UNIVERSIDAD DE OVIEDO

THERE ARE PHONETIC PATTERNS IN VOWEL SUBSTITUTION ERRORS IN TEXTS WRITTEN BY PERSONS WITH DYSLEXIA
Luz Rello[1] & Joaquim Llisterri[2]

[1]Universitat Pompeu Fabra, [2]Universitat Autònoma de Barcelona

Abstract. In this paper we describe the creation of a corpus of texts written by Spanish speaking children with dyslexia, establish a typology of errors found in the corpus and explore the phonetic basis of vowel substitution errors. The results of the analysis reveal that articulatory phonetic dimensions and frequency of occurrence are relevant factors to be taken into account to explain vowel substitution errors.

1. Introduction

The analysis of spelling errors produced by persons with dyslexia provides several kinds of relevant information, both from a theoretical and from an applied perspective. These errors have been traditionally studied considering only the graphical representation of the language. Since orthography is, to a certain extent, a reflection of the phonetics and the phonology of a language, it seems that an approach to spelling errors considering their relationship with the phonetic values associated to each grapheme may provide new insights into the nature of the deficit in the phonological component that is the cause of some of the difficulties experienced by persons with dyslexia.

In this work we present an attempt to analyze vowel substitutions found in a corpus of Spanish texts written by children with dyslexia, taking into account the phonetic nature of the errors. First, we present a brief characterization of dyslexia (Section 2), followed by a discussion of the relevance of errors produced by persons with dyslexia (Section 3). In Section 4, the corpus of Spanish texts written by children with dyslexia is described, and in Section 5 we put forward a typology of errors based on the corpus and provide the frequency of occurrence of each class of errors. A phonetic analysis of vowel substitution errors is presented in detail in Section 6. Finally, in Section 7 we summarize the conclusions of the study.

2. The concept of dyslexia

Dyslexia is a specific learning disability of neurological origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge (Lyon, Shaywitz, & Shaywitz, 2003).

---

[1] Correspondence to: Luz Rello, Natural Language Processing and Web Research Groups, Tànger 122, Office 55.422, 08018, Barcelona, Spain. E-mail: luzrello@acm.org

Although some authors refer to dyslexia as a specific reading disability (Vellutino, Fletcher, Snowling, & Scanlo, 2004) and to dysgraphia as its manifestation in writing (Romani, Ward, & Olson, 1999),[2] our study follows the standard classifications of ICD-10 (World Health Organization, 1993) and DSM-IV (American Psychiatric Association, 2000) in which dyslexia is listed as a reading and spelling disorder. Persons with dyslexia exhibit higher spelling error rates than non-dyslexic ones (Coleman, Gregg, McLain, & Bellair, 2009) and errors attributed to phonological impairment, to spelling knowledge and to lexical mistakes are also more frequent in persons with dyslexia (Sterling, Farmer, Riddick, Morgan, & Matthew, 1998).

Despite its universal neurocognitive basis, the manifestation of dyslexia varies among languages (Lindgrén & Laine, 2011) and cultures (Goulandris, 2003), but also among subjects and ages; for instance, the misspelling rate in children with dyslexia is higher than in adults (Sterling et al., 1998).

Language-specific variability is related to differences in the degree of consistency and regularity across spelling systems (Brunswick, 2010). English has an opaque –or deep– orthography, in which the correspondences between letters and sounds are inconsistent and many exceptions are permitted. For this reason, English presents a significantly greater challenge to the beginning reader than other languages, such as Spanish, with more regular and consistent mappings between letters and sounds, that is, a transparent –or shallow– orthography.

Due to cross-linguistic differences, the estimations on the prevalence of dyslexia may not coincide for different languages. As far as English is concerned, the Interagency Committee on Learning Disabilities reports that 10-17.5% of the population in the U.S.A. has dyslexia (ICOL, 1987); the model of Shaywitz, Escobar, Shaywitz, Fletcher, and Makuc (1992) predicts that 10.8% of English speaking children are dyslexic, while in Katusic, Colligan, Barbaresi, Schaid, and Jacobsen (2001) the rates vary from 5.3% to 11.8% depending on the formula used in the calculations; finally, Brunswick (2010) provides a 10% estimate for English. Data on the prevalence of dyslexia in Spanish speakers are much more scarce: Galván (2010) reports a 7.5% among school children in Perú; Carrillo, Alegría, Miranda, and Sánchez (2011) found that 11.8% of the school children examined in Murcia (Spain) exhibited difficulties associated to dyslexia and Jiménez, Guzmán, Rodríguez, and Artiles (2009) report an 8.6% for a similar population in the Canary Islands (Spain). As in the case of English, differences in the figures are due to the assessment methods applied and to the definition of dyslexia considered by the authors of the studies.

3. The relevance of dyslexic errors

Errors in the written production of persons with dyslexia have generally been used as a source of knowledge, since the specific types of dyslexic errors highlight different aspects of dyslexia (Treiman, 1997) such as a phonological processing deficit (Lindgrén & Laine, 2011; Moats, 1996). Moreover, diagnosis of dyslexia may be based on spelling scores (Schulte-Körne, Deimel, Müller, Gutenbrunner, & Remschmidt, 1996). Errors produced by persons with dyslexia have been also used for various accessibility related purposes such as the development of tools like spell checkers

---

[2] Dysgraphia refers to a writing disorder associated with the motor skills involved in writing, handwriting and sequencing, but also with orthographic coding (Romani et al., 1999). It is a medical condition that co-occurs with dyslexia (Nicolson & Fawcett, 2011).

(Pedler, 2007) or word processors (Gregor, Dickinson, Macaffer, & Andreasen, 2003).

Other applications include the estimation of the presence of dyslexia in the Web: among the different kind of errors found in the Web and in the social media (Rello & Baeza-Yates, 2012), at least 0.67% of them are specific to users with dyslexia (Baeza-Yates & Rello, 2011).

Although dyslexic error analysis is crucial for the applications mentioned above, to the best of our knowledge, dyslexic errors in Spanish have been studied using lists of errors, but not entire texts or corpora produced by persons with dyslexia (Aragón & Silva, 2000; Baeza-Yates & Rello, 2011). Our analysis will be based on a corpus of texts written by children with dyslexia (see Section 4) collected for the purpose of the present study and for future work on errors found in Spanish-speaking persons with dyslexia.

4. A Spanish corpus of texts written by dyslexic children

We have already mentioned in Section 2 that the manifestations of dyslexia differ among languages, individuals and age (Goulandris, 2003; Sterling et al., 2003; Vellutino et al., 2004). Due to this variability, our aim was to collect an homogeneous corpus of texts written by a similar population in terms of age, education, native language and diagnosed dyslexia.

We collected 16 Spanish texts produced by different children with dyslexia from 13 to 15 years old who have Spanish as their first language. The texts consist of homework writing exercises. They 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. An example of a transcribed story included in the corpus is given in Figure 1.

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

Figure 1. Transcription of a story written by a 14 years old child with dyslexia[3].

___

[3] Approximated literal translation: "A famous biologist, who lived in Bordeaux and was great-grandson of who probably was one of the wealthiest barons in France and 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 wise ideas to solve

Figure 1 exemplifies errors of all possible kinds: (i) substitution: *i (y)*, *realigaba (realizaba)*, *qreía (creía)*, *savias (sabias)*, *budú (vudú)*, *venerosas (venenosas)* and *baubab (baobab)*; (ii) insertion: *comprós (compró), *unos (uno)*, ; (iii) omission: *expermentos (experimentos)*, *beneficirio (beneficiario)*, *nausabundas (nauseabundas)*, and *del (de)*; and (iv) transposition: *pobrablemente (probablemente)*.

The average length per text is 67 words and the total corpus size amounts to 1,057 words. The reduced size of our Spanish corpus is explained by the difficulty in finding texts written by persons diagnosed with dyslexia. However, we believe that a corpus of these characteristics is valuable to analyze dyslexic errors in Spanish. The comparison of our corpus with one of similar characteristics for English, the Pedler's initial corpus (Pedler, 2007), reveals that both corpora are similar in terms of the type of errors detected (Rello, Baeza-Yates, Saggion, & Pedler, 2012).

5. A typology of errors produced by persons with dyslexia

As a first step in the analysis, errors found in the corpus were classified using three main criteria: (1) the degree of difference between the word actually written and the intended one (target word); (2) the correspondence with existing words; and (3) boundary letter errors. In this section we present the classification in detail and exemplify each error class with data from our corpus.

(1) Degree of difference between the error and the target word:

(1.1) Simple errors: the erroneous word differs from the intended word by a single letter. Simple errors can be due to:

(1.1.1) Substitution: *ja (ya)* 'already';
(1.1.2) Insertion: *sigilosarmente (sigilosamente)* 'silently';
(1.1.3) Omission: *imovilizó (inmovilizó)* 'immobilized';
(1.1.4) Transposition[4]: *ferai (feria)* 'fair'.

(1.2) Multi-errors: the erroneous word differs in more than one letter from the target word. The results of some errors, such as *guapoisismo (guapísimo)* 'gorgeous', closely resemble the intended word, while others are not so obvious: *Ilistnador (ilustrador)* 'illustrator'.

(1.3) Word boundary errors: these are mistakes (run-ons and split words) which are special cases of omission and insertion errors. A run-on is the result of omitting a space, such as *devidrio (de vidrio)*, 'of glass'. A split word occurs when a space is inserted in the middle of a word, such as in *per sona (persona)* 'person'.

(2) Correspondence with existing words:

---

health problems inspired by African voodoo, preparing nauseating infusions based on boiled baobab barks and skin of poisonous snakes."

[4] The standard definition of edit distance (Levenshtein, 1965) considers transpositions as two errors, while Damerau (1964) defines transposition as a single error.

(2.1) Real-word errors: misspellings that result in another correct word. For instance, *medias* 'means', being the intended word *mediados* 'mids'.

(2.2) Non-word errors: misspellings that do not result in another correct word.

(3) Boundary letter errors:

(3.1) First letter errors: they occur in cases such as *tro (otro)* 'other' in which there is an omission of the first letter of a word.

(3.2) Last letter errors: an omission is found in the last letter of a word, such as *comprós (compró),* 'bought'*.

Table 1 summarizes the frequency distribution of the errors found in the corpus, classified according to the criteria used in the analysis. Frequencies for word boundary errors and first and last letter errors are included in simple and multiple errors.

Category		Percentage
Multi-errors (23%) Simple errors (67%)	Substitution	62.5%
	Omission	20.6%
	Insertion	14.7%
	Transposition	2.2%
Boundary letter errors (26%)	First	39%
	Last	61%

Table 1. Frequency distribution of the errors found in the corpus.

6. Phonetic patterns in vowel substitution errors

Although it is widely recognized that dyslexia is associated to the phonological component of the language (see Section 2), the analysis of errors is usually performed in terms of graphemes, the units of the written language. In this section we attempt an analysis in terms of the phonetic realization of the graphemes, since the knowledge of the correspondence between sounds and letters is also part of the linguistic competence of our subjects.

In this first approach, the analysis focus on vowels and, more specifically, on substitution errors, since this category accounts for 62.5% of the simple errors found in the corpus and simple errors are the most common type of problem encountered (Table 1). We have only considered 22 vowel substitutions that involved different phonetic realizations; thus, confusions between <y> and <i> have not been taken into account, since both graphemes are used to represent the same phoneme /i/[5].

Spanish has a phonological five vowel system (/i/, /e/, /a/, /o/, /u/) with a regular one-to-one correspondence between phonemes and graphemes, except in the case

---

[5] Although Spanish is the first language of the children who produced the texts, they are also taught Catalan at school. Since in Catalan the connector 'and' (Spanish *y*) is written as *i*, cross-lingual interference may not be excluded when explaining *i (y)* substitution errors.

of /i/, which can be spelled either <i> or <y>. Traditional phonological analysis (Quilis, 1993) assumes that non syllabic glides [i̯] (*aire, hoy*), [j] (*Asia*), [u̯] (*auto*) and [w] (*bueno*) are allophones of /i/ and /u/; for this reason, we will consider them together with the syllabic vowels in our analysis. Other allophones, in particular open and close variants of /e/ and /o/ have been described, although they appear in free variation or are the expected result of coarticulatory processes.

Table 2 reveals that /a/ is the target vowel more frequently involved in substitution errors (36,36% of cases), followed by /o/ (27.27%), /u/ (13.64%) and /i/ (13.64%) and, finally, /e/ (9.09%). These errors result in the appearance of /e/ more than half of the times (59.09% of cases), followed by /o/ (22.73%) and, to a lesser extent, by /a/ (9.09%), /i/ (4.55%) and /u/ (4.55%).

	/a/	/e/	/i/	/o/	/u/
Target	36.36	9.09	13.64	27.27	13.64
Error	9.09	59.09	4.55	22.73	4.55

Table 2. Target and error vowels involved in vowel substitutions (in percentages relative to the total number of vowel substitution errors).

A closer analysis of the substitutions (Table 3) shows that /i/ is systematically replaced by /e/; also /a/ and /o/ are frequently replaced by /e/, while /u/ is changed to /o/ in two thirds of the cases.

Target	Error					
	/a/	/e/	/i/	/o/	/u/	*N*
/a/		75	0	25	0	8
/e/	50		0	50	0	2
/i/	0	100		0	0	3
/o/	16.7	66.7	0		16.7	6
/u/	0	0	33.3	66.7		3

Table 3. Vowel substitutions (in percentages relative to the total number of vowel substitution errors).

Phonetic and phonological contrasts among Spanish vowels are structured along three articulatory dimensions: place of articulation, degree of opening and lip rounding. As far as place of articulation is concerned, /i/ and /e/ are front vowels, /a/ is central and /o/ and /u/ are back vowels. In terms of opening or tongue height, /i/ and /u/ are close (high), /e/ and /o/ are mid and /a/ is open (low). Back vowels /u/ and /o/ are always produced with lip rounding, whereas in /i/, /e/ and /a/ lips are spread.

The most frequent substitution errors found in the corpus are those involving simultaneously place of articulation and degree of opening (31.82%, *N*=7), followed by those concerning degree of opening alone (27.27%, *N*=6) and by those in which place of articulation and lip rounding are simultaneously involved (27.27%, *N*=6); substitutions concerning the three dimensions are found only in 13.64% of cases (*N*=3).

Confusions among vowels sharing one phonetic feature appear in 59.09% of substitution errors (*N*=13), while vowels sharing two phonetic features are involved in 27.27% of errors (*N*=6); only 13.64% of cases (*N*=3) of confusions between vowels that do not have any phonetic feature in common are found in the corpus. Overall, graphemes corresponding to vowels sharing phonetic features are confused in 86.29% of cases.

Table 4 shows that in all cases of confusions among vowels sharing two phonetic features, place of articulation and rounding are always simultaneously involved; when vowels in confusions share only one feature, opening or rounding are the intervening dimensions.

Number of shared features	Opening	Rounding	Place and rounding	*N*
1	46.2	53.8	0	13
2	0	0	100	6

Table 4. Phonetic dimensions involved in vowel substitution errors (in percentages relative to the total number of vowel substitution errors).

In order to assess the role of each of the articulatory dimensions in vowel confusion errors, we have carried out separate analysis. Tables 5 and 7 present the global results, in which substitutions within the same category (e.g., between /i/ and /e/, both of them front vowels) are not being treated as errors; in tables 6, 8 and 9, these kind of substitutions have been considered to be an error.

Table 5 shows that, globally, back vowels are predominantly replaced by front ones (83,33%) and the central vowel is also most frequently changed into front vowels (75%). The more detailed analysis in Table 6 reveals that front vowels are confused with other front vowels in 60% of cases involving place of articulation errors. As a general trend, when substitutions occur, errors in place of articulation result in graphemes associated with front vowels; this is coherent with the results observed in Table 3: target /a/ (central) is written with the grapheme corresponding to /e/ (front) in 75% of cases and target /o/ (back) is written with the grapheme for /e/ in 66.7% of cases.

Target	Error			
		Back	Central	Front	N Errors
	Back		16.67	83.33	6
	Central	25		75	8
	Front	50	50		2

Table 5. Vowel substitutions in place of articulation (in percentages relative to the total number of errors in place of articulation).

Target	Error			
		Back	Central	Front	*N Target*
	Back	33.3	11.1	55.6	9
	Central	25		75	8
	Front	20	20	60	5

Table 6. Vowel substitutions in place of articulation (in percentages relative to the total number of vowel substitution errors).

The global analysis of the degree of opening (Table 7) reveals that both open and close vowels are systematically replaced by mid ones. Table 8 shows that mid vowels are also often confused with other mid vowels (62.5% of cases). A tendency to use graphemes representing mid vowels is then observed. This is in agreement with the fact that, as shown in Table 3, in confusion errors /i/ (close) is always written with the grapheme corresponding to /e/ (mid) and /a/ is also written with the grapheme for /e/ in 71.4% of cases.

Target	Error			
		Close	Mid	Open	*N Errors*
	Close		100	0	5
	Mid	33.33		66.67	3
	Open	0	100		8

Table 7. Vowel substitutions in degree of opening (in percentages relative to the total number of errors in the degree of opening).

Target	Error			
		Close	Mid	Open	*N Target*
	Close	16.7	83.3	0	6
	Mid	12.5	62.5	25	8
	Open	0	100		8

Table 8. Vowel substitutions in degree of opening (in percentages relative to the total number of vowel substitution errors).

Finally, confusions concerning lip rounding (Table 9) reveal that unrounded vowels are frequently replaced by other unrounded vowels and rounded vowels are also often changed into spellings corresponding to unrounded vowels. This is consistent with the high percentage of target /o/ (rounded) written with the grapheme corresponding to /e/ (unrounded) that can be observed in Table 3.

Target	Error		
	Rounded	Unrounded	*N Target*
Rounded	33.3	66.7	9
Unrounded	23.1	76.9	13

Table 9. Vowel substitutions in lip rounding (in percentages relative to the total number of vowel substitution errors).

Taken together, these results suggest a general trend favoring replacements by graphemes associated with front, mid and unrounded vowels. According to the data presented in Rojo (1991) and summarized in Table 10, front vowels are more frequent in Spanish than central and back ones, mid vowels are more frequent than open and close ones and unrounded are more frequent than rounded vowels.

	Frequency of occurrence		Frequency of occurrence		Frequency of occurrence
Back	26.94	Close	22.61	Rounded	26.94
Central	28.56	Mid	77.38	Unrounded	73.05
Front	44.58	Open	28.56		

Table 10. Frequency of occurrence of Spanish vowels (relative to the number of vowels) adapted from Rojo (1991).

The most frequent vowels in Spanish are /a/ and /e/, each of them accounting for 28.56% of the vowels in the corpus examined by Rojo (1991). The fact that /e/ appears in 59.09% of substitution errors (see Table 2) might not be unrelated to this fact. Along the same line, the second most frequent vowel appearing in substitution errors is /o/ (22.73%), which, after /a/ and /e/, is the third more frequent vowel in Spanish.

7. Conclusions

As explained in Section 3, to the best of our knowledge, no previous analysis of errors produced by Spanish speaking dyslexic persons have been carried out using a corpus-based approach. This methodology, widely used in other disciplines, has enabled us to propose a typology of errors based on three criteria: (1) the degree of difference between the word found in the corpus and the target word; (2) the correspondence between existing words and the words resulting from spelling errors; and (3) errors occurring in word-boundary position.

Simple substitution errors are the most frequent ones (62,5%) when the degree of difference between erroneous words and target words is considered; run-ons are more frequent than splits (84% vs. 16%) and errors are more common in the last letter of a word than in the initial one (61% vs. 39%).

The phonetic analysis of vowel substitution errors reveals that more than half of the confusions occur between vowels sharing one phonetic feature, while almost no confusions take place between vowels which do not have any phonetic dimension in common. Overall, graphemes corresponding to vowels sharing phonetic features are confused in 86% of cases.

As a general trend, errors in place of articulation result in graphemes associated with front vowels; target /a/ (central) tends to be written with the grapheme corresponding to /e/ (front) and target /o/ (back) tends to be spelled with the grapheme for /e/. As for the degree of opening, a tendency to use graphemes representing mid vowels is observed: /i/ (close) is always written with the grapheme corresponding to /e/ (mid) and /a/ (open) is frequently written with the grapheme for /e/. Concerning lip rounding, unrounded vowels are frequently replaced by other unrounded vowels and rounded vowels are also often changed into spellings corresponding to unrounded vowels: target /o/ (rounded) is often written with the grapheme corresponding to /e/ (unrounded). These results suggest a general trend favoring substitutions by graphemes associated with front, mid and unrounded vowels. The frequency of occurrence of vowels in Spanish might help to explain some of the phenomena observed.

Although the type of errors attested in our Spanish corpus is similar to those found in a similar corpus for English, further work will aim at extending the corpus. Moreover, a phonetic analysis of substitution errors in consonants will be attempted along the same lines as the one presented here for vowels. We hope that the combination of a corpus-based approach with an analysis that takes into consideration the phonetic representation associated to the graphic units of the language will contribute to a better understanding of the nature of errors produced by Spanish-speaking persons with dyslexia.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders (DSM-IV-TR)* (4th ed.). Arlington, VA: American Psychiatric Publishing.

Aragón Borja, L. A., & Silva Rodríguez, A. (2000). Análisis cualitativo de un instrumento para detectar errores de tipo disléxico (IDETID-LEA). *Psicothema*, *12*(2, Supl. 2), 35-38.

Baeza-Yates, R., & Rello, L. (2011). Estimating dyslexia in the Web. In *W4A 2001. 8th International Cross-Disciplinary Conference on Web Accessibility.* Hyderabad, Andhra Pradesh, India. March 28 - 29, 2011. (pp. 1-4). New York: ACM Press.

Brunswick, N. (2010). Unimpaired reading development and dyslexia across different languages. In N. Brunswick, S. McDougall, & P. de Mornay Davies (Eds.), *Reading and dyslexia in different orthographies.* (pp. 131-54). Hove: Psychology Press.

Carrillo, M. S., Alegría, J., Miranda, P., & Sánchez Pérez, N. (2011). Evaluación de la dislexia en la escuela primaria: Prevalencia en español. *Escritos de Psicología, 4*(2), 35-44.

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.

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*(3), 171-176.

Galván Gómez, R. (2010). Dislexia va en aumento. *El Peruano*, 8 de enero de 2010, p. 22.

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.

Interagency Committee on Learning Disabilities (1987). *Learning disabilities: A report to the U.S. Congress.* Bethesda, MD: National Institutes of Health.

Jiménez, J. E., Guzmán, R., Rodríguez, C., & Artiles, C. (2009). Prevalencia de las dificultades específicas de aprendizaje: La dislexia en español. *Anales de Psicología, 25*(1), 78-85.

Katusic, S. K., Colligan, R. C., Barbaresi, W. J., Schaid, D. J., & Jacobsen, S. J. (2001). Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn. *Mayo Clinic Proceedings*, *76*(11), 1081-1092.

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

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.

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

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

Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, *47*(1), 117-127.

Pedler, J. (2007). *Computer correction of real-word spelling errors in dyslexic text.* PhD Thesis, Birkbeck College, London University.

Quilis, A. (1993). *Tratado de fonología y fonética españolas.* Madrid: Gredos.

Rello, L., & Baeza-Yates, R. (2012). Social media is not that bad! The lexical quality of social media. In *ICWSM 2012. The 6th International AAAI Conference on Weblogs and Social Media.* Trinity College, Dublin, Ireland. June 4, 2012.

Rello, L., Baeza-Yates, R., Saggion, H., & Pedler, J. (2012). A first approach to the creation of a Spanish corpus of dyslexic texts. In *NLP4IT. Workshop on Natural Language Processing for Improving Textual Accessibility in LREC 2012, 8th International Conference on Language Resources and Evaluation.* Istanbul, Turkey. May 27, 2012*.*

Rojo, G. (1991). Frecuencia de fonemas en español actual. In M. Brea & F. Fernández Rei (Eds.), *Homenaxe ó profesor Constantino García.* (pp. 451-67). Santiago de Compostela: Servizo de Publicacións e Intercambio Científico, Universidad de Santiago de Compostela.

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.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*(3), 145-150.

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

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 Associates.

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.

World Health Organization (1993). *International classification of diseases and related health problems (ICD-10)* (10th ed.). Geneva: World Health Organization.