

A longitudinal study of L2 historical writing: lexical richness and writing proficiency in Content and Language Integrated Learning

Adrián Granados, María Dolores López-Jiménez & Francisco Lorenzo

Universidad Pablo de Olavide (Spain)

agranav@upo.es, mdlopezji@upo.es, fjlorber@upo.es

Abstract

The aim of this paper is to offer a much-needed longitudinal description of lexical richness in the L2 historical writing of CLIL bilingual secondary school students over a three-year period. The automated tool Coh-Metrix 3.0 was used to analyse the evolution of the lexical diversity, density and sophistication of a learner corpus made up of 75 history essays composed by the same 15 students as part of their L2 (English-taught) history lessons. The results show an increase in the number of lexical items employed by the students and in the abstractness and associability of these items. This indicates that students improved their lexical richness, while developing their writing proficiency and history literacy skills.

Keywords: lexical richness, L2 writing proficiency, history literacy, Coh-Metrix, CLIL.

Resumen

Estudio longitudinal de la escritura histórica en una L2: la riqueza léxica y la competencia escrita en el Aprendizaje Integrado de Contenidos y Lenguas Extranjeras

El propósito de este estudio es ofrecer una descripción longitudinal de la riqueza léxica en el discurso sobre Historia en L2 de estudiantes de enseñanza secundaria bilingüe AICLE a lo largo de tres años. Se ha empleado la herramienta computacional Coh-Metrix 3.0 para analizar la evolución de la diversidad, densidad y sofisticación léxicas de un corpus de aprendices

compuesto por 75 textos producidos por 15 estudiantes en sus clases de Historia en inglés como L2. Los resultados muestran un incremento en la cantidad de unidades léxicas empleadas por los estudiantes, así como en la abstracción y asociabilidad de los términos, lo cual indica que los estudiantes mejoraron su riqueza léxica, a la vez que desarrollaron su competencia escrita y su literacidad histórica.

Palabras clave: riqueza léxica, competencia escrita en la L2, literacidad histórica, Coh-Metrix, AICLE.

1. Introduction

The writing-to-learn approach (e.g., Britton, Martin & Rosen, 1966) and its secondary trends, ‘writing across the curriculum’ (e.g., Young & Fulwiler, 1986) and ‘writing in the disciplines’ (e.g., Deane & O’Neill, 2011), share the assumption that disciplinary content is deeply ingrained in the very act of literacy development. Furthermore, they are concerned with disciplinary discourses and their enactment in genres, which have their own particular characterisation across all language levels (Rose, 2008; Shanahan & Shanahan, 2008). These trends as a whole have paved the way for the current consideration of disciplinary content and writing conventions as focal points in literacy research, so much so that *disciplinary literacy* is often added to the traditional distinction between *basic interpersonal communicative skills* (BICS) and *cognitive academic language proficiency* (CALP) (Dressen-Hammouda, 2008; Harwood & Hadley, 2004; see Cummins, 1979, for BICS and CALP). In this study, the focus is on historical literacy, and the writing proficiency of a group of students is examined regarding the specialised language of history.

Writing proficiency is a subset of language competence in which the mastery of genres and rhetorical devices should be combined with language-specific abilities, such as the use of a range of vocabulary and syntactic structures (Wolfe-Quintero, Tnagaki and Kim, 1998; in Lahuerta, 2015). Indeed, the linguistic features employed by most writing researchers fall into three areas: lexis, syntax and cohesion (McNamara, Crossley & McCarthy, 2010). Of these constructs, lexical richness is considered one of the most important proxies for text quality (e.g., Crossley & McNamara, 2011; Engber, 1995; Grobe, 1981; Jarvis, 2002; Malvern, Richards, Chipere & Durán, 2004; McNamara et al., 2010; Nold & Freedman, 1977) and is perhaps the most commonly used one (Crossley, 2020). Both usage-based and psycholinguistic

approaches (see Ellis, 2002, 2012, respectively) assume that “more proficient writers produce words that are more difficult to process and recognise, either because of exposure to the words or because of properties inherent to the words” (Crossley, 2020, p. 418).

The aim of this paper is therefore to provide a longitudinal description of the lexical richness of secondary school students’ L2 history essays (see Breeze & Gerns, 2019, for the impact of academic writing instruction on this population). The description of L2 development in a school setting is still incomplete for several reasons. Firstly, language assessment is not often content-bound and therefore the evidence of competence derives from language output disconnected from both the discipline in question (e.g., history, science, mathematics, etc.) and the curriculum (the Industrial Revolution, ecosystems, integers, etc.). Secondly, corpora produced in formal learning settings are mostly cross-sectional, which means that writers’ development is difficult to track (Dóczy & Kormos, 2016; Nikula, 2017; Pellicer-Sánchez, 2018). By tracing how the lexical richness of secondary education students’ L2 historical essays developed over a period of three years, we wanted to gain further insights into the evolution of academic writing proficiency at a critical moment in students’ literacy development, namely from early to mid-adolescence.

2. The development of lexical richness

2.1. Lexical richness

Despite the popularity of linguistic lexical features as a measure of text quality and learners’ writing proficiency, there is still a certain disparity in the conceptualisation of lexical richness. For Crossley (2020), it consists of the number of unique words (lexical diversity), the proportion between content and function words (lexical density) and the proportion of advanced words (lexical sophistication) in a text. The problem lies in the operationalisation of the advanced word construct. Traditionally, research has focused on the number of low frequency words (Laufer & Nation, 1995), but, as Crossley puts it, this construct has evolved to encompass a vast number of word properties (Crossley, 2020, p. 418):

Sophisticated words have been defined as words that are more likely found in academic texts (Coxhead, 2000), words that are less concrete, imageable, and familiar (Crossley & Skalicky, in press; Saito et al., 2016; Salsbury,

Crossley, & McNamara, 2011), words that have fewer phonological and orthographical neighbors, words that have higher latencies in word naming and lexical decision tasks (Balota et al., 2007), more specific words (Fellbaum, 1998), and words that are less diverse based on context (McDonald & Shillcock, 2001).

For Jarvis (2013, 2017; in Vanhove, Bonvin, Lambelet & Berthele, 2019), the lexical richness of a text may be reflected in six dimensions: the total number of words (volume), the lexical diversity (variability), the equal or unequal repetition of words (evenness), the frequency of words in the language as a whole (rarity), the similarity of words (disparity) and the distribution of repeated words in the text (dispersion). The problem with this theoretical model is, once again, the operationalisation of the dimensions, particularly those pertaining to the textual relationship between words: evenness and dispersion. For evenness, Jarvis (2013b) uses the standard deviation of the counts of tokens per type. Regarding dispersion, he considers it to be the mean distance between different tokens of the same type, averaged over all types in the text, but admits that he presently computes it as “the number of times that types are repeated within the next n (e.g., 20) tokens” (Vanhove et al., 2019: 502).

As can be appreciated, both conceptualisations are totally compatible, only differing in the grouping of indices in dimensions and the extent to which textual relations between words are considered. Nevertheless, studies focusing on the development of lexical richness tend to use a combination of the aforementioned parameters.

2.2. Lexical richness and writing proficiency

A comprehensive review of the empirical findings regarding the relationship between text quality and lexical richness can be found in Crossley (2020). With respect to L1 writing, research has shown that more proficient writers tend to use more academic words (Douglas, 2013), more specific and less¹ polysemous words and more imageable and concrete words (Crossley, Roscoe, McNamara & Graesser, 2011; McNamara, Crossley & Roscoe, 2013), less meaningful words (McNamara et al., 2013), longer words (Crossley, Weston, McLain & McNamara, 2011; Gardner, Nesi & Biber, 2019; Haswell, 2000), less familiar words (Crossley, Weston, McLain & McNamara, 2011) and more infrequent words (Crossley, Roscoe, McNamara & Graesser, 2011; McNamara et al., 2010).

As to L2 writing, similar patterns have been reported. According to the research, more proficient L2 writers tend to use more specific and less polysemous words (Guo, Crossley & McNamara, 2013; Kyle & Crossley, 2016), less meaningful words (Crossley & McNamara, 2012), longer words (Grant & Ginther, 2000; Reppen, 1994) and less familiar and more infrequent words (Crossley & McNamara, 2012). The only difference in L1 writing has been observed in the imageability and concreteness of words, as more proficient L2 writers have been found to use less imageable words (Crossley, Kyle, Allen, Guo & McNamara, 2014).

According to Jarvis's theoretical model (2013, 2017), three of his six dimensions predict expert ratings of overall text quality (Vanhove et al., 2019): variability or lexical diversity, (Crossley & McNamara, 2011; Engber, 1995; Grobe, 1981; Jarvis, 2002; Kuiken & Vedder, 2014; Malvern et al., 2004; McNamara et al., 2010), rarity or the number of less frequent words (Crossley & McNamara, 2011; Guo et al., 2013; Malvern et al., 2004; McNamara et al., 2010) and volume or the overall number of words (Grobe, 1981; Jarvis, Grant, Bikowski & Ferris, 2003; Nold & Freedman, 1977).

Given the considerable number of lexical indices, Crossley, Salsbury, McNamara and Jarvis (2010) attempted to identify those that better predict human ratings of lexical proficiency. After analysing word length, lexical diversity, word frequency, hypernymy, polysemy, semantic co-referentiality, word meaningfulness, word concreteness, word imageability and word familiarity, they concluded that the best predictors of written lexical proficiency were lexical diversity, hypernymy and frequency, which explained 44 per cent of the variance in the human evaluations.

3. The development of history literacy

The development of lexical richness and writing proficiency needs to be framed within the development of disciplinary literacy. Biliteracy –and therefore disciplinary literacy in an L1 and L2– is a continuum (Hornberger, 2004). The transition from plain, here-and-now conversational language (BICS) to mature there-and-then academic language (CALP) is a watershed in individual language use, especially in writing. The longitudinal study of general academic language has brought out various aspects of language growth and development across life stages (see Biber, 1992, on academic genre acquisition; Christie, 2012, on language education from a functional

perspective; Grabe, 2002, on the transition from narrative to expository texts; Ortega & Byrnes, 2008, on advanced discourse). The development of academic language results from the consolidation of the language of a discipline in the form of sentential components (lexicogrammar) and discourse aspects (the discipline's functions and genres). These language features shape knowledge structures in each academic area and constitute disciplinary literacy, to wit, literacy in biology, mathematics, history, etc. (Shanahan & Shanahan, 2008).

Thus, just as descriptions and definitions are key to science (Mohan, Leung & Davison, 2001) and argumentation is important in algebra (Prediger & Hein, 2017), so too is history characterised by particular language features in which lexical richness plays a substantial role: nominalisations, implicit causal and temporal organisation and cause-effect relations within clauses (see Achugar & Schleppegrell 2005; Coffin, 2006, 2009; Lorenzo, 2017; Nokes, 2013; Schleppegrell & Colombi, 2002; see also Achugar & Carpenter, 2014, for a description of language in history, both as an L1 and L2). Furthermore, history is fundamentally a written discipline, to the point that the historical periods for which there are no written testimonies are referred to as prehistory. The linguistic turn of this discipline has even led historiographers to declare that 'history as science' is nothing more than a 'literary artefact' (White, 2010).

History literacy is an evolutionary construct. Coffin (2006) proposes a three-stage model of historical thinking: first, a purely narrative period (recording, corresponding to the 11-13 age bracket); then, an exploration of causes and consequences including multifactorial causality (explaining, corresponding to the 14-16 age bracket) and, finally, personal judgement, plus an ideological stance (arguing, corresponding to the 16-18 age bracket). When learning history, students need to leverage arguments, evaluations, generalisations, and abstractions in order to progress (Christie & Maton, 2011). Mature history narratives employ a higher concentration of nominals, more morphological narrative complexities and more present and past participles over time (Asención-Delaney & Collentine, 2011). These features modify the writing styles of students, who can consequently meet more complex academic and discursive requirements.

4. Methodology

4.1. Research background

This research builds on a series of previous studies in the European field of CLIL (Granados, Lorenzo-Espejo & Lorenzo, 2021; Lorenzo & Granados, 2020; Lorenzo & Moore, 2009; Lorenzo & Rodríguez, 2014; Lorenzo, Granados & Rico, 2021), inquiring into issues such as the advantages of CLIL versus monolingual education and the description of incidental learning and positive transfer between an L1 and an L2. Particularly, Lorenzo, Granados and Ávila (2019) explored the development of fluency, syntactic complexity, and text easability of the learner corpus analysed in this paper, and Granados and Lorenzo (2021) described the development in the use of connectives.

In the research context described above, the aim of this study is to analyse the development of L2 written lexical richness in the discipline of history and how it fluctuated in a formal bilingual setting within an established time frame: three academic years.

4.2. Research questions

In order to fulfil this objective, several research questions were formulated:

4.2.1. RQ 1. Did the lexical richness of the students' L2 historical writing develop over time?

Employing Crossley's (2020) conceptualisation, written lexical richness was analysed as the number of unique words (lexical diversity), the proportion between content and function words (lexical density) and the proportion of advanced words (lexical sophistication) in the essays.

4.2.2. RQ 2: If so, which lexical dimensions evolved in the students' L2 historical writing?

Each dimension was studied separately to determine whether or not they developed differently as students matured. If a dimension evolved, it meant that it developed during the critical period of adolescence and that it was sensitive enough to maturation to vary over a three-year period.

4.2.3. RQ 3: If so, was there any lexical dimension that did not evolve in the students' L2 historical writing?

This would imply that there are dimensions that did not develop during the

critical period of adolescence or which were not sensitive enough to maturation to vary over a three-year period.

4.3. Sampling and participants

This study was performed on a sample of students from a state secondary school in Andalusia (southern Spain), which has been running an optional English-Spanish bilingual programme for the past 15 years, in keeping with the Spanish and European trend towards CLIL-type multilingual education. The study sample was made up of 20 students enrolled in the bilingual programme, all of whom were L1 Spanish speakers and belonged to the same grade and class. This made it possible to neutralise many of the variables present in learning environments (the teaching methodology, the quality and quantity of language exposure, the number and nature of courses taught in English, etc.), thus providing an adequate setting for a longitudinal study.

Since students were in the Andalusian bilingual programme, they received 4–5 weekly contact hours of explicit L2 instructions (depending on the school year they were in) and at least two content subjects were taught in the L2 (one of them always being Social Sciences). A minimum of 30% and a maximum of 50% of the curriculum had to be taught in English (see Andalusian Department of Education, 2017, for more information).

These students were tested five times over a three-year period. When the first test (Test 1) was administered, they were all ninth-graders (aged between 14 and 15) who had already received two years of education in the bilingual programme. They had an English level of A2 according to the Common European Framework of Reference for Languages (CEFR). By the time the fifth test (Test 5) was administered, they had become eleventh-graders (aged between 16 and 17) and were expected to oscillate between a B1 and B2 level of English within the following two academic years. Nevertheless, the sample suffered attrition. Three of the students abandoned the bilingual programme in the tenth grade (because of the extra cognitive demands involved, the amount of work required or other academic reasons) and two had to retake a year (ninth grade). As a result, a total of 15 students sat the five tests.

4.4. Instrumentation and data collection

Five tests in the form of history essays were administered to the students without prior notice. The test topics were in keeping with the official history

curriculum being taught in class, specifically, 9/11 and the Clash of Civilisations, the Avant-Gardes, the Industrial Revolution, the American Revolution, and the Spanish Civil War. During the three-year study time frame, the history teacher informed the research team when each topic was being studied. On the basis of this information, Tests 1 and 2 were administered in the ninth grade (14-15 years old), Tests 3 and 4 in the tenth grade (15-16 years old) and Test 5 in the eleventh grade (16-17 years old).

The essays composed by the students were based on what they had learnt in the ordinary history course. They were not allowed to consult any additional materials when sitting the tests, which were supervised. Besides the minimum length for the essay, they were only given the following guidelines: (a) define the given concept or historical period; (b) explain its causes and consequences; and (c) give your opinion on its historical implications.

The data collection process resulted in a learner corpus made up of 75 essays, totalling 12,000 words, which were organised in three periods of composition (Year 1, Year 2, and Year 3), corresponding to the three years of the study. Year 1 encompassed the combined results of Tests 1 and 2, and Year 2, those of Tests 3 and 4. In the third and final year (Year 3), only one test (Test 5) was administered in order to avoid further dropouts which would have seriously compromised the study.

4.5. Computational analysis

The students' essays were coded and processed with the Coh-Metrix computational tool, which produces indices of the linguistic and discursive representations of a text in five major dimensions: narrativity, syntactic simplicity, word concreteness, referential cohesion and deep (causal) cohesion (McNamara, Louwse, Cai & Graesser, 2014). Coh-Metrix has been validated by numerous researchers, including Polio and Yoon (2018). These authors compared Coh-Metrix results with hand-coding, confirming that it is a non-redundant and reasonably transparent tool for measuring cohesion, complexity, and coherence metrics, as well as being capable of reflecting differences in genres among English-as-an-L2 (ESL) writers with reliability. Similarities in results and metrics have also been found with other software tools used to analyse the lexical complexity of history essays (Lorenzo & Rodríguez, 2014).

In this study, Crossley's (2020) conceptualisation was used. Written lexical richness was therefore analysed as the number of unique words (lexical

diversity), the proportion between content and function words (lexical density) and the proportion of advanced words (lexical sophistication) in the essays. The following Coh-Metrix indices were employed:

4.5.1. Written lexical diversity

A. The type-token ratio for content words and for all words

Lexical diversity assesses the range of vocabulary employed in a text (McNamara et al., 2014). The most reputed measure of lexical diversity is the type-token ratio (hereinafter TTR), a coefficient resulting from dividing the number of unique words in a text (i.e., types) by the overall number of words (i.e., tokens). The type token ratio can be measured by means of two Coh-Metrix indices: (a) Coh-Metrix index 48, which only processes content words (i.e., nouns, verbs, adjectives, and adverbs) sharing a common lemma (e.g., tree/treed; mouse/mousey; price/priced, etc.); and (b) Coh-Metrix index 49, which measures the type-token ratio for all words.

B. The Measure of Textual Lexical Diversity (MTLD) and vocd lexical diversity measures for all words

TTR has proved to be extremely sensitive to text length and, therefore, a poor predictor of lexical proficiency when text length is not constant. In fact, “as the number of word tokens increases, there is a lower likelihood of those words being unique” (McNamara et al., 2014, p. 67) and TTR tends to be lower. In order to overcome these metric limitations, Coh-Metrix includes indices that use estimation algorithms such as the Measure of Textual Lexical Diversity (hereinafter MTLD) and vocd, indices 50 and 51, respectively.

The MTLD is calculated as the mean length of sequential word strings in a text which maintain a given TTR value (McNamara et al., 2014, p. 67). Similarly, vocd is calculated through a computational procedure that matches TTR random samples with ideal TTR curves (McNamara et al., 2014, p. 67). Both indices allow researchers to compare the lexical diversity of texts differing in length, although validation studies tend to favour MTLD over vocd (McCarthy & Jarvis, 2010).

4.5.2. Written lexical density

Lexical density is the proportion between content and function words. Coh-Metrix indicates the incidence of nouns, verbs, adjectives and adverbs

(indices 84-87) per 1000 words. By combining these indices, the proportion between content and function words can be calculated.

4.5.3. Written lexical sophistication

This dimension was operationalised by means of the following indices:

A. Familiarity, concreteness, imageability and meaningfulness for content words

Familiarity (Coh-Metrix index 98) refers to how familiar a word seems to an adult on a 700-point scale (100 for unheard words and 700 for those heard almost every day), according to the MRC Psycholinguistic Database (McNamara et al., 2014).

Concreteness (Coh-Metrix index 99) indicates how concrete or non-abstract a word is on the same scale –100 for words that score low in concreteness, like ‘protocol’ (264), and 700 for words referring to things that can be touched, heard, or tasted, like ‘box’ (597)– according to the MRC Psycholinguistic Database (McNamara et al., 2014).

Meaningfulness (Coh-Metrix index 100) refers to the extent to which one word can be associated with others, on the same scale –100 for words with a weak association, like ‘abbess’ (218), and 700 for those with a strong association, like ‘people’ (612)– according to the MRC Psycholinguistic Database (McNamara et al., 2014).

Imageability (Coh-Metrix index 101) indicates how easy it is to construct a mental image of a word, on a similar scale –100 for low-imagery words, like ‘reason’ (267), and 700 for high-imagery words, like ‘hammer’ (618)– according to the MRC Psycholinguistic Database (McNamara et al., 2014).

B. Word length

Word length is calculated in relation to the mean number of syllables per word (Coh-Metrix index 8) and the mean number of letters per word (Coh-Metrix index 10).

C. Word frequency

Word frequencies for content words (Coh-Metrix index 94) are given in accordance with the CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1995).

D. Polysemy and hypernymy

Polysemy (Coh-Metrix index 102) is computed as the mean number of senses (core meanings) of the content words used in a text, according to the WordNet lexicon.

Hypernymy indicates the rank of a word on the hierarchical scale of the WordNet lexicon. For instance, ‘entity’ is considered as the hypernym of all nouns and, therefore, has a hypernymy value of 1. For its part, ‘chair’ has many higher hypernymy categories (only as regards the object, e.g., ‘furniture’, ‘furnishing’, ‘instrumentality’, ‘artefact’, ‘whole’, ‘object’, ‘physical entity’ and ‘entity’) and, therefore, has a hypernymy value of 8.5. Coh-Metrix provides the hypernymy values for nouns (index 103) and for verbs (index 104).

The results and the mean values of each Coh-Metrix index were studied in order to perform a descriptive analysis on lexical development, with the aim of revealing quantitative and qualitative tendencies in multiple case studies.

5. Results

5.1. Written lexical diversity

5.1.1. The type-token ratio for content words and for all words

As can be seen in Figure 1, the mean type-token ratio for content words and the mean type-token ratio for all words followed the same decreasing pattern, although the type-token ratio for all words registered lower results. This gap in Figure 1 is perfectly logical: non-content words such as conjunctions (‘and’, ‘but’, etc.), prepositions (‘in’, ‘out’, etc.), or pronouns (‘he’, ‘who’, etc.) are much more frequent and therefore repeated in a prototypical text. When considered, lexical diversity is necessarily lower. At first sight, therefore, the students’ lexical diversity decreased as the study advanced.

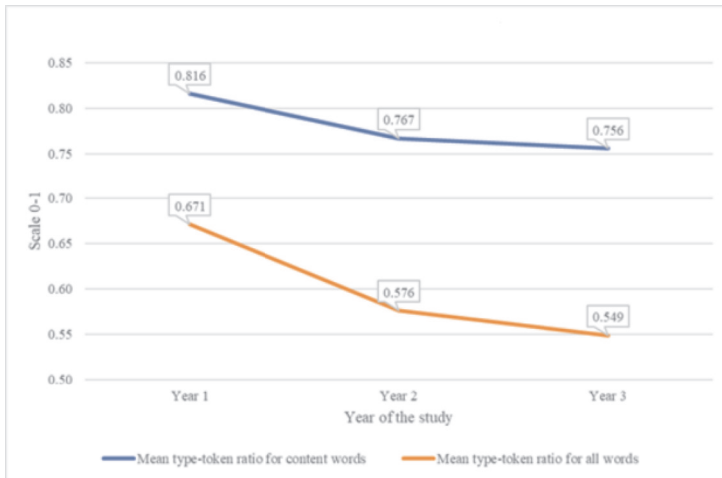


Figure 1. Mean type-token ratio of the students' essays.

However, as already explained in section 3.5, TTR has proved to be extremely sensitive to text length and, therefore, a poor predictor of lexical proficiency when textual length is not constant. That is precisely the case here, as the students displayed higher levels of conceptual fluency over time, which led to much longer texts. In order to remedy the metric limitations, Coh-Metrix includes indices that use estimation algorithms such as the MTLT and vocd.

5.1.2. MTLT and vocd lexical diversity measures for all words

As shown in Figure 2, even though the degree to which lexical diversity increased differed noticeably from MTLT to vocd (the former being regarded as more reliable), both measures indicated a clear growth (by more than 20%, according to MTLT; and by almost 50%, according to vocd). These gains were constant over time and developed gradually from Year 1 to Year 2, and from this mid-point to Year 3. One full-text example of this development is shown in Table 2, at the end of the results section.

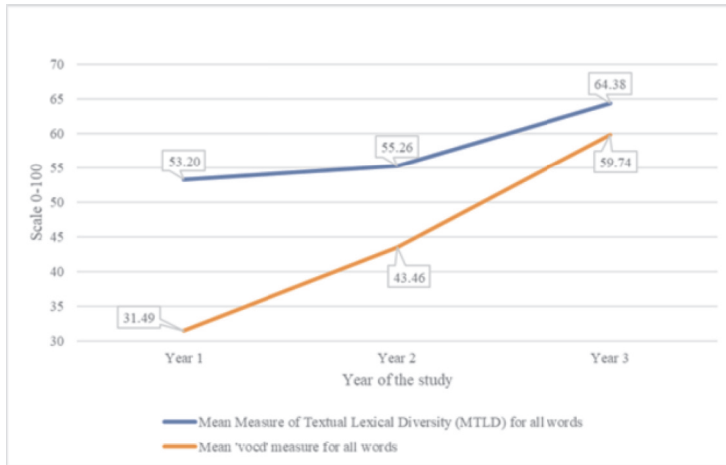


Figure 2. Mean MTLD and vocd of the students' essays.

From this steady increase in lexical diversity, it can be inferred that the students became more proficient as they grew older and progressed in the bilingual education system (Jarvis, 2002; McNamara et al., 2010; Crossley, Weston, McLain & McNamara, 2011; Crossley & McNamara, 2012).

5.2. Written lexical density

The proportion between content and function words is shown in Table 1. The overall proportion of content words increased slightly from Year 1 to Year 3. This was due to the considerable growth in the proportion of adjectives, which compensated for the slight decreases in the proportion of nouns and verbs. This growth in the proportion of adjectives could be related to the expansion of noun phrases, a feature of academic writing. Despite these minor variations, however, no remarkable development was observed in this dimension.

| | Mean number of words/text | Proportion of content words | Proportion of nouns | Proportion of verbs | Proportion of adjectives | Proportion of adverbs |
|--------|---------------------------|-----------------------------|---------------------|---------------------|--------------------------|-----------------------|
| Year 1 | 101.472 | 52.95% | 30.65% | 13.14% | 5.64% | 3.51% |
| Year 2 | 177.470 | 53.45% | 30.41% | 11.49% | 8.32% | 3.24% |
| Year 3 | 242.400 | 53.56% | 28.83% | 12.55% | 8.36% | 3.81% |

Table 1. Mean proportion of content words in the students' essays.

5.3. Written lexical sophistication

5.3.1. Familiarity, concreteness, imageability and meaningfulness of content words

The study results displayed in Figure 3 show that familiarity levels remained very similar, with a variation of less than one point up or down the scale over time (573, 572 and 574, respectively), meaning that word difficulty remained constant. The concreteness and imageability indices fell considerably during the first two years, while remaining constant during the final year, thus implying a higher degree of abstraction. The essays became less picturesque and anecdotal, hinting at a transition from more narrative to more expository texts. Finally, lexical meaningfulness increased moderately but steadily, pointing to the construction of more cohesive texts with words better knitted in clusters and lexical bundles, thus proving that lexical development is not random but develops in semantic networks. A full-text example can be consulted in Table 2, at the end of the results section.

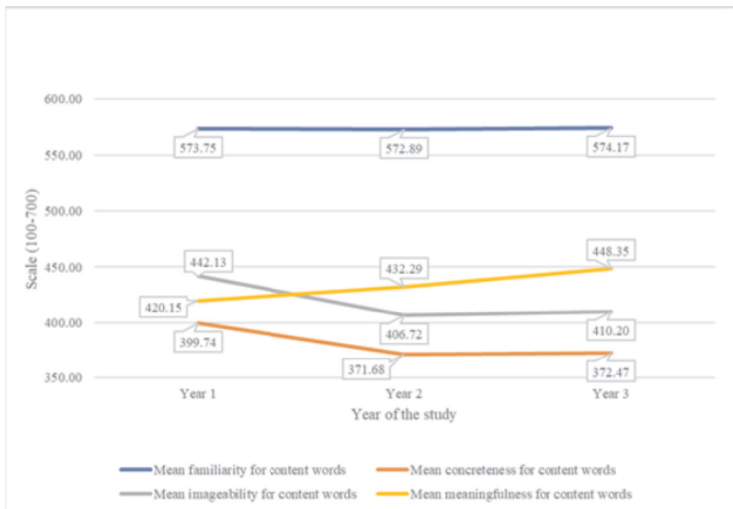


Figure 3. Mean familiarity, concreteness, imageability and meaningfulness of the students' essays.

5.3.2. Word length, word frequency, polysemy, and hypernymy

Finally, no remarkable development was observed for word length, word frequency, polysemy, and hypernymy. Their evolution was subtle and irregular. Their values can be consulted in the Appendix.

| Student 10's first and last essays | |
|---|---|
| Year 1 | <p>On September 11th 2001, some terrorists hijacked two planes and crashed them into the Twin Towers in New York. The planes crashed on the 90th floor, so many people¹ were trapped and they can't went out. Some people jump to street and died. Some minutes after the crashed, the towers fell down.</p> <p>This happened because some years before, USA had a war with Iraq because Americans think Iraq had nuclear weaponry. As a revenge Iraq hijacked some planes and wanted to crashed them into some important buildings of USA, two of them were The Twin Towers.</p> <p>Because many innocent people died, because some unforgiving terrorists want to harm Americans.</p> |
| Year 3 | <p>The Spanish Civil War began on July 1936. The rising took place in Melilla, where the general Francisco Franco rebelled against the current Republic. This rebellion had been planned by different generals from different parts of Spain. Therefore, it quickly spreaded across the country, having under control a third of the country after a few days. On the one hand, the national group, Franco's one, received the support of fascist countries such as Germany or Italy, as well as Portugal. On the other hand, the republican army did not get the support of other democratic countries but did get it from the Soviet Union, due to the numerous communists politicians that conformed the government.</p> <p>Franco won almost every battle they had had so, by the end of 1939, all Spain was under his control. As a consequence of the battles many people died, but many others were killed by their neighbours, who acused each other as a result of past disagreements. In addition, republicans were shoted when a new town was occupied.</p> <p>Some important consequences of this war was the high number of deaths and the separation that suffered the society. This lead to difficulties such as to get a job to those people who were in the republican part at the beginning of the war. Moreover, there was an extremely high number of disappeared people.</p> |

Table 2. Student 10's first and last essays. Original grammar and spelling.

6. Discussion

This study has analysed the development of lexical richness in L2 historical writing according to three dimensions: the number of unique words (lexical diversity), the proportion between content and function words (lexical density) and the proportion of advanced words (lexical sophistication). In these dimensions, a clear evolution was observed in only two: lexical diversity and lexical sophistication. The second dimension, lexical density, remained constant during the three-year study period. Within lexical sophistication, development was detected in the familiarity, concreteness, imageability and meaningfulness of words. Word length, word frequency, polysemy and hypernymy also remained unaffected. The development of each dimension will now be addressed in turn. Furthermore, in order to flesh out the raw data, the results will be discussed on the basis of a corpus sample of one of the student's essays from Year 1 and Year 3 (i.e., from the first and last tests administered), which are included in Table 2 (results section).

High levels of lexical diversity entail lower cohesion and higher difficulty: there are more unique words introducing new information that needs to be processed and integrated into the discourse by the reader (McNamara et al., 2014). In contrast, the greater the frequency with which the same words are used multiple times across the text, the lower the lexical diversity and the higher the text cohesion will be. In this study, the increase in the number of lexical items employed by the students has been confirmed, a trait that indicates that they were becoming more proficient L2 writers (Crossley & McNamara, 2011; Crossley & McNamara, 2012; Engber, 1995; Grobe, 1981; Jarvis, 2002; Kuiken & Vedder, 2014; Malvern et al., 2004; McNamara et al., 2010).

Moreover, if the nature of this increase in the breadth of vocabulary is examined in the students' essays, the first feature to emerge is the persistence of semantic extension over time (Harmon & Kapatsinski, 2017). Learners initially extend the L1 semantic load of lexical items to L2 equivalents. This was represented in the students' essays by the presence of calques, like 'conform', which is employed with the meaning of the Spanish verb *conformar* ('form', 'make up', 'constitute'). Semantic extensions decline over time, however, when L1 intake is blocked out and L2 intake is connected only to L2 relations. Research has called this process a move from 'word association representation' to 'conceptual mediation representation' (Spöttl & McCarthy, 2004). These results show that in this bilingual model, once compulsory education has been completed, there are still indications of overreliance on L1 for word generation in the form of transliteration, calques or extreme translanguageing. These are different forms of the 'one-to-one principle', namely, the naïve belief that lexical units in the two languages match perfectly. High idiomaticity levels are an indication of L2 proficiency, but here the first language still contaminates L2 production, especially as regards academic vocabulary (see, for example, the misspelling of cognates like *'comunists' and *'acused', and the structural calques *'the separation that suffered the society' and *'get a job to those people').

The second dimension in which evolution was observed is lexical sophistication, particularly as regards the familiarity, concreteness, imageability and meaningfulness of words. These are key indices for writing proficiency: less familiar words are more difficult to learn and take longer to process (McNamara et al., 2014), word concreteness and word imageability are indirectly proportional to abstraction (Barber, Otten, Kousta & Vigliocco, 2013) and the average meaningfulness of a text is indirectly

proportional to text difficulty, since words with a stronger association imply that readers need to process and integrate less new information into the discourse (Crossley & McNamara, 2012). In this regard, research has found that more proficient L2 writers tend to use less concrete and imageable words (Crossley et al., 2014) and less familiar and meaningful words (Crossley & McNamara, 2012). In this study, the students did indeed use less concrete and imageable words; that is, there was a greater degree of abstractness. Contrary to Crossley and McNamara's (2012) findings, however, familiarity remained constant and there was an increase in the meaningfulness of lexical items; namely, terms had a greater degree of associativity. This divergence might have been due to the age and the developmental stage of the students making up the sample, as they are far from reaching the top proficiency levels.

In terms of associativity, one implication of the net gains reported is that lexical growth is not random but develops in semantic networks. Example 2, from Year 3 (Table 2, results section), includes a wide variety of words related to conflict: 'war', 'battle' and 'rebellion'. Indeed, lexical development goes hand in glove with a better control of derivational mechanisms which improve the quality of academic writing. In Example 2, three different word forms for the same word family concur: noun ('rebellion'), adjective ('rebel') and verb ('rebelled'). Derivational expertise goes a long way to helping text cohesion and textual cross-references. The new constellation of semantic fields not only includes nominal groups, because grammar words for expressing functional categories also increase over time, as will be seen below regarding the expression of causality.

As to abstractness, research has observed that abstraction in academic writing is achieved by means of signalling nouns, namely, abstract nouns which refer to a general area of meaning whose specific meaning is found elsewhere in the clause or text (Flowerdew, 2014: 96). One such example can be detected in Example 2 (Year 3), in the account of a historical episode in which 'difficulties' are mentioned 'for the armies involved in warfare'. The actual embodiment of such difficulties is only found further on in the sentence. This dummy word exists mostly for the sake of anticipating semantic processes, here of a historical nature. Lexical gains, therefore, follow a tendency towards more abstract language.

The development of abstraction in written language relates in part to that of nominalisation. Nominalisation characterises mature academic language like

no other construction (Lorenzo et al., 2019, Granados et al., 2021). At earlier ages, as in Test 1, language includes more verbs and more prototypical theme/rheme sentences. Over time, language evolves and becomes more nominal. Terms like ‘separation’ and ‘support’ represent the typical grammatical metaphor, whereby noun phrases are used instead of verb-like sentences. As is well-known in functional systemics, nominalisations freeze actions and transform eventful episodes into non-temporal abstract processes: as in the use of ‘rising’ (as in a coup d’état) in Example 2, as opposed to a non-nominalised ‘X rose against Y’ pattern, which would have been more typical in the case of a younger student. When describing this composition device, Halliday (2004) posited that when writers express a process by means of nominalisation, a rhetorical tension is created between the semantic level (which describes a process as if it were an agent undertaking an action) and the lexicogrammatical level (the actual nominal word forms which embody the action). He went on to say that this is regarded as a metaphor because the end result is a virtual entity which only exists as semiosis. The use of ‘rising’, instead of military insurrection, in Example 2 further elaborates on the metaphor within. The fact that the action described (‘the military rose in arms’) is represented by a neutral or even positive action (‘rising’) ties in with the fascist propaganda following the military coup. This bilingual student’s command of history vocabulary demonstrates not only advanced lexical knowledge, but also the consolidation of abstract thought in ideological writing (e.g., ‘the support of fascist countries such as Germany or Italy, the numerous *comunist politicians that *conformed the government, and as a result of past disagreements’).

To conclude the discussion on lexical richness, it should also be noted that it was possible to glimpse the bigger picture by just glancing at the general discourse structure of the essays. In addition to the differences in lexical constituents, the essays composed in Test 5 show variations in discourse texture. In later stages, they are more densely packed with lexical collocations (e.g., adverb + adjective, like the phrase ‘extremely high number of *disappeared people’, in Example 2 from Test 5). This implies a new approach to text construction involving longer units with more pre- and post-modifications.

Having said that, the results should be interpreted with caution due to two research limitations. Firstly, in longitudinal multiple-case studies like this one, which aimed at neutralising contextual variables (by testing students taught

under the same conditions), the results are merely descriptive. The findings discussed here should pave the way for future large-scale, cross-sectional analyses aimed at testing their generalisability. Secondly, even though the study included five tests per student over a three-year period, individual text topics might have affected the range of vocabulary used by the students (Tracy-Ventura, Mitchell & McManus, 2016). Despite these limitations, this paper describes a pioneering longitudinal study of lexical development in L2 historical writing in a formal bilingual setting, at a crucial moment for the academic language development of students.

7. Conclusion

Writing proficiency is usually analysed by means of three constructs: lexical, syntactic and cohesion (McNamara et al., 2010). Our study has focused on the lexical construct and has examined the lexical richness of secondary school students' L2 historical writing in relation to their lexical diversity, density, and sophistication. Our results show that, after three years of formal bilingual education, the students in the sample used more lexical items and employed terms with a larger degree of abstractness and associability. This points to their greater lexical richness, their higher writing proficiency, and their more profound grasp of history literacy.

History is a fundamentally written discipline, a literary artefact to postmodern historiography (White, 2010). Following this conception, international institutions such as the Council of Europe have even developed language descriptors which evaluate writing in relation to the discipline of history. In this context, the fact that the students' history literacy and writing proficiency matured during the three-year study period may help us to understand content learning and to support further academic success.

Article history:
Received 03 March 2022
Received in revised form 02 May 2022
Accepted 03 May 2022

References

- Achugar, M., & Carpenter, B. (2014). Tracking movement toward academic language in multilingual classrooms. *Journal of English for Academic Purposes*, 14, 60-71. <<https://doi.org/10.1016/j.jeap.2013.12.002>>

- Achugar, M. & Schleppegrell, M.J. (2005). Beyond connectors: The construction of cause in history textbooks. *Linguistics and Education*, 16(3), 298-318. <<https://doi.org/10.1016/j.linged.2006.02.003>>
- Andalusian Department of Education (2017). Acuerdo de 24 de enero de 2017. *BOJA*, 24, 10-57.
- Asención-Delaney, Y., & Collentine, Y. (2011). A multidimensional analysis of a written L2 Spanish corpus. *Applied Linguistics*, 32, 299-322. <<https://doi.org/10.1093/applin/amq053>>
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. University of Pennsylvania, Linguistic Data Consortium.
- Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain & Language*, 125, 47-53. <<https://doi.org/10.1016/j.bandl.2013.01.005>>
- Biber, D. (1992). The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities*, 26(5/6), 331-345.
- Breeze, R., & Gerns, P. (2019). Building literacies in secondary school history: The specific contribution of academic writing support. *E-JournALL, EuroAmerican Journal of Applied Linguistics and Languages*, 6(1), 21-36. <<https://doi.org/10.21283/2376905X.10.149>>
- Britton, J., Martin, N., & Rosen, H. (1966). *Multiple Marking of Compositions*. Her Majesty's Stationary Office.
- Christie, F., & Maton, K. (2011). *Disciplinary: Functional linguistic and sociological perspectives*. Continuum.
- Christie, F. (2012). *Language education: A functional perspective*. Wiley-Blackwell.
- Coffin, C. (2006). Reconstructing personal time as collective time: Learning the dis-course of history. In R. Whittaker, M. O'Donnell, M. & A. McCabe (Eds.), *Language and literacy: Functional approaches* (pp. 15-45). Continuum.
- Coffin, C. (2009). *Historical discourse*. Continuum.
- Crossley, S. A., & McNamara, D.S. (2011). Understanding expert ratings of essay quality: Coh-Matrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning*, 21(2-3), 170-191. <<https://doi.org/10.1504/IJCELL.2011.040197>>
- Crossley, S. A., & McNamara, D.S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115-135. <<https://doi.org/10.1111/j.1467-9817.2010.01449.x>>
- Crossley, S. A., Kyle, K., Allen, L. K., Gou, L., & McNamara, D.S. (2014). Linguistic microfeatures to predict L2 writing proficiency: A case study in automated writing evaluation. *Journal of Writing Assessment*, 7(1).
- Crossley, S. A., Roscoe, R. D., & McNamara, D. S. (2011). Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay & A. Mitrovic (Eds.), *Artificial Intelligence in education (AIED 2011)* (pp. 438-440). Springer. <https://doi.org/10.1007/978-3-642-21869-9_62>
- Crossley, S. A., Salsbury, T., McNamara, D. S. & Jarvis, S. (2010). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561-580. <<https://doi.org/10.1177/0265532210378031>>
- Crossley, S. A., Weston, J., McLain Sullivan, S. T., & McNamara, D.S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28(3), 282-311. <<https://doi.org/10.1177/0741088311410188>>
- Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415-443. <<https://doi.org/10.17239/jowr-2020.11.03.01>>
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 121-129.
- Deane, M., & O'Neill, P. (2011). *Writing in the disciplines*. Palgrave Macmillan.
- Dóczy, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. Oxford University Press.
- Douglas, R. D. (2013). The lexical breadth of undergraduate novice level writing competency. *The Canadian Journal of Applied Linguistics*, 16(1), 152-170.
- Dressen-Hammouda, D. (2008). From novice to disciplinary expert: disciplinary identity and genre mastery. *English for Specific Purposes*, 27(2), 233-252. <<https://doi.org/10.1016/j.esp.2007.07.006>>
- Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition.

- Studies in Second Language Acquisition*, 24(2), 143-188. <<https://doi.org/10.1017/S0272263102002024>>
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17-44. <<https://doi.org/10.1017/s0267190512000025>>
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139-155. <[https://doi.org/10.1016/1060-3743\(95\)90004-7](https://doi.org/10.1016/1060-3743(95)90004-7)>
- Flowerdew, J. (2014). Corpus-based approach to language description for specialized academic writing. *Language Teaching*, 50, 90-106. <<https://doi.org/10.1017/S0261444814000378>>
- Gardner, S., Nesi, H., & Biber, D. (2019). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*, 40(4), 646-674. <<https://doi.org/10.1093/applin/amy005>>
- Grabe, W. (2002). Narrative and expository macrogenres. In A. N. Johns (Ed.), *Genre in the classroom: Multiple perspectives*. Lawrence Erlbaum Associates Publishers.
- Granados, A., & Lorenzo, F. (2021). English L2 connectives in academic bilingual discourse: a longitudinal computerised analysis of a learner corpus. *Revista Signos*, 54(106), 626-644. <<https://doi.org/10.4067/S0718-09342021000200626>>
- Granados, A., Lorenzo-Espejo, A., & Lorenzo, F. (2021). Evidence for the interdependence hypothesis: A longitudinal study of biliteracy development in a CLIL/bilingual setting. *International Journal of Bilingual Education and Bilingualism*. <<https://doi.org/10.1080/13670050.2021.200142>>
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9, 123-145. <[https://doi.org/10.1016/s1060-3743\(00\)00019-9](https://doi.org/10.1016/s1060-3743(00)00019-9)>
- Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English*, 15(1), 75-85.
- Guo, L., Crossley, S. A., & McNamara, D.S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18(3), 218-238. <<https://doi.org/10.1016/j.asw.2013.05.002>>
- Halliday, M. A. K. (2004). *An introduction to functional grammar*. Oxford University Press.
- Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, 98, 22-44. <<https://doi.org/10.1016/j.cogpsych.2017.08.002>>
- Harwood, N., & Hadley, G. (2004). Demystifying institutional practices: critical pragmatism and the teaching of academic writing. *English for Specific Purposes*, 23(4), 355-377. <<https://doi.org/10.1016/j.esp.2003.08.001>>
- Haswell, R. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication*, 17(3), 307-352. <<https://doi.org/10.1177/0741088300017003001>>
- Hornberger, N. (2004). The continua of biliteracy and the bilingual educator: Educational linguistics in practice. *International Journal of Bilingual Education and Bilingualism*, 7(2), 155-171. <<https://doi.org/10.21832/9781853597565-004>>
- Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57-84. <<https://doi.org/10.1191/0265532202lt220oa>>
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(1), 87-106. <<https://doi.org/10.1111/j.1467-9922.2012.00739.x>>
- Jarvis, S. (2017). Grounding lexical diversity in human judgments. *Language Testing*, 34(4), 537-553. <<https://doi.org/10.1177/0265532217710632>>
- Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, 12(4), 377-403. <<https://doi.org/10.1016/j.jslw.2003.09.001>>
- Kuiken, F.- & Vedder, I. (2014). Rating written performance: What do raters do and why? *Language Testing*, 31(3), 329-348. <<https://doi.org/10.1177/0265532214526174>>
- Kyle, K.- & Crossley, S.A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34(4), 12-24. <<https://doi.org/10.1016/j.jslw.2016.10.003>>
- Lahuerta Martínez, A. C. (2015). The written competence of Spanish secondary education students in bilingual and non-bilingual programs. *Porta Linguarum*, 24, 74-61. <<https://doi.org/10.30827/Digibug.53796>>
- Laufer, B., & Nation, P. (1995). Vocabulary size

- and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322. <<https://doi.org/10.1093/applin/16.3.307>>
- Lorenzo, F., & Granados, A. (2020). One generation after the bilingual turn: Results from a large-scale CLIL teachers' survey. *Estudios de Lingüística Inglesa Aplicada*, 20, 77-101. <<https://doi.org/10.12795/elia.2020.i20.04>>
- Lorenzo, F., & Rodríguez, L. (2014). Onset and expansion of L2 cognitive academic language proficiency in bilingual settings: CALP in CLIL. *System*, 47, 64-72. <<https://doi.org/10.1016/j.system.2014.09.016>>
- Lorenzo, F., & Moore, P. (2009). European language policies in monolingual southern Europe: Implementation and outcomes. *European Journal of Language Policy*, 1(2), 121-135.
- Lorenzo, F. (2017). Historical Literacy in bilingual settings: Cognitive academic language in L2 History narratives. *Linguistics and Education*, 37(1), 32-41. <<https://doi.org/10.1016/j.linged.2016.11.002>>
- Lorenzo, F., Granados, A., & Ávila, I. (2019). The development of cognitive academic language proficiency in multilingual education: Evidence of a longitudinal study on the language of history. *Journal of English for Academic Purposes*, 41, 100767. <<https://doi.org/10.1016/j.jeap.2019.06.010>>
- Lorenzo, F., Granados, A., & Rico, N. (2021). Equity in bilingual education: socioeconomic status and content and language integrated learning in monolingual Southern Europe. *Applied Linguistics*, 42(3), 393-413. <<http://dx.doi.org/10.1093/applin/amaa037>>
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan. <<https://doi.org/10.1007/978-0-230-51180-4>>
- McCarthy, P. M., & Jarvis, S. (2010). MTL, vocD, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381-392. <<https://doi.org/10.3758/BRM.42.2.381>>
- McNamara, D. S., Crossley, S. A., & McCarthy, P.M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86. <<https://doi.org/10.1177/0741088309351547>>
- McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural Language Processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45(2), 499-515. <<https://doi.org/10.3758/s13428-012-0258-1>>
- McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- Mohan, B., Leung, C., & Davison, C. (2001). *English as a Second Language in the mainstream*. Pearson Education.
- Nikula, T. (2017). Emerging themes, future research directions. In A. Llinares & T. Morton (Eds.), *Applied Linguistics Perspectives on CLIL* (pp. 307-313). John Benjamins.
- Nokes, J. D. (2013). *Building students' historical literacies: Learning to read and reason with historical texts and evidence*. Routledge.
- Nold, E. W., & Freedman, S.W. (1977). An analysis of readers' responses to essays. *Research in the Teaching of English*, 11(2), 164-174.
- Ortega, L., & Byrnes, H. (Eds.). (2008). *The Longitudinal Study of Advanced L2 Capacities*. Routledge.
- Pellicer-Sánchez, A. (2018). Examining second language vocabulary growth. Replications of Schmitt (1998) and Webb & Chang (2012). *Language Teaching*, 52(4), 512-523. <<https://doi.org/10.1017/S026144481800037X>>
- Prediger, S., & Hein, K. (2017). Learning to meet language demands in multi-step mathematical argumentations: Design research on a subject-specific genre. *European Journal of Applied Linguistics*, 5(2), 309-337. <<https://doi.org/10.1515/eujal-2017-0010>>
- Reppen, R. (1994). *Variation in elementary student language: A multi-dimensional perspective*. [Unpublished doctoral dissertation]. Northern Arizona University.
- Rose, D. (2008). Writing as linguistic mastery: The development of genre-based literacy pedagogy. In R. Beard, D. Myhill, J. Riley & M. Nystrand (Eds.), *Handbook of Writing Development* (pp. 151-166). Sage.
- Schleppegrell, M.J., & Colombi, M.C. (2002). *Developing advanced literacy in first and second languages*. Lawrence Erlbaum.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents. Rethinking content-area literacy. *Harvard Educational Review*, 78, 40-59. <<https://doi.org/10.17763/haer.78.1.v62444321p602101>>
- Spöttl, C., & McCarthy, M. (2004). Comparing the knowledge of formulaic sequences across L1, L2, L3 and L4. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 191-225). John Benjamins.

- Tracy-Ventura, N., Mitchell, R., & McManus, K. (2016). The LANGSNAP longitudinal learner corpus. Design and use. In M. Alonso-Ramos (Eds.), *Spanish Learner Corpus Research. Current trends and future perspectives* (pp. 117-142). John Benjamins. 10.17239/jowr-2019.10.03.04>
- Vanhove, J., Bonvin, A., Lambelet, A., & Berthele, R. (2019). Predicting perceptions of the lexical richness of short French, German, and Portuguese texts using text-based indices. *Journal of Writing Research*, 10(3), 499-525. <<https://doi.org/10.17239/jowr-2019.10.03.04>>
- White, H. (2010). *The fiction of narrative: Essays on history, literature, and theory, 1957-2007*. The Johns Hopkins University Press.
- Wolfe-Quintero, K, Inagaki, S., & Hae-Youn, K. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. University of Hawaii Press.
- Young, A., & Fulwiler, T. (1986). *Writing across the disciplines: Research into practice*. Boynton/Cook.

Adrián Granados is a postdoctoral researcher at Universidad Pablo de Olavide (Seville, Spain). His research focuses on the study of second language acquisition and bilingualism, and he is specialised in processing academic corpora with linguistic analysis software. He has published in journals such as *Applied Linguistics*, *Journal of English for Academic Purposes*, and *International Journal of Bilingual Education and Bilingualism*.

María Dolores López-Jiménez is a senior lecturer at Universidad Pablo de Olavide (Seville, Spain). She has also held visiting scholar positions at Indiana University. Her research focuses on the teaching and learning process of vocabulary and (inter)cultural aspects in an L2, publishing in journals such as *Porta Linguarum*, *Revista de Educación*, and *RESLA*.

Francisco Lorenzo is a full professor at Universidad Pablo de Olavide (Seville, Spain). He has held visiting scholar positions at Harvard University, University of London and University of Jyväskylä. His research focuses on the study of second language acquisition and bilingualism, sociolinguistics and sociology of language, and European language policies. He has authored more than sixty publications in journals such as *Applied Linguistics*, *Language Policy*, *System*, and *Language and Education*.

NOTES

¹ Throughout the manuscript, when *less* is followed by an adjective, it is always functioning as an adverb, not a determiner (e.g., by *less polysemous words* we mean *words which are less polysemous*, and not *fewer polysemous words*).

Acknowledgements

This research was supported by the European Regional Development Fund (ERDF, 80%) and by the Department of Economy, Knowledge, Business and University of the Andalusian Regional Government (20%), within the framework of research project UPO-1380541.

Appendix. Mean values of word length, word frequency, polysemy, and hypernymy

| | Mean number of syllables per word | Mean number of letters per word | Mean word frequency | Mean polysemic value of content words | Mean hypernymic value of nouns | Mean hypernymic value of verbs |
|--------|-----------------------------------|---------------------------------|---------------------|---------------------------------------|--------------------------------|--------------------------------|
| Year 1 | 1.51 | 4.48 | 2.41 | 3.64 | 5.31 | 1.36 |
| Year 2 | 1.62 | 4.93 | 2.37 | 3.96 | 5.94 | 1.27 |
| Year 3 | 1.54 | 4.63 | 2.43 | 3.74 | 5.15 | 1.33 |

Table 3. Mean values of word length, word frequency, polysemy, and hypernymy.

