69	1.84	84.48		5.84	1.84	92.32	

table (Table 11) for the percentages. Most of the errors occur with fuzzy target letters, 68.95 % for typewriting and 67.11 % for handwriting.

5.3 Phonetic analysis

After transcribing and phonetically annotating each target-error word pair according to the criteria described in Sect. 4.3, we performed an analysis of the errors from the point of view of the phonetic realization of the graphemes involved. Given the prevalence of substitution errors in *DysList* (see Table 5, left), we present the results concerning this category.

5.3.1 Vowels

Overall, vowel substitutions account for 5.38 % ($N = 63$) of the total number of errors in *DysList*. Table 12 shows the percentage of substitution errors for each vowel with respect to the total number of vowel substitutions.

The analysis of the phonetic features associated to each vowel reveals that substitutions involve pairs of phones that share the feature [rounding] in 43.40 % of cases. The second most frequent confusion occurs between vowels that share the features [rounding] and [place] (24.53 %). Substitutions involving the three phonetic dimensions considered are found in 15.09 % of cases, and they correspond to confusions between the graphemes ⟨i⟩ and ⟨y⟩.[11] It is interesting to note that only 15.87 % of the vowel substitution errors correspond to phones that do not have any feature in common.

The pattern arising from the study of the phonetic features is consistent with the most frequent substitutions found in the corpus (Table 12), and suggests a general trend favoring substitutions by graphemes associated with front, mid and unrounded vowels:

[11] In Catalan, the sound [i] is always represented by the letter ⟨i⟩, while in Spanish it might be also represented by ⟨y⟩; moreover, ⟨y⟩ in Catalan is only present in the digraph ⟨ny⟩ used to represent the nasal palatal consonant [ɲ]. Thus, transfer from Catalan might explain the errors.

Table 10 Percentages of errors with rotation visual features

Feature	Rotation (type)			Rotation (hand)		
	Error		Total (%)	Error		Total (%)
	Yes	No		Yes	No	
Correct						
Yes	7.01	24.54	31.55	10.18	28.05	38.23
No	18.03	50.42	68.45	18.20	43.57	61.77
Total (%)	25.04	74.96		28.38	71.62	

[a] ([unrounded]) → [e] ([unrounded] [mid] [front]) (20.63%)
[e] ([unrounded] [mid] [front]) → [a] ([unrounded]) (15.87%)
[i] ([unrounded] [front]) → [e] ([unrounded] [mid] [front]) (7.94%)
[o] ([mid]) → [e] ([unrounded] [mid] [front]) (4.76%)

Substitutions in vowel combinations forming a diphthong account for the 0.94 % ($N = 11$) of the errors found in *DysList*. The most frequent errors in this category—two cases of each—are the substitution of [i̯a] by [ea] and of [i̯o] by [eo]. The highest proportion of errors is observed in target [i̯a] and [o̯e] combinations. In terms of the result of the substitutions, [ea] and [eo] are the two most frequent errors. Given the small size of the sample, no further analysis has been performed, but the trend is coherent with the prevalence of errors involving [e] and [a] described for vowels.

5.3.2 Consonants

Substitution errors in single consonants correspond to the 46.37 % ($N = 543$) of the total number of errors in the resource. They represent, then, the largest category of errors present in *DysList* and are summarized in Table 13.

It can be observed that the most frequent errors in consonants are related to the cases in which a one-to-one correspondence between graphemes and phones is not maintained. This results in two different graphemes having the same phonetic value (see Table 3):

- ⟨b⟩ and ⟨v⟩: both realized as a bilabial plosive [b] or a bilabial approximant [β̞] according to the phonetic context.
- ⟨j⟩ followed by ⟨a⟩, ⟨o⟩ or ⟨u⟩ and ⟨g⟩ followed by ⟨e⟩ or ⟨i⟩: both realized as a velar fricative [x].
- ⟨z⟩ followed by ⟨a⟩, ⟨o⟩ or ⟨u⟩ and ⟨c⟩ followed by ⟨e⟩ or ⟨i⟩: both realized as an interdental fricative [θ̪].
- ⟨c⟩ followed by ⟨a⟩, ⟨o⟩ or ⟨u⟩ and ⟨qu⟩ followed by ⟨i⟩ or ⟨e⟩: both are realized as a velar plosive [k].
- ⟨r⟩ in word-initial position and after nasals or lateral consonants or ⟨s⟩ and ⟨rr⟩ between vowels: both are realized as an alveolar trill [r].

Table 11 Percentages of errors with fuzzy visual features

Feature	Fuzzy (type)			Fuzzy (hand)		
	Error		Total	Error		Total
	Yes	No	(%)	Yes	No	(%)
Correct						
Yes	49.75	16.86	66.61	43.07	18.53	61.60
No	19.20	14.19	33.39	24.04	14.36	38.40
Total (%)	68.95	31.05		67.11	32.89	

Table 12 Percentage of vowel substitutions

	Error							Total (%)
	a	e	i	j̱	o	u	u̱	
Correct								
a	0	20.63	3.17	0	9.52	0	0	33.33
e	15.87	0	4.76	1.59	6.35	0	0	28.57
i	0	7.94	6.35	0	0	0	0	14.29
j̱	0	0	0	6.35	0	0	0	6.35
o	3.17	4.76	0	0	0	1.59	1.59	11.11
u	0	0	1.59	0	3.17	0	0	4.76
u̱	0	0	0	1.59	0	0	0	1.59
Total (%)	19.05	33.33	15.87	9.52	19.05	1.59	1.59	100

This is the reason of the high percentage of errors in target consonants [β̞] (18.23 %), [x] (14.36 %), [θ̪] (12.15 %), and [k] (7.18 %) and also in the consonants resulting from a substitution error: [β̞] (18.42 %), [θ̪] (13.08 %) [x] (10.50 %), and [k] (6.45 %) (Table 13). The lack of biunivocal correspondence between phones and graphemes is also patent in the most frequent confusions in manner of articulation within the class of fricative consonants (24.68 %)—to which [x] and [θ̪] belong—within the group of approximant consonants (20.07 %)—[β̞]—and within plosive consonants (14.55 %)—[k]. Taps and trills are also involved as target phones or as errors, although to a lesser extent. The same trend is observed when place of articulation is considered: the largest number of confusions occur within the class of bilabials (26.70 %)—which includes [β̞]—and inside the group of velars (19.15 %)—which includes [x] and [k]. The interdental consonant [θ̪] appears as the result of substitution errors in 13.08 % of cases and as target phone in 11.97 % of cases.

Confusions between [s] and [θ̪] (4.42 %) and between [θ̪] and [s] (3.31 %) observed in Table 13 might be in part explained by the geolectal phenomenon known as *seseo*, which consists in the systematic substitution of [θ̪] (interdental fricative) by [s] (alveolar fricative) so that [θ̪] is absent from the phonetic inventory of the speakers of the geographic areas in which this phenomenon occurs. The

Table 13 Percentages of consonant substitutions

	Error																					Total (%)
	b	β̪	d̪	ð̪	f	g	ɣ̪	j	ɲ	k	ks	l	ʎ	m	n	p	ɾ	r	s	t̪	θ̪	x	
Correct																						
b	7.73	0	0.37	0	0	0	0	0	0	0	0	0	0	0.18	0	0.18	0	0	0	0	0	8.47
β̪	0.18	16.57	0	0.74	0	0	0	0	0	0	0	0.37	0	0	0.18	0	0.18	0	0	0	0	18.23
d̪	0.37	0	0	0	0	0.18	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0	0.74
ð̪	0	0.74	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0	0.74	0	0	0.37	0	2.03
f	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0	0.92	1.10
g	0	0	0	0	0	0	0	0	0	0.55	0	0	0	0	0	0	0	0	0	0.18	0	0.37	1.10
ɣ̪	0	0.55	0	0	0	0	0.37	0	0	0.18	0	0.18	0	0	0.18	0	0	0	0	0	0	0.55	2.03
j	0	0	0	0	0	0	0.18	0.92	0	0	0	0	2.03	0	0	0	0	0	0	0	0	0	3.13
ɲ	0	0	0	0	0	0	0.37	0.18	2.58	0	0	0	0.18	0	0.37	0	0	0	0	0	0	0.18	3.87
ɟ̪	0	0	0	0	0	0	0	0	0	0	0	0	0.55	0	0	0	0	0	0	0	0	0.37	0.92
k	0	0	0	0	0	0.18	0.18	0	0	3.13	0	0	0	0	0	0.55	0.18	0	0.18	0.18	2.39	0.18	7.18
l	0	0.18	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0.18	0.37	0	0.18	0.37	0	0	1.47
ʎ	0	0	0	0	0	0	0	0.37	0	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0.55
m	0	0	0	0	0	0	0	0	0	0	0	0	0	1.66	0.37	0	0.18	0	0	0	0	0	2.21
n	0	0	0	0	0	0	0	0	0	0	0	0	0	1.10	0	0	1.10	0	0.55	0	0	0	2.76
ŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0	0	0	0.18
p	0.18	0	0	0	0	0	0	0	0	0.18	0	0.18	0	0	0	0	0.18	0	0	0	0	0	0.74
ɾ	0	0.37	0	0	0	0	0	0	0	0	0	0.37	0	0	0.37	0.37	0	0.74	0.55	0.18	0	0	2.95
r	0	0	0	0	0	0	0	0	0	0	0	0	0.37	0	0	0	1.66	1.10	0	0.18	0.18	0	3.50
s	0	0	0	0	0	0	0	0	0	0.74	0	0.37	0	0	0.37	0	0.37	0.18	1.47	0	4.42	0.37	8.29
t̪	0.18	0	0.18	0.18	0	0.18	0	0	0	0	0	0.37	0	0	0	0	0	0	0	0	0	0	1.10
θ̪	0	0	0.37	0.55	0.37	0	0.18	0	0	1.29	0	0	0	0	0	0.18	0.55	0.18	3.31	0.18	4.60	0.37	12.15

Table 13 continued

	Error																					Total (%)
	b	β̪	d̪	ð̪	f	g	ɣ̪	j	ɲ	k	ks	l	ʎ	m	n	p	ɾ	r	s	t̪	θ̪	x	
ʧ	0	0	0	0.18	0	0.18	0.18	0	0	0	0	0	0	0	0	0	0	0	0	0	0.18	0.18	0.92
x	0	0	0	0	0	1.84	3.50	0	0	0.18	0.18	0	0	0	0	0	0.18	0	0.18	0	0.37	7.92	14.36
Total (%)	8.66	18.42	0.92	1.66	0.37	2.58	4.97	1.47	2.58	6.45	0.18	1.84	3.13	3.31	2.03	1.47	5.89	2.21	6.45	1.84	13.08	10.50	100

analysis of features of manner and place also point out in this direction if the substitutions in the class of fricatives and in alveolar and interdental consonants are considered.

The presence of a 3.13 % of cases in which [ʎ] appears as the result of a substitution error and the confusions between [j] and [ʎ] (2.03 %) shown in Table 13 might be partially accounted for by the presence of *yeísmo*, *i.e.* a neutralization of the contrast between [j] (palatal approximant) and [ʎ] (palatal lateral) in favor of [j] which is common in most geographical varieties of Spanish. When substitutions in manner of articulation are considered, 2.58 % of cases of confusions between laterals and approximants are found; part of the substitutions within the class of palatals (6.63 %) may be also accounted for by the presence of *yeísmo*.

The 2.58 % of substitution errors in [ɲ] (palatal nasal) that appear in Table 13 may be explained by the decision taken for the phonetic transcription of the texts concerning a potential transfer from Catalan spelling rules. Since [ɲ] is spelled as ⟨ñ⟩ in Spanish and as ⟨ny⟩ in Catalan, it was considered that both ⟨ñ⟩ and ⟨ny⟩ were intended to represent the palatal nasal consonant.

Almost half of the substitutions found in consonants occur between phones that share their three phonetic features (48.43 %), while confusions between consonants sharing one (19.52 %) or two (26.15 %) features are less commonly encountered. It is worth noting that confusions between consonants that do no have any phonetic feature in common take place in only 5.52 % of cases.

Half of the consonant substitutions in the resource affect simultaneously voicing, manner and place features, a fact to be explained by the spelling irregularities mentioned earlier. When two features are involved in errors, manner and place are simultaneously affected in 16.99 % of cases, and voicing and place in 9.77 % of cases. If the substitution involves only one feature, it can be either place of articulation (9.96 %) or voicing (9.57 %).

In summary, the analysis of consonant substitutions reveals that the spelling mistakes in cases of lack of one-to-one correspondence between phones and graphemes are an important source of confusions within the same class of consonants and are phonetically motivated.

Substitutions affecting combinations of consonants represent a 0.60 % ($N = 7$) of the total number of errors in the collection. More than half of the errors within this category—four cases—correspond to the target sound [ks], spelled as ⟨x⟩ in Spanish. The rest of the errors are found in heterosyllabic clusters formed by a plosive (or their approximant realizations) plus a liquid (*i.e.* a lateral or a rhotic consonant). No further phonetic analysis has been carried out due to the small size of the sample.

6 Conclusions

In this paper we have presented the analysis of a new resource for Spanish, *DysList*, an annotated list of errors written by children and teenagers with dyslexia that is freely available in the Web.[12] The main conclusion is that dyslexic written errors are both phonetically and visually motivated. The phonetic analyses show that the most frequent errors involve letters that do not maintain a one-to-one correspondence between graphemes and phones. Regarding vowels and consonants, most of the errors occur when they share phonetic features. For instance, substitution errors occur most frequently between unrounded vowels than between rounded and unrounded vowels. The visual analyses show how errors involve more frequently letters with similar visual features. For instance, in handwritten letters, 46.91 % of the errors had a mirror feature, 38.23 % had a rotation feature, and 67.11 % are fuzzy letters. Further analyses taking into account each letter frequency would be needed for a stronger assumption.

Our Spanish list of dyslexic errors is still small, but large enough to find insights about dyslexic written errors and to settle the annotation criteria. Another limitation of our study is that we could not compare our resource with one composed of errors written by children without dyslexia. Even if there are child language corpora in Spanish, most of them consist of spoken language transcriptions (see, for example, Fernández 2011; Garrote 2010 and the resources collected in CHILDES); for this reason, we were unable to find a corpus-based list of errors produced by children without dyslexia. However, the resource and the analyses presented are useful by themselves as they allow the development of assistive technologies for people with dyslexia (Korhonen 2008; Li et al. 2013; Pedler 2007). For instance, Pedler (2007) developed a spellchecker for people with dyslexia using a similar language resource for English. In fact, *DysList* has already been used to create language exercises to support the spelling skills of children with dyslexia. These exercises were later integrated in the game *DysEggxia* (*Piruletras* in Spanish) (Rello et al. 2012; Rello et al. 2014), with more that 17,000 downloads. Moreover, these findings have theoretical implications. In cognitive sciences it is intensively debated whether dyslexia has a phonological foundation (Ramus 2003; Snowling 1998) or a visual basis (Franceschini et al. 2012; Vidyasagar and Pammer 2010). Our analyses suggest that both factors have an effect on the written errors produced by people with dyslexia. However, since visually similar letters tend to represent similar sounds, e.g. ⟨p, b, d⟩ are realized as plosive consonants in certain contexts, it is impossible to separate both conditions in an analysis that takes letters into account.

Future work will address the limitations of our current approach by enlarging our collection of errors written by both people with and without dyslexia. This will allow us to carry out more comparisons. First, we shall address a comparison of errors written by people with dyslexia with a control group (texts written by people without dyslexia).[13] Second, since there could be differences between errors made

[12] http://grupoweb.upf.es/WRG/resources/DysWebxia/DysList_resource.csv.gz.

[13] Similarly to the corpus created by Jara for adults in Costa Rica (Jara Murillo 2013), we will create a list of errors from a control group comparable in age and Spanish variant.

when texts are handwritten or typed, future work would also consider typed texts in comparison with handwritten texts. Finally, by enlarging our corpus we will be able to perform a more balanced comparison with the existing English collections. We plan to enlarge our collection by adding more texts written by people with dyslexia and also using the Web as corpus.

Acknowledgments We thank Martí Mayo for computing some of the features and the first author thanks the partial funding of a doctoral fellowship (FI-DGR) of the *Generalitat de Calaluyna* (Government of Catalonia). We thank Yolanda Otal de la Torre, teacher and professional of *CREIX—Centro de Desarrollo Infantil Barcelona* for helping us to collect texts written by people with dyslexia. We also thank the anonymous reviewers for their comments.

Appendix: Comparing English and Spanish errors

We took a subgroup of texts from our corpus composed of 1075 word and performed a comparison of the error distribution with a similar one in English.

English and Spanish languages are archetypes of deep and shallow orthographies, respectively. Along an orthographic transparency scale for European languages, English appears as the language with the deepest orthography and Spanish as the second most shallow one after Finnish (Seymour et al. 2003).

In Tables 14 and 15 we compare the data of the English corpus described in Pedler (2007) with our Spanish texts. We compute the error ratio as the fraction of errors over the correctly spelt words we observe. As expected, Spanish dyslexics make less spelling errors (15 %) than English dyslexics (20 %), due to the different orthographies of their languages. However, the percentage of distinct errors is almost the same.

Table 15 presents the distribution of the different types of dyslexic errors for both languages. To determine if an error was a real world error we checked its existence in the *Diccionario de la lengua española* (*Dictionary of the Spanish Language*) (Real Academia Española 2001), the standard normative dictionary for Spanish.

As expected, there is a greater percentage of multi-errors in a language with deep orthography—English—than in Spanish, e.g. **qría (creía, 'thought')*. However, first letter errors are almost two times more frequent in Spanish, e.g. **tula (ruta, 'way')*. This may look surprising according to Yannakoudakis and Fawthrop (1983), whose findings report that the first letter of a misspelling is correct in the majority of cases,

Table 14 Error ratio and percentage of total errors (with repetitions) and distinct errors in English and Spanish texts written by people with dyslexia

Category	English	Spanish
Total words	3134	1075
Total errors	636	157
Error ratio	0.20	0.15
Distinct errors	577	144
Percentage	90.7	91.7

Table 15 Distribution of errors in English and Spanish corpora

Category	English		Spanish	
	Number	%	Number	%
Simple errors	307	53	96	67
Multi-errors	227	39	33	23
Word boundary errors	47	8	15	10
Real-word errors	100	17	30	21
Non-word errors	477	83	114	79
First letter errors	30	5	16	11
Total	577	100	144	100

but in Spanish the letter ⟨h⟩ at the beginning of a word is not pronounced and this generates many more errors (4.6 %) in that position (see Table 5).

The rest of the dyslexic error types are similar in both languages. There are slightly more real-word errors in Spanish, **dijo (digo, 'said')* or **llegada (llegaba, 'arrived')*. Simple errors are the most frequent ones in both languages. However, each error type has a different frequency. A detailed analysis of the different kind of dyslexic errors and their occurrence in the Web is given in Rello and Baeza-Yates (2012).

Even if both corpora are composed of text written by children with dyslexia in English and Spanish this comparison is not definitive because the two corpora are not fully comparable. For instance, text types and text size, among other characteristics, were not controlled. However, this comparison is still useful for cross-linguistic studies (Brunswick 2010; Goulandris 2003; Seymour et al. 2003), as a preliminary approach for a qualitative cross-linguistic comparison of dyslexic errors written in both languages, as English and Spanish present similar distributions frequencies; and, as expected, differences are due to the different orthographies of the two languages.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Arlington, VA: American Psychiatric Publishing.

Aragón, L. E., & Silva, A. (2000). Análisis cualitativo de un instrumento para detectar errores de tipo disléxico (IDETID-LEA) (Qualitative analysis of an instrument to detect dyslexic errors, IDETID-LEA). *Psicothema*, *12*(Supl. 2), 35–38.

Baeza-Yates, R. & Rello, L. (2011). Estimating dyslexia in the Web. In *Proceedings of the 8th international cross-disciplinary conference on web accessibility (W4A 2011)*. ACM Press, Hyderabad, India.

Brunswick, N. (2010). Unimpaired reading development and dyslexia across different languages. In S. McDougall & P. del Mornay Davies (Eds.), *Reading and dyslexia in different orthographies* (pp. 131–154). Hove: Psychology Press.

Carrillo, M. S., Alegría, J., Miranda, P., & Sánchez, N. (2011). Evaluación de la dislexia en la escuela primaria: Prevalencia en español (Evaluation of dyslexia in primary school: The prevalence in Spanish). *Escritos de Psicología*, *4*(2), 35–44.

Col·legi de Logopedes de Catalunya (2011). PRODISCAT: Protocol de detecció i actuació en la dislèxia. Àmbit educatiu (Protocol for detection and management of dyslexia. Educational scope). Departament d'Educació, Generalitat de Catalunya (Department of Education, Catalan Government), Barcelona.

Coleman, C., Gregg, N., McLain, L., & Bellair, L. W. (2009). A comparison of spelling performance across young adults with and without dyslexia. *Assessment for Effective Intervention*, *34*(2), 94–105.

Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2006). Contribution of lower order skills to the written composition of college students with and without dyslexia. *Developmental Neuropsychology*, *29*(1), 175–196.

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*, 171–176.

Fernández, M. (2011). *Lingüística de corpus y adquisición de la lengua (Corpus linguistics and language acquisition)*. Madrid: Arco/Libros.

Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facoetti, A. (2012). A causal link between visual spatial attention and reading acquisition. *Current Biology*, *22*(9), 814–819.

Garrote, M. (2010). *Los corpus de habla infantil. Metodología y análisis (Children's spoken language corpora. Methodology and analysis)*. Madrid: Ediciones de la Universidad Autónoma de Madrid.

Gelman, I. A., & Barletta, A. L. (2008). A "quick and dirty" website data quality indicator. In *The 2nd ACM Workshop on Information Credibility on the Web (WICOW '08)* (pp. 43–46). Napa Valley, USA.

Gil, J. (2007). *Fonética para profesores de español: de la teoría a la práctica (Phonetics for teachers of Spanish: from theory to practice)*. Madrid: Arco/Libros.

Goulandris, N. (Ed.). (2003). *Dyslexia in different languages: Cross-linguistic comparisons*. London: Whurr Publishers.

Gregor, P., Dickinson, A., Macaffer, A., & Andreasen, P. (2003). Seeword: A personal word processing environment for dyslexic computer users. *British Journal of Educational Technology*, *34*(3), 341–355.

Hernández García, C. (1998). Una propuesta de clasificación de la interferencia lingüística a partir de dos lenguas en contacto: el catalán y el español (A proposal for the classification of linguistic transfer based on two languages in contact: Catalan and Spanish). *Hesperia. Anuario de filología hispánica*, *1*, 61–80.

Holbrook, D. (1964). *English for the rejected: Training literacy in the lower streams of the secondary school*. Cambridge: Cambridge University Press.

Interagency Commission on Learning Disabilities. (1987). *Learning disabilities: A report to the U.S. Congress*. Government Printing Office, Washington, DC.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jara Murillo, C. V. (2013). *COCAE: corpus cacográfico adulto del español de Costa Rica (COCAE: a cacographical corpus of adults in Costa Rica Spanish)*. Universidad de Costa Rica, San José de Costa Rica. http://hdl.handle.net/10669/8928

Korhonen, T. (2008). Adaptive spell checker for dyslexic writers. In *Proceedings of the 11th international conference on Computers Helping People with Special Needs (ICCHP '08)* (pp. 733–741). Berlin, Heidelberg: Springer.

Levenshtein, V. (1965). Binary codes capable of correcting spurious insertions and deletions of ones. *Problems of Information Transmission*, *1*, 8–17.

Li, A. Q., Sbattella, L., & Tedesco, R. (2013). Polispell: An adaptive spellchecker and predictor for people with dyslexia. In S. Carberry, S. Weibelzahl, A. Micarelli, & G. Semeraro (Eds.), *User Modeling, Adaptation, and Personalization* (pp. 302–309). Berlin, Heidelberg: Springer.

Lindgrén, S. A., & Laine, M. (2011). Multilingual dyslexia in university students: Reading and writing patterns in three languages. *Clinical Linguistics & Phonetics*, *25*(9), 753–766.

Llisterri, J. & Mariño, J. B. (1993). *Spanish adaptation of SAMPA and automatic phonetic transcription*. Tech. rep., ESPRIT project 6819 SAM-A Speech Technology Assessment in Multilingual Applications. http://liceu.uab.cat/ joaquim/publicacions/SAMPA_Spanish_93.pdf.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, *53* (1), 1–14.

Machuca, M. J. (2000). Problemas de pronunciación (Pronunciation problems). In S. Alcoba (Ed.), *La expresión oral (Oral expression)* (pp. 71–88). Barcelona: Ariel.

Meng, H., Smith, S., Hager, K., Held, M., Liu, J., Olson, R., et al. (2005). DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proceedings of the National Academy of Sciences*, *102*, 17053–17058.

Mitton, R. (1996). *English spelling and the computer*. Harlow: Longman.

Moats, L. C. (1996). Phonological spelling errors in the writing of dyslexic adolescents. *Reading and Writing*, *8*(1), 105–119.

Navarro Tomás, T. (1980). *Manual de pronunciación española (Manual of Spanish pronunciation)* (20th ed.). Madrid: Consejo Superior de Investigaciones Científicas. (Original work published in 1918).

Orton Dyslexia Society Research Committee. (1994). Operational definition of dyslexia. In C. Scruggs (Ed.), *Perspectives*, *20*(5), 4.

Pedler, J. (2007). Computer correction of real-word spelling errors in dyslexic text. Ph.D. thesis, Birkbeck College, London University.

Piskorski, J., Sydow, M., & Weiss, D. (2008). Exploring linguistic features for web spam detection: a preliminary study. In *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb '08)* (pp. 25–28). New York, NY: ACM Press.

Pujol, M. (2004) Análisis de errores grafemáticos en textos libres de estudiantes de enseñanzas medias (Analysis of graphematic errors in free texts by secondary school students). Ph.D. thesis, Departament de Didàctica de la Llengua i la Literatura, Universitat de Barcelona, Barcelona. http://www.tdx.cat/handle/10803/1283.

Ramírez, F., & López, E. (2006). Spelling error patterns in Spanish for word processing applications. In *Proceedings of the 5th international conference on language resources and evaluation (LREC 2006)* (pp. 93–98). Genoa, Italy.

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*(2), 212–218.

Real Academia Española. (2001). *Diccionario de la lengua española (Dictionary of the Spanish Language)* (22nd ed.). Madrid: Espasa-Calpe.

Real Academia Española. (2005). *Diccionario panhispánico de dudas (Pan-Hispanic Dictionary of Doubts)*. Madrid: Santillana.

Rello, L., & Baeza-Yates, R. (2012). The presence of English and Spanish dyslexia in the Web. *New Review of Hypermedia and Multimedia*, *8*, 131–158.

Rello, L., Baeza-Yates, R., & Llisterri, J. (2014). DysList: An annotated resource of dyslexic errors. In *Proceedings of the 9th international conference on language resources and evaluation (LREC 2014)* (pp. 1289–1296). Reykjavik, Iceland.

Rello, L., Baeza-Yates, R., Saggion, H., & Pedler, J. (2012). A first approach to the creation of a Spanish corpus of dyslexic texts. *LREC Workshop Natural Language Processing for Improving Textual Accessibility (NLP4ITA)* (pp. 22–27). Istanbul, Turkey.

Rello, L., Bayarri, C., & Gorriz, A. (2012). What is wrong with this word? Dyseggxia: A game for children with dyslexia (demo). In *Proceedings of the 14th international ACM SIGACCESS conference on computers and accessibility (ASSETS '12)* (pp. 219–220). Boulder, USA: ACM Press.

Rello, L., Bayarri, C., Otal, Y., Pielot, M. (2014). A computer-based method to improve the spelling of children with dyslexia. In *Proceedings of the 16th international ACM SIGACCESS conference on computers & accessibility, ASSETS '14* (pp. 153–160). Rochester, NY, USA, October 20–22, 2014.

Romani, C., Ward, J., & Olson, A. (1999). Developmental surface dysgraphia: What is the underlying cognitive impairment? *The Quarterly Journal of Experimental Psychology*, *52*(1), 97–128.

Schulte-Körne, G., Deimel, W., Müller, K., Gutenbrunner, C., & Remschmidt, H. (1996). Familial aggregation of spelling disability. *Journal of Child Psychology and Psychiatry*, *37*(7), 817–822.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Snowling, M. (1998). Dyslexia as a phonological deficit: Evidence and implications. *Child and Adolescent Mental Health*, *3*(1), 4–11.

Spooner, R. (1998). A spelling aid for dyslexic writers. Ph.D. thesis, University of York

Sterling, C., Farmer, M., Riddick, B., Morgan, S., & Matthews, C. (1998). Adult dyslexic writing. *Dyslexia*, *4*(1), 1–15.

Toro, J., & Cervera, M. (1984). *TALE: Test de Análisis de Lectoescritura (TALE: Literacy Analysis Test)*. Madrid: Visor.

Treiman, R. (1997). Spelling in normal children and dyslexics. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 191–218). Mahwah, NJ: Lawrence Erlbaum.

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.

Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: A deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Sciences*, *14*(2), 57–63.

Wells, J. C. (2000). *Computer-coding the IPA: A proposed extension of SAMPA*. Division of Psychology and Language Sciences, University College London, London. http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm.

Wells, J. C. (2005). *SAMPA—Computer readable phonetic alphabet*. Division of Psychology and Language Sciences, University College London, London. http://www.phon.ucl.ac.uk/home/sampa/index.html.

World Health Organization. (1993). *International statistical classification of diseases, injuries and causes of death (ICD-10)* (10th ed.). Geneva: World Health Organization.

Yannakoudakis, E. J., & Fawthrop, D. (1983). The rules of spelling errors. *Information Processing & Management*, *19*(2), 87–99.