

Linguistic indicators of text quality in analytical texts: developmental changes and sensitivity to pedagogical work (*Indicadores lingüísticos de la calidad textual en los textos analíticos: cambios evolutivos y sensibilidad al trabajo pedagógico*)

Melina Aparici^a, Rocío Cuberos^b, Naymé Salas^c and Elisa Rosado^b

^aDepartament de Psicologia Bàsica, Evolutiva i de l'Educació, Universitat Autònoma de Barcelona;

^bDepartament d'Educació Lingüística i Literària, i Didàctica de les Ciències Experimentals i de la Matemàtica, Universitat de Barcelona; ^cDepartament de Didàctica de la Llengua i la Literatura, i de les Ciències Socials, Universitat Autònoma de Barcelona

(Received 1 August 2019; accepted 14 January 2020)

ABSTRACT

Mastering analytical writing involves the proficient use of varied later-acquired grammatical, lexical and discourse forms and functions. Developmental studies have identified specific linguistic features as diagnostic of increasing proficiency. This study examines how these features change throughout educational levels and before and after the implementation of a set of classroom activities aimed at homogenizing participants' pedagogical input while raising their awareness of analytical texts main functions. Two hundred and twelve Spanish speakers from primary, secondary and university levels participate in the study, each producing two analytical texts. The corpus consists of 424 texts produced on the same topic. Results indicate that not all the features pertaining to the same domain show identical changes across educational levels or are identically sensitive to pedagogical work. Productivity measures increase with educational level and after pedagogical work, except for university students. Most lexical, syntactic, and discourse-level features measures improve with educational level, though their sensitivity to pedagogical work is not straightforward. Findings suggest the need of evaluating writing quality at a high level of granularity.

KEYWORDS

analytical texts, text quality, Spanish, primary to higher education

RESUMEN

El dominio de la escritura analítica requiere un uso avanzado de diversas funciones gramaticales, léxicas y discursivas de adquisición tardía. Los estudios sobre el desarrollo han identificado algunos aspectos lingüísticos concretos como indicadores de un dominio creciente. En este estudio se analiza el cambio que experimentan esos aspectos en diferentes niveles educativos y antes y después de la implementación de una serie de actividades en el aula, diseñadas para homogeneizar la exposición pedagógica de los participantes e incrementar sus conocimientos de las características de los textos analíticos. En el estudio participaron 212 estudiantes hispanohablantes de educación primaria, secundaria y educación superior; cada uno de ellos produjo dos textos analíticos. El corpus está compuesto por 424 textos sobre un mismo tema. Los resultados indican que no todos los aspectos de un mismo dominio lingüístico muestran los mismos cambios en todos los niveles educativos o son igualmente sensibles al trabajo pedagógico realizado. Los indicadores de productividad aumentan con el nivel académico y tras la implementación de las actividades pedagógicas, excepto en el caso de los estudiantes universitarios. La mayoría de los indicadores de aspectos léxicos, sintácticos y discursivos mejoran con el nivel académico, aunque su sensibilidad al trabajo pedagógico no es homogénea. Los resultados sugieren la necesidad de evaluar la calidad del texto escrito con un alto nivel de granularidad.

PALABRAS CLAVE

Textos analíticos, calidad textual, español, educación primaria a universitaria

Translation from English / *Traducción del inglés*: Mercè Rius

CONTACT Melina Aparici, Departament de Psicologia Bàsica, Evolutiva i de l'Educació,
Universitat Autònoma de Barcelona, Edifici B, 08193 Bellaterra, Barcelona, España. E-mail:
melina.aparici@uab.cat

When composing a text, writers must integrate local decisions—such as lexical choices, syntactic structuring of clauses or linkage between clauses—into the overall discourse organization of the text. Developmental studies of writing have examined the specific ways in which linguistic devices are realized in different discourse genres, and have come up with a set of text-embedded features that serve as indicators of writing quality (Berman & Nir-Sagiv, 2009; Crossley, Weston, McLain, Sullivan, & McNamara, 2011). The notion of discourse genre is central in this endeavour. *Genre* refers to the text types that fulfill particular functions, as defined by sociocultural norms and conventions (Halliday & Hasan, 1985). It exposes speaker/writer's expectations as to what should be written. Proficient writers' local decisions in text composing are not only guided by their knowledge of the topic and the particular features of the writing task, but also by genre-specific writing plans stored in long-term memory (Alamargot & Chanquoy, 2001).

There are different approaches to genre analysis but the important differences among them revolve around the relative attention given to text (i.e., a linguistic orientation) or the social context (i.e., a sociological orientation or contextual approach) (Flowerdew & Wan, 2010). In the present study we take a linguistic orientation; we examine a selection of lexical, morphosyntactic and discourse indicators of writing quality and follow the development of such indicators in the texts produced by children, adolescents, and young adults in response to prompts proposed to elicit texts in analytical genres¹, that is, texts that focus on analysis and argumentation (Schlepppegrell, 2004; Uccelli, Deng, Phillips Galloway, & Qin, 2019) and combine expository and argumentative components (Vilar, Tolchinsky & Aparici, submitted). They are characterized by the need for the writer to express their point of view on a given topic, present arguments and

counterarguments, and provide adequate evidence and/or possible rebuttals (Toulmin, 2002).

Their expected result is that the reader is led to share the author's viewpoint, persuaded by the writer's solid argumentation (Driver, Newton, & Osborne, 2000).

Language demands imposed by the nature of analytic texts are challenging not only for secondary but also higher education students (Tolchinsky, Aparici, & Rosado, 2017). Progress in mastering different genres has been shown to move from: (1) personal genres, e.g., narratives and retellings to (2) factual genres, e.g., procedures and reports, and, finally, to (3) analytic genres, e.g., explanations, expositions, argumentations (Schleppegrell, 2004). While written narrative organization tends to be well mastered by around age 10, analytic writing constitutes a protracted accomplishment. This is because it requires a proficient use of varied later-acquired grammatical, lexical, and discourse forms and functions (Uccelli, Dobbs, & Scott, 2012). In analytical texts writers are expected to use diverse but precise vocabulary, package information (e.g., by means of complex syntax or nominalizations) and make discourse organization explicit (e.g., by using markers to signal text transitions) (Schleppegrell, 2004; Snow & Uccelli, 2009).

Learning to write analytical texts is a major requirement for academic and professional development. We aimed to test the "macrodevelopmental changes" that these indicators may suffer with age/school level, and whether they are sensitive to pedagogical action, that is, the implementation of a set of instructional activities ("microdevelopmental changes").

Distinguishing between types of indicators (i.e., micro- and macrodevelopmental), should be key in informing psycholinguists and educational practitioners. Specifically, identifying which text-features develop with educational level would be useful for establishing patterns of linguistic and pragmatic development, for comparing discourse competence between typical and atypical populations, and between native and non-native writers. On the other hand, gaining information

about the sensitivity of these features to pedagogical work might be relevant for defining benchmarks at different educational levels, and for designing and evaluating future interventions.

Measures of text quality in analytical writing development

The quality of a text can be partially explained by examining its linguistic features (McNamara, Crossley, & McCarthy, 2010; Salas, Llauradó, Castillo, Taulé, & Martí, 2016). Different linguistic measures have been suggested as indicators of development and quality at different text levels. A general proxy of text quality is *productivity*, a feature that refers to the number of written words, sentences, and ideas in a composition (e.g., Kim et al., 2011), and is a robust measure of writing development across languages (Wagner, Puranik, Foorman, Foster, Tschinkel, & Kantor, 2011). The strength of productivity as a proxy for text quality (e.g., Berman & Nir-Sagiv, 2009) diminishes during elementary school for narrative and expository texts (Abbott & Berninger, 1993). Since a minimum length is necessary to elaborate a line of argumentation as the one required in writing analytical texts, we assume that productivity would still be a relevant indicator of quality beyond elementary school.

Lexical indicators of text quality

At the lexical level, word length, lexical diversity and lexical density are among the most popular indicators of text quality. All these measures have shown developmental patterns and are affected by text genre.

Word length, operationalized by average number of letters per content word, is regarded as diagnostic of literacy level (Malvern, Richards, Chipere, & Durán, 2004). It is an indicator of lexical complexity (Penris & Verspoor, 2017), and capitalizes on the fact that word length and frequency of use are inversely related: as length increases, frequency decreases (Zipf, 1932). It is widely used in corpus linguistics research as a way of gauging lexical sophistication

(Riedemann, 1996); it reflects both the advanced use of sophisticated, precise, low-frequency terms (Biber, 1995), and an increased command of structurally complex derivatives (Anglin, 1993). Word length shows a clear developmental pattern and is also affected by text genre: as a function of age, the length of words has been shown to increase (Llauradó & Tolchinsky, 2013) and, when compared with narratives, expository texts show significantly longer words (Cutillas, Tolchinsky, Rosado, & Perera, 2014).

Lexical density accounts for the relative weight of content words in the text. Indeed, nouns, adjective and verbs have a higher informative load than function words such as prepositions or articles. A text with a high proportion of content, informative words contains more information as compared to one with a high proportion of function words (Johansson, 2008; Gregori & Clavel, 2015).

Finally, lexical diversity is an estimate of the richness of vocabulary in a text. It is usually assessed as a type-token ratio (TTR), that is, the ratio obtained by dividing the number of different words (types) by the total number of words (tokens) (Bates, Bretherton, & Snyder, 1988). Both measures increase as a function of age (Berman & Nir-Sagiv, 2010; Malvern et al., 2004; McCarthy & Jarvis, 2010) and have shown higher indexes in expository than in narrative texts (Cutillas et al., 2014).

Two other indicators of an advanced, literate lexicon relate to the use of adjectives and nominalizations. The two devices affect nominal phrase (NP) complexity while realizing relevant features of analytical texts (Baratta, 2010; Ravid & Levie, 2010). NP complexity has been shown to increase with age and schooling experience in different languages, particularly, in written expository (e.g. Ravid, van Hell, Rosado, & Zamora, 2002, for Hebrew, Dutch, Spanish, and English).

Adjectives favor a more precise identification of referents, while nominalizations turn actions and qualities into abstract objects, remove human participants and compact information to make the material more concise (Schleppegrell, 2004). The increase in adjectives coincides with the consolidation of a high-register lexicon (Ravid & Levie, 2010). Likewise, production of nominalization indicates sophisticated and complex lexical uses, as well as analytic discursive functions required to express writer stance, for example, persuasive argumentation, personal opinion and/or evaluation (Charles, 2003; Hyland, 2005).

Syntactic indicators of text quality

The degree of syntactic complexity in a text is also assumed to fulfill rhetorical aims (Langacker, 1999). Compacting ideas in a single unit enables readers/writers to process them simultaneously in a more integrated way (Chafe, 1994). Measures of complexity look at the kind of *syntactic linkage* between units (e.g., juxtaposition, coordination, subordination) or at the number of lower-level units (e.g., words) included in higher-level units (e.g., phrases, clauses, discourse-level units); that is, the *syntactic density* within a given unit (Berman & Nir-Sagiv, 2009). Both the kind of linkage and the syntactic density, have been found to be associated with writing quality and proved sensitive to educational level and genre (Biber, 1995; Crossley & Mcnamara, 2010; Snow & Uccelli, 2009).

A number of studies found that, with age and schooling, writers tend to include a higher proportion of subordinate clauses (Crossley et al., 2011; Wagner et al., 2011), and this finding holds for various languages such as Hebrew (Berman & Nir-Sagiv, 2009), French (Mazur-Palandre et al., 2011) or Spanish (Aparici, 2010). Syntactic complexity in writing, mainly indexed by use of subordinate clauses, is also sensitive to genre. Indeed, the use of subordination has been considered a marker of genre adequacy (Snow & Uccelli, 2009). In English, more

complex structures were found in argumentative texts than in descriptive, narrative, or compare-and-contrast genres (Beers & Nagy, 2009). In Spanish, Aparici (2010) found more subordination in expository than in narrative texts. In addition, syntactic complexity, indexed by the use of subordinating conjunctions, positively correlates with text quality ratings (Crossley & McNamara, 2010). In analytical texts, adverbial subordinate clauses are the syntactic locus for reasons, explanations, and causes, which are essential for building a sound line of reasoning and support the writer's point of view.

Similarly, several measures of syntactic density have shown to be sensitive to the age of the writers and to the genre being produced, as well as to text quality ratings. For example, McNamara et al. (2010) found that syntactic density in terms of number of words before the main verb was one of the most predictive indices of the quality of essays produced by college students. In Crowhurst's (1983) developmental study, argumentative essays with high grammatical complexity received significantly higher quality ratings than essays with low grammatical complexity. The association between complexity and quality increased significantly with educational level.

Discursive indicators of text quality

Beyond clause linkage, texts' texture is accomplished by different connectivity devices, such as discourse markers (DM). DMs are a set of a relatively closed class of linguistic elements that function to signal relationships between units of discourse and the meanings they encode (Choi, 2007). Sanders and Noordman (2000) argue that DMs 'guide the reader in selecting the right relation' (p. 56). Previous research suggests that the ability to mark relationships between units of discourse is developed relatively late (Katzenberger & Cahana-Amitay, 2002). However, Crossley and McNamara (2010) examined argumentative essays written by college freshmen and

scored by expert raters and found no evidence that cohesion cues (explicit cue-based, local cohesion) and essay quality are related.

Besides DMs, texts unfold through a number of differentiated rhetorical moves in order to fulfill their communicate goal. Rhetorical moves refer to segments of discourse that fulfill a particular communicative function (Upton & Cohen, 2009). Due to their nature, the characterization of rhetorical moves is genre-based. Irrespective of their quality or the function they fulfill in the text, rhetorical moves can be characterized by their density, that is, by their number of words or clauses. Rhetorical moves containing more clauses and/or words are presumably more informative and may contribute to a greater extent to the quality of a text.

To sum up, previous studies have identified a series of linguistic indicators of quality that comprise purely quantitative measures such as productivity, and specific linguistic features that operate both at local and text level (i.e. vocabulary, syntax, and discourse organization). Mastery of such linguistic devices arguably results in a more felicitous arrangement of writers' arguments and evidences and, therefore, in a more successful accomplishment of the function of the text. The two-faceted approach, including the assessment of macro- and microdevelopmental changes should put us in a privileged position to distinguish linguistic features that require long developmental windows to be mastered, from those that are sensitive to pedagogical work in the short run.

This study

This study is aimed at addressing written composition of analytical texts, and our goal was twofold: (1) identifying macrodevelopmental (age/schooling-related) changes in a set of text-embedded lexical, syntactic and discourse features that research has shown to be diagnostic of written language quality by comparing these features in analytical texts produced by primary,

secondary and university students; and (2) examining whether these features change in a limited period of time, that is, during the application of a set of instructional activities (microdevelopmental changes) by comparing these features in the analytical texts produced by students before and after instructional activities.

The design of a common pedagogical environment enabled us to take this dual (macro- and microdevelopmental) perspective. This was not an intervention study since our purpose was not to demonstrate that this set of activities works better than other possible treatments, but to provide a common framework for examining the development of the target features. During the time of data collection participants were subject to identical pedagogical work at school to ensure that the developmental interpretation be rigorous and unadulterated. We assumed that writers' knowledge of the function of a given genre mobilizes the linguistic devices that are best suited to fulfill such function (Berman, Ragnarsdóttir, & Strömqvist, 2002).

Accordingly, we implemented a set of instructional activities aimed at raising students' awareness of the function of analytical texts. In order to establish their baseline level in the use of the target linguistic devices, we obtained text samples before the pedagogical work as well as samples after the last session. In line with the studies reviewed above, we expected:

(1) that the educational level will have an effect on productivity measures and lexical features: older students would produce longer texts, including higher levels of lexical density and diversity, more occurrences of nominalizations and adjectives than younger students. We also expected an effect of educational level for the syntactic and discourse level measures.

(2) the target measures to show moderately higher values upon the implementation of the activities aimed to raise students' awareness of the features of analytical texts.

In contrast with the hypotheses for macrodevelopmental changes, our expectations regarding microdevelopmental changes are theoretically rather than empirically motivated, due to the lack of empirical studies on the effect of pedagogical work on the use of specific linguistic indicators in analytical texts.

Method

Participants

Two-hundred and twelve students from three educational levels—6th grade, 10th grade and university—participated in the study, a cross-sectional design with repeated measures. In Spanish curriculum, 10th grade is the 4th and last year of obligatory secondary education. Participants were recruited and tested in Castilla-León and Castilla-La Mancha, Spain, in two primary public schools and three secondary public schools, and one university faculty.

After convenience sampling, the sample consisted of 65 primary-school students, 78 secondary-school, and 69 university students. Mean age, age range and gender by educational level are displayed in Table 1. No participant was excluded due to sociolinguistic background or for having learning disabilities.

<INSERT TABLE 1>

Materials and procedure

Participants were exposed during seven lessons to a set of instructional activities carried out by their regular teacher in their intact classrooms. Teachers were trained to apply the same set instructional activities, and were instructed on wording and prompting the tasks following the study design, in order to guarantee the same pedagogical context in all the groups. The activities

were identical for the three educational levels but teachers' verbal instructions were slightly modified to provide adequate scaffolding to the younger group of participants.

The instructional activities aimed to raise students' awareness of the features of analytical texts, and included readings about the topic at stake (*freedom of movement across countries*), class discussions, and production of analytical texts. Besides, some sessions were devoted to evaluating texts of students' peers by means of a rubric, and reflecting on key aspects of the produced texts.² During the study, each student produced five texts: Text 1 on topic A (*freedom of dress*), before the onset of instructional activities; texts 2, 3, and 4 on topic B (*freedom of movement*), before, during, and immediately after the application of the set of activities, respectively; and text 5 on topic C (*rewards and punishments*), one month after the end of the instructional activities. In the present study, we examined data belonging to Texts 2 and 4, both on the same prompt, produced before the onset of activities and immediately after completion of the instruction.

All texts obtained were transcribed and coded using CHAT format (CHILDES project; MacWhinney, 2000), and data were analyzed using CLAN programs. The corpus consisted of 424 texts that were segmented into clauses, which was the basic unit of analysis, following Berman and Slobin (1994).

Measures

Measures were obtained from the texts to capture linguistic features of text quality in the texts produced by participants, across four dimensions: productivity measures, lexical measures, syntactic- and discourse-level measures (see previous section on Measures of Text Quality in Analytical Writing Development).

Productivity measures

Number of words. The total number of words was obtained as a measure of productivity at the word level. Word counts were obtained automatically, with CLAN commands, from the transcribed texts.

Number of clauses. The total number of clauses was obtained as a measure of productivity at the sentence level. Clause counts were obtained automatically, with CLAN commands, from the transcribed texts. Twenty percent of the texts were segmented twice by two research assistants with a background on Linguistics. Inter-rater reliability was assessed using the intra-class correlation coefficient (ICC, two-way mixed effects-absolute agreement). ICC was .998 (95% CI: .997, .999).

Number of rhetorical moves. We identified manually argumentative (expressing the author's positioning), and expository moves (supporting the author's positioning). Inter-rater agreement was calculated in five percent of the texts considering the opening and closure of the rhetorical movements. The proportion of absolute agreement was .89. Since each rhetorical move had been coded in the CLAN transcriptions, they were also counted automatically.

Lexical level measures

Word length. We calculated the mean number of letters of content words; that is, words tagged as (common) nouns, (main) verbs, adjectives and adverbs. Non-grammaticalized (deadjectival) adverbs were considered content words. The morphological analyzer of CLAN, plus manual disambiguation allowed to obtain this measure.

Lexical density. This measure was the ratio of content words by the total number of words, and was obtained automatically with CLAN programs.

Lexical diversity. We used the D measure, implemented in CLAN under the name of VOCD (MacWhinney, 2000), an estimate of lexical diversity that is not affected by the length of the texts, as other measures of lexical diversity (Malvern et al., 2004).

Use of adjectives. We computed the proportion of adjectives out of the total number of words. Only descriptive adjectives were considered.

Use of nominalizations. We used CLAN morphological analyzer and manual disambiguation to obtain this measure, and then computed the proportion of nominalizations out of the total number of words. Both deverbal and deadjectival nouns were considered nominalizations.

Syntactic level measures

Syntactic density. We computed the mean number of words per clause, using the automatic counts of both words and clauses.

Use of subordination. We identified and manually coded subordinate clauses in the texts and computed the proportion of subordinate clauses out of the total number of clauses. All types of subordinate clauses, both finite and nonfinite, were counted for each text, but we further distinguished adverbial and relative clauses in the following measures.

Adverbial clauses. We identified all adverbial clauses in the texts and computed the proportion of adverbial clauses out of the total number of clauses. This type of clauses carries important semantic information in the texts analyzed here, such as causes and consequences. They add additional information and are complementary to the sentence argument structure; their use is, thus, optional, and subject to the rhetorical preferences of the writer.

Relative clauses. We identified all relative clauses in the texts and computed the proportion of relative clauses out of the total number of clauses. This type of clauses function as noun complements creating complex NPs and typically enrich and refine information carried by nouns.

Discourse level measures

Discourse markers. We identified all discourse markers (DMs) in the texts following Martín Zorraquino & Portolés' (1999) criteria and computed the proportion of DMs out of the total number of clauses.

Discourse density. We computed the mean number of words per rhetorical move. We included this measure as an indicator of how much information is packed into the boundaries of a discourse-level unit.

Results

Table 2 displays the means and standard deviations of all measures by educational level and text. We report on a series of two-way repeated-measures ANOVAs, with Educational level as the between-subject variable (primary vs. secondary school, vs. university) and Text as the within-subject variable (Text 1 vs. Text 2, henceforth T1 and T2, respectively). The data analyzed showed, in general, to be normally distributed, with skewness values under 3 and kurtosis values under 10 (absolute values; Kline, 2011). We, however, identified a few extreme outliers ($> 2,5$ SDs), which corresponded to participants that had produced considerably more text (in words, clauses, or rhetorical moves) than their peers within the same educational level. In addition, the assumption of homoscedasticity of variances was not met for several variables. For these reasons, we performed a two-way mixed-ANOVA analyses in R, using function *bwtrim* from the package WSR2, which is robust to deviations from normality and to heteroscedasticity, and

provides the statistic Q for each of the main effects, and for the interaction between them. In the case of significant interactions, we calculated the simple effect of Educational level with one-way ANOVAs using function *t1way* from the same package. Similarly, we calculated the simple effect of Text using function *yuend*, which provides a robust t-test for related samples (Mair & Wilcox, 2016). In these cases, multiple comparisons were applied the Bonferroni correction to avoid Type I errors.³

<INSERT TABLE 2>

Macro- and microdevelopmental changes in productivity

All productivity measures, at the word, sentence, or discourse level, were affected by the educational level of the writers, so that more text was generated by more experienced participants (Table 3). Increasing the amount of text was also the typical response to the pedagogical work, since there were significant Text effects for all productivity measures (Table 3). However, these main effects were moderated by a significant interaction, $Q(2, 110.42) = 25.42, p < .001$, for number of words; $Q(2, 109.97) = 20.88, p < .001$, for number of clauses; and $Q(2, 109.90) = 8.85, p < .001, \eta^2_p = .063$, for number of rhetorical moves. A follow-up of these interactions revealed that the Educational level effect was significant for both T1 and T2 ($p < .001$). However, *t*-tests of the simple effect of Text showed that, while younger participants increased the amount of text generated in T2 in terms of words or clauses ($ps < .001$), university students included a significantly smaller number ($ps \leq .017$). As for the number of rhetorical moves, primary-school children and university students did not significantly vary the number of rhetorical moves from T1 to T2 ($ps > .05$), in contrast to secondary students ($p < .001$).

<INSERT TABLE 3>

<INSERT TABLE 4>

<INSERT FIGURE 1>

Macro- and microdevelopmental changes in lexical measures

Most measures of participants' lexical choices (i.e., word length, lexical diversity, proportion of adjectives, and of nominalizations) experienced significant macrodevelopmental changes, given that a significant main effect of Educational level was found, except for lexical density (Table 3). However, only lexical density and lexical diversity changed after the pedagogical work that the students received, although while lexical density decreased for T2, lexical diversity increased (Table 4). A significant Educational level x Text interaction was found for average word length, $Q(2, 104.02) = 5.69, p = .005$; and the proportion of nominalizations, $Q(2, 100.51) = 5.29, p < .001$. Follow-up ANOVAs, to determine the simple effect of Educational level, showed that the effect was significant for T1 and T2 across both measures. T-tests of the simple effect of Text revealed different effects of the pedagogical work depending on the educational level of the participants. For the average length of words, primary-school children increased significantly the average length ($p = .006$), both secondary and university students showed no change across T1 and T2 ($ps > .05$). Finally, there were no significant differences at any educational level in the proportion of nominalizations in T1 as compared with T2; observing the means by group and text, however, younger students included more nominalizations in T2, on average, while University students appeared to show the reverse trend.

<INSERT FIGURE 2>

Macro- and microdevelopmental changes in syntax measures

Most measures of text-based syntax increased with educational level (Table 3). Nevertheless, only the proportion of relative clauses showed microdevelopmental changes (Table 4). In addition, syntactic density, that is, the average number of words per clause, was modified by a significant Educational level x Text interaction, $Q(2, 109.17) = 4.86, p < .001$. A follow-up of the interaction revealed that the Educational level effect was significant for both T1 and T2 when comparing primary with older groups (all $ps < .001$), while secondary-school and university students did not differ from each other in T1 or T2 ($ps > .016$). As for the simple effect of Text, only primary school children included significantly more words per clause in T2 than in T1 ($p = .003$), but secondary school and university students included a similar proportion of words per clause in both texts ($p > .05$).

<INSERT FIGURE 3>

Macro- and microdevelopmental changes in discourse measures

Both discourse-level measures, the proportion of discourse markers and discourse density (number of clauses per rhetorical move), experimented a general increase with Educational level (Table 3). The nature of the changes as a function of the pedagogical work that participants received was against our initial expectations. First, there was no effect of Text in the proportion of discourse markers (Table 4), nor was this measure affected by a significant interaction, $Q(2, 104.31) = 2.36, p = .099$. Second, there was no effect of Text in the density of rhetorical moves (Table 4). However, this measure was affected by a significant interaction, $Q(2, 104.91) = 4.70, p = .011$ (Figure 3). One-way ANOVAs of the simple effect of Educational level indicated that the effect was significant for both T1 and T2, when comparing primary with university students in T1 ($p < .001$), and when comparing primary and secondary with university students in T2 (all $ps < .016$). Moreover, t-tests of the simple effect of Text revealed that only primary school children

displayed significantly higher levels of discourse density in T2 than in T1 ($p = .015$), while it remained stable in the secondary ($p = .545$) and the university groups ($p = .358$).

Discussion

We examined a set of lexical, morphosyntactic, and discourse features in the written essays of students from primary, secondary, and university level controlling for pedagogical input. This enabled us to identify, on the one hand, the text-features that develop with educational level, that is, the macrodevelopmental changes in the indicators of quality of analytical texts. And, on the other hand, the text-features that develop in the context of the instructional activities, that is, features that undergo microdevelopmental changes.

Our study yielded two main results. Firstly, productivity, i.e., the amount of text that students produced, a feature which involves every domain of language, confirms its suitability as a measure of evolving writing quality. Secondly, not all text-features that pertain to the same language domain followed the same developmental patterns, or showed similar sensitivity to pedagogical work. In the following lines we elaborate on these findings and discuss their implications for the evaluation of writing quality and for educational interventions.

Macro- and microdevelopmental of written productivity

In line with findings on the development of writing in different discourse genres, productivity reflected both macro- and microdevelopmental changes (Berman & Nir-Sagiv, 2009; Crossley & McNamara, 2010; Salas et al., 2016) and may well be a good proxy for overall text quality, as it is sensitive to changes across several domains. This recurrent finding suggests that schooling (and increasing experience with written language) facilitates writers to expand on topics and arguments and, thus, to produce longer texts (Crossley et al., 2011). Other factors (e.g., age,

content knowledge) being equal, producing more text can arguably lead to statements and reasons being more precise. Likewise, an increase in syntactic complexity requires additional words and clauses (e.g., relative clauses; Scott, 2009).

The pedagogical activities we implemented, aimed to discuss and read texts about a topic, impinged on text productivity, but mainly during the learning-to-write years. University students may need more intensive or different kinds of activities in order to lengthen their texts. An alternative explanation is that these students might be making changes to the texts that are essentially different from those made by younger students. For example, while younger writers could be adding more information by simply adding more words/clauses, older students could equally make their texts more informative without necessarily making them longer. In other words, their rhetorical choices may imply a more sophisticated reorganization of their linguistic resources, rather than the production of a larger quantity of text. This interpretation, that university-level students are concerned with their rhetorical choices during text composition, is well aligned with models of competent writing (Alamargot & Chanquoy, 2001; Hayes, 2011). Future research should strive to fine-grain the precise ways in which text productivity indicates aspects of text quality.

Different sensitivities within the same domain

A second crucial finding of the present study was that, with the exception of productivity, measures pertaining to the same domain (i.e., lexicon, syntax, discourse) showed differing sensitivities at both a macro- and a microdevelopmental level.

In the lexical domain, it was clear from our findings that more experienced writers used more sophisticated words, which were moreover more diverse. In addition, they also tended to use more complex morphological words, such as nominalizations, and more lexical items that

help increase the precision of statements and introduce nuances in discourse, such as adjectives, in line with previous research on the development of lexical choices in writing (Berman & Nir-Sagiv, 2009; Crossley et al., 2011; McNamara et al., 2010; Nippold, 2007). The only exception was lexical density which, in line with Llauradó and Tolchinsky (2015), but contrary to Cutillas et al. (2014), showed no developmental changes. Lexical density has been used as a measure of lexical development in different languages (mostly English, but also Catalan, French, Hebrew, and Swedish) with uneven findings, which might point to language-specific trends (Johansson, 2008). Indeed, it is a measure that is based on the use of function words. For example, faltering use of function words by young children (in some languages) may make their texts appear denser than those produced by older, more proficient writers.

Some lexical measures were sensitive to the pedagogical work and showed micro-developmental changes. In particular, participants appeared to make different lexical choices: lexical diversity increased at all educational levels from T1 to T2. Participants also chose more sophisticated tokens, as evidenced by the longer average length of the content words included in the texts produced after the pedagogical work; however, only primary school children showed this effect. We would speculate that the more reduced knowledge-base in terms of vocabulary tokens of younger participants means that they also have larger gains to make in terms of learning of new or more complex vocabulary than older participants.

Different syntactic measures also showed distinct behaviors. Generally speaking, and in line with previous research on the topic (e.g., Beers & Nagy, 2009), more experienced writers tended to provide more information per clause, and produced more subordinated clauses; specifically, more relative clauses. Relative clauses, similarly to adjectives, add to NPs complexity, while they facilitate identification and characterization of entities in discourse. This

aspect of language structure and use goes hand in hand with a general increase in clausal complexity and is considered a yardstick in syntactic and discourse development (Ravid & Berman, 2010). Unexpectedly, the use of adverbial clauses decreased with educational level. Adverbial clauses are a preferred site for writers to provide their reasons and explanations to support their arguments, however, this category has proven to be an extremely wide umbrella term comprising many different forms and functions. For instance, in the developmental study by Aparici (2010), age effects were only observed when discriminating between different subcategories of adverbial clauses. Further work is necessary to distinguish among subcategories of adverbial clauses that convey information relevant to argumentative writing such as causal or consecutive clauses, from those that, for instance, locate (the reported) events in time and/or space.

Only two syntactic measures were sensitive to the pedagogical work implemented. First, relative clauses were used more profusely by participants across groups in T2, that is, after a series of lessons aimed to raise awareness about the specific features of analytical texts. Second, primary-school children, but not secondary or university-level students, included more words per clause on average in T2 as compared to T1. We have mentioned above that more complex clauses should be more information-rich, which might be one of the more immediate effects of the pedagogical work on the younger participants. Secondary and university-level students, in contrast, might be more prone to resort to other resources for increasing informational load, which do not necessarily increase the number of words in the clause.

The two features selected for measuring writing quality at a discourse level showed similar developmental trends but a different sensitivity to the pedagogical work. Both discourse density and the use of DMs increased with educational level, thus confirming that the ability to

skillfully connect pieces of discourse develops with age (e.g., Katzenberger & Cahana-Amitay, 2002). In terms of microdevelopmental trends, and contrary to expectations, the use of DMs did not change from T1 to T2. Both the fact that non-conventional markers operating at the sentence level (Cuenca, 2013) were not considered, and that non-canonic (i.e., semantically inadequate) uses of such markers were not identified, might account for these results. The use of explicit devices such as DMs is but one of the several means for rising the coherence of a text. In this sense, we concur with previous assessments that argued that the coherence of a text may not reside in the incidence of cohesive devices, but more likely in their absence (Crossley et al. 2011). The results we obtained for syntactic and discourse density (clear developmental changes and limited effect of pedagogical work), as well as for the use of DMs call for a thorough exploration of the different devices that writers marshal for “expressing notions and for compacting semantic content” (Slobin, 1977, p. 187). In this respect, further research might consider, for instance, the use of qualitative measures concerning the quality of the argumentation structure and the adoption of discourse stance (Uccelli et al., 2019; Vilar et al., submitted), the establishment of illegal semantic relations, or non-canonical uses of lexical devices (e.g., intra- and intersentential connectives).

Finally, with regards to micro-developmental effects on discourse density, only primary-school children showed an increase in the density of rhetorical moves from T1 to T2. Similarly to what we reported for the clausal level, at the discourse level other resources apart from addition of same-level units (i.e., more words in clauses or more clauses in rhetorical moves) are mobilized by participants to be maximally expressive.

In order to fully understand the process of mastery of a discourse genre (e.g., analytical texts) it is crucial to identify which text features change with age/schooling level and which ones

change in a more limited period of time, in the context of monitored, explicit pedagogical work. The use of an identical sequence of activities allowed us to single-out developmental changes in features of text quality. Nevertheless, the different behavior of the text-features comprised in each language domain suggests that both evaluation of writing quality and pedagogical work must be approached at a high level of granularity (i.e., changes may not have been found in the lexical domain as a whole, but we did find them in specific devices within this domain). Detecting which specific features function as the best indicators of text quality and which are more responsive to pedagogical work may contribute to a better understanding of the processes underlying the development of analytical writing throughout different educational levels. But, crucially, this information would also inform future intervention studies and should be helpful in orienting writing assessment practices.

Limitations of the Study

As with any study of a similar nature, this one suffers from a series of limitations. First, writing development is typically characterized by great inter-individual variation. Despite our sample not being small, a larger one would have probably allowed us to obtain more solid results and increased statistical power overall, so that seeming trends may have received support or could have been effectively refuted. Second, even though our study design was not meant to be an intervention, it would be interesting to compare the results with a control group, receiving a more traditional approach to writing instruction. Future research should strive to replicate the educational setting designed for the present study and compare it with a no-intervention condition. Finally, one of the strengths of the study, that is, its genre-focused approach, compromises the generalization of these results to other genres (e.g., narrative, description) which is, therefore, not warranted.

Notes

¹ The term genre is used to refer to a category of texts—oral or written—assigned on the basis of external criteria such as intended audience, purpose, and activity type, that is, it refers to a conventional, culturally recognized grouping of texts based on properties other than lexical or grammatical.

² For further detail about the instructional activities, see Vilar et al. (submitted).

³ The threshold for significance, after the Bonferroni correction, was $p < .016$.

Table 1

Mean age, age range and gender by Educational level

Educational level	Mean age	Age range	Participants		
			male	female	total
Primary level	11.6	11.1 – 12.1	30	35	65
Secondary level	15.8	15.0 – 18.3	33	45	78
University level	21	19.10 – 31.8	27	42	69

Table 2

Descriptive statistics for all measures in T1 and T2

Ed. Level	Text-based measures	mean (SD)	Min.-Max.	T1 mean (SD)	T2 mean (SD)
PRI	N.º words	207.38 (84.36)	84 – 388	85.89 (43.58)	121.49 (56.96)
SEC		467.09 (131.76)	179 – 936	201.36 (66.58)	265.73 (76.69)
UNI		702.18 (225.57)	312 – 1911	370.75 (148.10)	331.42 (95.10)
PRI	N.º clauses	38.11 (14.91)	14 – 74	16.58 (7.90)	21.52 (9.86)
SEC		74.37 (21.08)	18 – 138	31.88 (11.10)	42.49 (13.43)
UNI		105.49 (39.99)	45 – 312	55.81 (25.44)	49.68 (16.95)
PRI	N.º rhetorical moves	9.57 (3.93)	4 – 18	4.62 (2.44)	4.95 (2.26)
SEC		13.97 (3.04)	8 – 22	5.94 (2.32)	8.04 (2.17)
UNI		16.58 (4.94)	10 – 40	8.41 (4.26)	8.17 (2.95)
PRI	Word length	4.32 (2.64)	3.80 – 5.09	4.38 (0.35)	4.24 (0.29)
SEC		4.58 (2.30)	4.10 – 5.12	4.56 (0.27)	4.59 (0.24)
UNI		4.71 (1.61)	4.38 – 5.24	4.69 (0.19)	4.73 (0.18)
PRI	Lexical density	39.85 (0.39)	31.01 – 56.34	40.79 (4.34)	38.90 (4.75)
SEC		39.83 (0.28)	34.43 – 45.86	40.11 (3.16)	39.56 (3.40)
UNI		40.40 (0.23)	36.00 – 47.10	40.64 (2.76)	40.17 (0.27)
PRI	Lexical diversity	66.28 (25.27)	0 – 111.93	57.78 (33.16)	74.77 (28.37)
SEC		86.99 (15.31)	52.19 – 130.49	84.08 (17.55)	89.90 (18.67)
UNI		87.27 (12.11)	67.46 – 128.64	87.93 (16.28)	86.59 (12.93)
PRI	% adjectives	4.18 (2.00)	0.45 – 10.26	3.69 (2.27)	3.57 (2.22)
SEC		5.13 (1.89)	0.86 – 9.24	5.19 (2.31)	5.60 (1.91)
UNI		5.96 (1.42)	3.14 – 9.48	6.25 (1.79)	6.08 (1.58)
PRI	% nominalizations	3.62 (1.73)	0 – 7.00	2.81 (2.30)	2.86 (1.97)
SEC		4.51 (2.05)	0 – 9.09	5.02 (2.05)	5.27 (1.94)
UNI		5.98 (1.33)	2.89 – 9.04	6.25 (1.79)	5.76 (1.45)
PRI	Syntactic density	5.75 (.99)	3.75 – 9.94	5.20 (0.91)	5.71 (1.00)
SEC		6.18 (.88)	4.60 – 8.81	6.45 (1.02)	6.38 (0.94)
UNI		6.72 (.79)	5.04 – 9.19	6.89 (1.03)	6.82 (0.91)
PRI	% subordinate clauses	62.09 (9.99)	22.22 – 81.25	61.63 (12.14)	60.42 (13.45)
SEC		60.14 (8.73)	26.76 – 74.23	63.13 (11.23)	60.69 (10.18)
UNI		62.71 (6.67)	51.04 – 80.46	60.26 (12.17)	62.80 (7.97)
PRI	% relative clauses	13.59 (6.02)	0 – 23.53	10.08 (7.61)	13.97 (8.61)
SEC		17.02 (6.44)	2.78 – 32.47	15.74 (8.33)	17.57 (7.07)

UNI		18.58 (5.84)	6.41 – 34.21	17.51 (6.16)	23.25 (7.74)
PRI		26.40 (9.01)	5.56 – 52.38	28.08 (13.59)	26.20 (11.31)
SEC	% adverbial clauses	26.34 (7.45)	7.04 – 43.40	27.92 (11.78)	26.75 (8.28)
UNI		18.58 (5.84)	6.41 – 34.21	23.90 (7.70)	23.16 (7.23)
PRI		1.49 (.86)	0 – 3.45	1.02 (1.03)	1.16 (1.00)
SEC	% discourse markers	1.52 (.75)	0 – 3.61	1.88 (1.01)	1.98 (0.88)
UNI		1.75 (.66)	0.39 – 3.75	1.76 (0.73)	1.53 (0.80)
PRI		5.34 (2.06)	2.00 – 12.80	5.05 (3.35)	6.37 (3.57)
SEC	Discourse density	5.17 (2.20)	1.67 – 12.54	5.62 (2.58)	5.08 (2.46)
UNI		6.24 (3.01)	2.38 – 20.80	7.18 (4.23)	6.48 (3.12)

Note. PRI primary school, SEC secondary school, UNI university adults

Table 3

Robust 2-way mixed-ANOVA test results for the main effect of Educational level

Text-based measure	Q(2)	p	Post-hoc tests (Bonferroni)		
			PRI-SEC	PRI-UNI	SEC-UNI
No. words	229.94	< .001	$p < .001$	$p < .001$	$p < .001$
No. clauses	142.07	< .001	$p < .001$	$p < .001$	$p < .001$
No. rhet. moves	49.01	< .011	$p < .001$	$p < .001$	$p < .001$
Word length	59.48	< .001	$p < .001$	$p < .001$	$p < .001$
Lexical density	1.08	.342		n/a	
Lexical diversity	16.63	< .001	$p < .001$	$p < .001$	$p = 1.00$
% adjectives	43.50	< .001	$p < .001$	$p < .001$	$p < .001$
% nominalizations	88.23	< .001	$p < .001$	$p < .001$	$p < .001$
Syntactic density	53.94	< .001	$p < .001$	$p < .001$	$p = .006$
% subord. clauses	0.29	.746		n/a	
% relative clauses	33.53	< .001	$p < .001$	$p < .001$	$p < .001$
% adverbial clauses	5.94	.004	$p = 1.00$	$p = .017$	$p = .007$
% discourse markers	24.37	< .001	$p < .001$	$p < .001$	$p = .046$
Discourse density	9.89	< .001	$p = .975$	$p < .001$	$p < .001$

Note. PRI primary school, SEC secondary school, UNI university adults

Table 4

Robust 2-way mixed-ANOVA test results for the main effect of Text

Text-based measure	Q(1)	<i>p</i>	Direction of effect
No. words	17.52	< .001	T1 < T2
No. clauses	12.77	< .001	T1 < T2
No. rhet. moves	12.18	< .001	T1 < T2
Word length	1.24	.267	n/a
Lexical density	10.25	.002	T1 > T2
Lexical diversity	10.41	.002	T1 < T2
% adjectives	< 0.01	.966	n/a
% nominalizations	0.02	.877	n/a
Syntactic density	1.64	.203	n/a
% subord. clauses	1.62	.205	n/a
% relative clauses	26.03	< .001	T1 < T2
% adverbial clauses	0.74	.392	n/a
% discourse markers	< 0.01	.992	n/a
Discourse density	0.633	.428	n/a

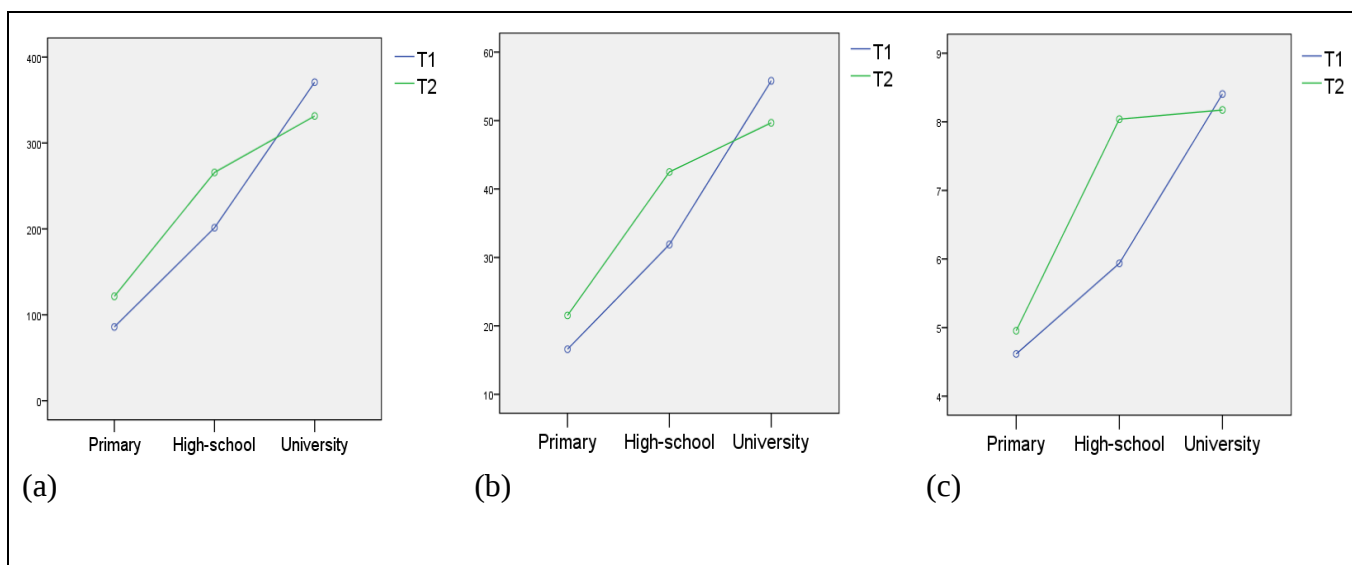


Figure 1. Graphs of the significant Educational level x Text interaction for the productivity measures: (a) number of words; (b) number of clauses; (c) number of rhetorical moves.

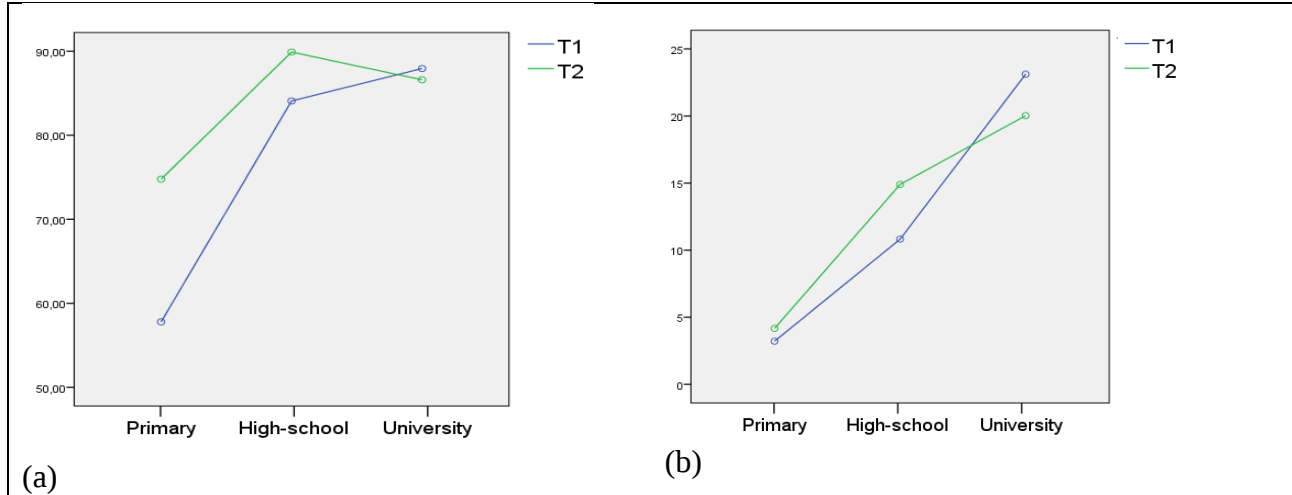


Figure 2. Graphs of the significant Educational level x Text interaction for the lexical measures:
(a) lexical diversity; (b) average content-word length.

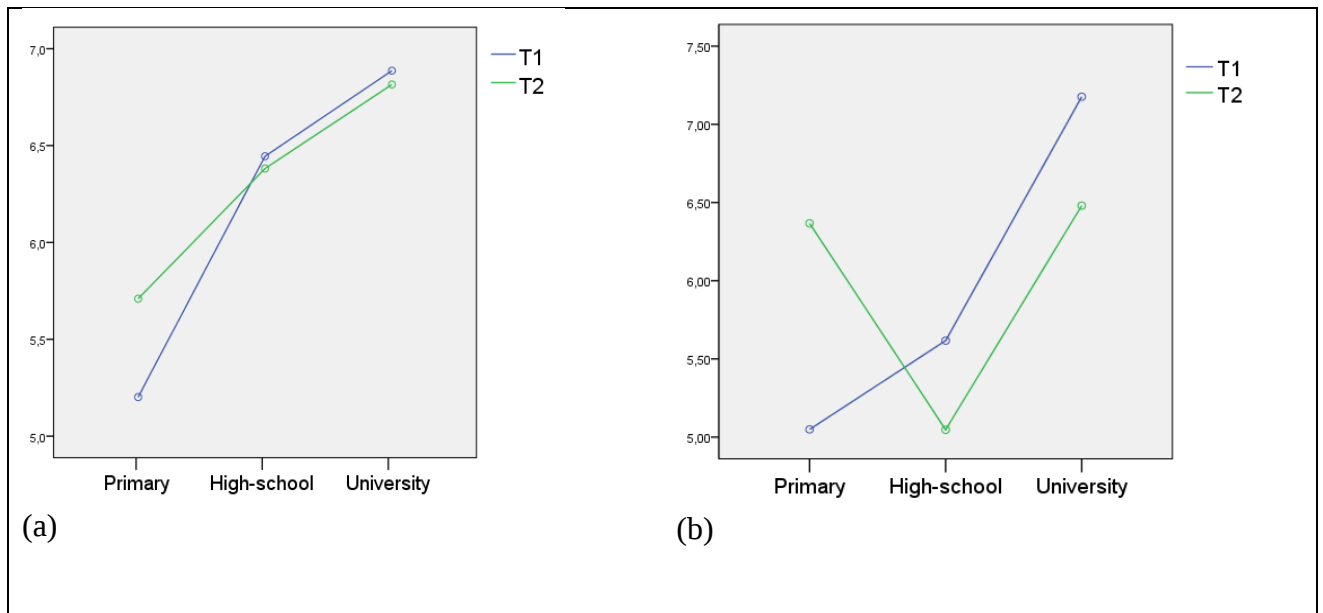


Figure 3. Graphs of the significant Educational level x Text interaction for the syntax and discourse measures: (a) syntactic density; (b) discourse density.

Indicadores lingüísticos de la calidad textual en los textos analíticos: cambios evolutivos y sensibilidad al trabajo pedagógico

Para producir un texto, los autores deben integrar decisiones locales —tales como sus elecciones léxicas, las estructuras sintácticas de las cláusulas o los enlaces entre ellas— en la organización discursiva global del texto. Los estudios del desarrollo de la escritura han analizado las formas concretas de ejecución de los mecanismos lingüísticos en los distintos géneros discursivos y han elaborado un conjunto de aspectos textuales que funcionan como indicadores de la calidad textual (Berman & Nir-Sagiv, 2009; Crossley, Weston, McLain, Sullivan, & McNamara, 2011). El concepto de género discursivo resulta clave en este sentido y refiere a los tipos textuales que desempeñan funciones determinadas, definidas por normas y convenciones socioculturales establecidas (Halliday & Hasan, 1985). Por medio del género se ponen de manifiesto las expectativas del hablante/escritor sobre el contenido del texto. Las decisiones locales de los escritores avezados en la composición textual no solo se guían por su conocimiento del tema o las características específicas de la tarea, sino por los planes de escritura específicos de género almacenados en la memoria a largo plazo (Alamargot & Chanquoy, 2001).

Las principales diferencias entre los diversos tipos de análisis del género textual giran en torno a la atención que se dedica al texto (enfoque lingüístico) o al contexto social (enfoque sociológico o contextual) (Flowerdew & Wan, 2010). En este estudio adoptamos un enfoque lingüístico, de modo que analizamos una serie de indicadores léxicos, morfosintácticos y discursivos de la calidad textual y seguimos su desarrollo en los textos elaborados por niños, adolescentes y jóvenes adultos producidos a partir de una serie de indicaciones proporcionadas para obtener textos analíticos¹, esto es, textos centrados en el análisis y la argumentación

(Schleppegrell, 2004; Uccelli, Deng, Phillips Galloway, & Qin, 2019) y que presentan componentes tanto expositivos como argumentativos (Vilar, Tolchinsky & Aparici, en revisión). Estos textos se caracterizan por la necesidad del autor de expresar su punto de vista sobre un tema, presentar argumentos a favor y en contra y ofrecer evidencia adecuada y/o posibles refutaciones (Toulmin, 2002). El resultado esperado es que el lector comparta el punto de vista del autor, persuadido por la solidez de su argumentación (Driver, Newton, & Osborne, 2000).

Las exigencias lingüísticas que impone la naturaleza de los textos analíticos suponen un desafío, no solo para estudiantes de secundaria sino incluso para universitarios (Tolchinsky, Aparici, & Rosado, 2017). Se ha demostrado que la evolución en el dominio de los diversos géneros textuales avanza del siguiente modo: (1) de géneros personales, p.ej., narraciones a (2) géneros fácticos, p.ej., procedimientos e informes y, por último, (3) géneros analíticos, p.ej., explicaciones, exposiciones, argumentaciones (Schleppegrell, 2004). Mientras que la organización de la narración escrita tiende a dominarse alrededor de los 10 años, la escritura analítica constituye un logro más tardío. Esto se debe a que requiere un uso avanzado de formas y funciones gramaticales, léxicas y discursivas cuya adquisición es también más tardía (Uccelli, Dobbs, & Scott, 2012). En los textos analíticos se espera que los autores hagan uso de un vocabulario diverso pero preciso, condensen la información (p.ej., mediante una sintaxis compleja o nominalizaciones) y expliciten la organización del discurso (p.ej., utilizando marcadores para indicar transiciones textuales) (Schleppegrell, 2004; Snow & Uccelli, 2009).

El aprendizaje de la composición de textos analíticos es esencial para el desarrollo académico y profesional. En este estudio nos proponemos, por una parte, identificar cuáles son los “cambios macroevolutivos” que experimentan estos indicadores en los diferentes cursos escolares o a distintas edades y, por otra, determinar si estos cambios son sensibles al trabajo

pedagógico; es decir, a la implementación de una serie de actividades educativas (“cambios microevolutivos”). Poder distinguir entre tipos de indicadores (p.ej., micro y macroevolutivos), debería ser clave para la formación de psicolingüistas y educadores. En particular, identificar qué aspectos textuales se desarrollan con la edad o nivel educativo sería muy útil para poder establecer patrones de desarrollo lingüístico y pragmático, comparar distintos niveles de competencia discursiva entre poblaciones típicas y atípicas o entre escritores nativos y no nativos. Por otro lado, disponer de información sobre cuán sensibles son estos aspectos a la intervención pedagógica resultaría relevante para definir estándares en distintos niveles educativos y diseñar y evaluar adecuadamente futuras intervenciones.

Medidas de la calidad textual en el desarrollo de la escritura analítica

La calidad de un texto puede explicarse parcialmente a partir del análisis de sus características lingüísticas (McNamara, Crossley, & McCarthy, 2010; Salas, Llauradó, Castillo, Taulé, & Martí, 2016). Se han propuesto diversas medidas lingüísticas como indicadores de desarrollo y calidad textual en diversos niveles. Un indicador general de la calidad textual es la **productividad**, que hace referencia al número de palabras, oraciones e ideas presentes en el texto (p. ej., Kim et al., 2011) y que constituye una medida robusta del desarrollo del lenguaje escrito en la mayoría de las lenguas (Wagner, Puranik, Foorman, Foster, Tschinkel, & Kantor, 2011). El poder explicativo de la productividad como indicador de calidad textual (p. ej., Berman & Nir-Sagiv, 2009) disminuye durante la educación primaria en el caso de los textos narrativos y expositivos (Abbott & Berninger, 1993). Dado que para elaborar una línea de argumentación como las exigidas en la composición de un texto analítico es necesaria cierta extensión mínima, asumimos que la productividad seguiría siendo un indicador relevante de la calidad textual más allá de la educación primaria.

Indicadores léxicos de la calidad textual

En el nivel léxico, la longitud de palabra, la diversidad y densidad léxicas son los indicadores más populares de la calidad textual. Todas estas medidas han mostrado patrones de desarrollo y se ven afectadas por el género textual.

La longitud de palabra, representada por la media del número de letras de las palabras contenido, se considera diagnóstica del nivel de competencia letrada (Malvern, Richards, Chipere, & Durán, 2004). Su consideración como indicador de complejidad léxica (Penris & Verspoor, 2017) se ve reforzada por el hecho de que la extensión de las palabras está inversamente relacionada con su frecuencia de uso de forma que esta disminuye según aumenta su longitud (Zipf, 1932). Esta medida se ha utilizado ampliamente en la investigación de la lingüística de corpus para evaluar el nivel de sofisticación léxica (Riedemann, 1996), ya que refleja tanto el uso avanzado de términos sofisticados, precisos y menos frecuentes (Biber, 1995) como el dominio creciente de palabras derivadas estructuralmente complejas (Anglin, 1993). La longitud de palabra es una medida que muestra un patrón evolutivo claro y que está también sujeta a la influencia del género textual: se incrementa en función de la edad (Llauradó & Tolchinsky, 2013) y, en comparación con los textos narrativos, los expositivos exhiben palabras significativamente más largas (Cutillas, Tolchinsky, Rosado, & Perera, 2014).

La densidad léxica da cuenta del peso relativo de las palabras contenido en el texto. Nombres, adjetivos y verbos tienen mayor carga informativa que elementos funcionales como las preposiciones o los artículos. Así, un texto con una proporción más elevada de palabras contenido contiene más información que uno con una gran proporción de palabras funcionales (Johansson, 2008; Gregori & Clavel, 2015).

La diversidad léxica, por su parte, constituye una estimación de la riqueza de vocabulario del texto. Su evaluación suele realizarse mediante el cálculo de la *token/type ratio* o TTR, medida que se obtiene resultado de dividir el número de palabras distintas (tipos) por el número de palabras totales (*token* o caso) (Bates, Bretherton, & Snyder, 1988). Ambas medidas aumentan en función de la edad (Berman & Nir-Sagiv, 2010; Malvern et al., 2004; McCarthy & Jarvis, 2010) y muestran índices más elevados en los textos expositivos que en los narrativos (Cutillas et al., 2014).

Por último, otros dos indicadores de desarrollo léxico son el uso de adjetivos y nominalizaciones; ambos repercuten en la complejidad del sintagma nominal (SN) y son característicos de los textos analíticos (Baratta, 2010; Ravid & Levie, 2010). En numerosas lenguas se ha demostrado que la complejidad de los SN aumenta con la edad y experiencia académica y, particularmente, en los textos expositivos (véase Ravid, van Hell, Rosado, & Zamora, 2002, para el hebreo, neerlandés, español e inglés).

Mientras los adjetivos facilitan una identificación más precisa de los referentes, las nominalizaciones convierten en abstractas acciones y cualidades, eliminando o minimizando los agentes humanos y condensando la información (Schleppegrell, 2004). El aumento en la producción de adjetivos coincide con la consolidación de un léxico de registro más elevado (Ravid & Levie, 2010). Del mismo modo, la producción de nominalizaciones indica unos usos léxicos más complejos y sofisticados, así como funciones discursivas analíticas necesarias para expresar la posición del autor, por ejemplo, la argumentación persuasiva, la opinión personal o la evaluación (Charles, 2003; Hyland, 2005).

Indicadores sintácticos de la calidad textual

También se asume que el grado de complejidad sintáctica de un texto desempeña funciones retóricas (Langacker, 1999). La condensación de ideas en una unidad permite a lectores/escritores procesarlas simultáneamente de manera más integrada (Chafe, 1994). Las medidas de complejidad examinan tanto el tipo de *vínculo sintáctico* entre las unidades (p. ej., yuxtaposición, coordinación, subordinación) como el número de unidades de bajo nivel (p. ej., palabras) incluidas en unidades de nivel superior (p. ej., frases, cláusulas, unidades de nivel discursivo), es decir, la *densidad sintáctica* dentro de una unidad determinada (Berman & Nir-Sagiv, 2009). Ambos tipos de medidas han demostrado estar relacionadas tanto con la calidad textual, como con el desarrollo educativo y las características del género textual en cuestión (Biber, 1995; Crossley & McNamara, 2010; Snow & Uccelli, 2009).

Diversos estudios revelan que, con la edad y la formación académica, los autores tienden a utilizar una mayor proporción de cláusulas subordinadas (Crossley et al., 2011; Wagner et al., 2011); y esto es así con respecto a diversas lenguas, entre ellas el hebreo (Berman & Nir-Sagiv, 2009), el francés (Mazur-Palandre et al., 2011) o el español (Aparici, 2010). En el caso de los textos escritos, la complejidad sintáctica, representada principalmente por el uso de oraciones subordinadas, es también sensible al género discursivo y, de hecho, se considera un indicador de adecuación al género (Snow & Uccelli, 2009). Por ejemplo, en el caso del inglés aparecen más estructuras complejas en los textos argumentativos que en los textos descriptivos, narrativos o comparativos (Beers & Nagy, 2009). También en español, Aparici (2010) observó una proporción mayor de subordinación en los textos expositivos que en los narrativos. Asimismo, la complejidad sintáctica, indicada por el uso de conjunciones subordinadas, se relaciona positivamente con las evaluaciones de la calidad textual (Crossley & McNamara, 2010). En los

textos analíticos, las cláusulas adverbiales subordinadas constituyen el locus sintáctico de razones, explicaciones y causas, elementos esenciales para crear una línea de razonamiento sólida, y respaldan el punto de vista del autor.

De la misma manera, se ha demostrado que diversas medidas de la densidad sintáctica son sensibles a la edad del autor y al género textual producido, así como a las valoraciones de la calidad textual. Por ejemplo, McNamara et al. (2010) observaron que la densidad sintáctica, en términos del número de palabras antes del verbo principal, era uno de los índices más predictivos de la calidad de los escritos producidos por estudiantes universitarios. En el estudio evolutivo de Crowhurst (1983), los textos argumentativos con un alto nivel de complejidad gramatical recibían puntuaciones de calidad significativamente más altas que los que presentaban un nivel de complejidad gramatical más bajo. La relación entre complejidad y calidad, pues, aumenta significativamente con el nivel educativo.

Indicadores discursivos de la calidad textual

Más allá de la conexión interclausal, el entramado de un texto se realiza por medio del uso de mecanismos de conexión tales como los marcadores discursivos (MD). Los MD constituyen una clase semicerrada de elementos lingüísticos que señalan las relaciones entre las unidades del discurso y el significado de las mismas (Choi, 2007). Para Sanders y Noordman (2000) los MD “guide the reader in selecting the right relation” (guían al lector en la selección correcta) (p. 56). Investigaciones previas sugieren que la capacidad de señalar relaciones entre unidades discursivas es un desarrollo relativamente tardío (Katzenberger & Cahana-Amitay, 2002). Sin embargo, Crossley y McNamara (2010) analizaron textos argumentativos redactados por universitarios de primer año y puntuados por evaluadores expertos, y no hallaron relación entre los indicadores de cohesión explícitos y la calidad textual.

Además de los MD, los textos se organizan y estructuran en movimientos retóricos diferenciados que sirven para conseguir sus propósitos comunicativos. Estos movimientos retóricos son segmentos del discurso que realizan funciones comunicativas definidas (Upton & Cohen, 2009). Debido a su particular naturaleza, la caracterización de estos movimientos retóricos viene determinada por el género textual. Independientemente de su calidad o de la función que desempeñan en el texto, los movimientos retóricos pueden caracterizarse por su densidad, es decir, el número de palabras o cláusulas que los componen. Se asume que aquellos que contienen mayor número de cláusulas o palabras resultan más informativos y contribuyen en mayor medida a la calidad estimada de un texto.

En resumen, en estudios previos se han identificado indicadores de la calidad textual basados estrictamente en medidas cuantitativas (p. ej. la productividad), e indicadores lingüísticos específicos que operan tanto a nivel local como a nivel textual (p. ej., vocabulario, sintaxis y organización discursiva). El dominio de estos mecanismos lingüísticos resulta en una mejor disposición de los argumentos y evidencias que proporciona el autor y, por tanto, en una consecución más eficaz de la función del texto. Un enfoque dual, que incluya un análisis de los cambios evolutivos tanto a nivel macro como micro, nos permitirá distinguir aspectos cuyo dominio requiere periodos prolongados de desarrollo de aquellos que son sensibles a una exposición pedagógica a corto plazo.

El presente estudio

En este estudio analizamos la composición de los textos analíticos con un doble objetivo: (1) identificar los cambios macroevolutivos (relacionados con la edad y/o el nivel educativo) en una serie de elementos léxicos, sintácticos y discursivos que han demostrado ser indicadores de la

calidad del texto escrito, comparando los textos elaborados por estudiantes de educación primaria, secundaria y universitarios; (2) analizar si estos aspectos experimentan cambios en un periodo de tiempo limitado, es decir, durante la aplicación de una serie de actividades pedagógicas (cambios microevolutivos), comparando dichos aspectos en los textos analíticos que producen los estudiantes antes y después de estas actividades.

El diseño de un entorno pedagógico común nos permite asumir esta doble perspectiva (microevolutiva y macroevolutiva). Dado que no es este un estudio de intervención, nuestro objetivo no es demostrar que estas actividades son más efectivas que otros tratamientos posibles, sino ofrecer un marco común para analizar el desarrollo de los aspectos objeto de estudio. Durante la recolección de los datos, los participantes realizaron la misma secuencia pedagógica en su centro educativo para garantizar que la identificación e interpretación de posibles tendencias evolutivas no se viera afectada o adulterada. Asumimos que el conocimiento de los participantes sobre la función de un género determinado activa los mecanismos lingüísticos más apropiados para llevar a cabo dicha función (Berman, Ragnarsdóttir, & Strömqvist, 2002).

De esta forma, implementamos en el aula una serie de actividades pedagógicas diseñadas con el objeto de despertar o generar conciencia en los estudiantes sobre la función y las características de los textos analíticos. Para determinar su nivel inicial en el uso de los mecanismos lingüísticos examinados, se pidió a los participantes que produjeran un texto antes de realizar las actividades pedagógicas, y otro después de la última sesión. En línea con los estudios revisados, esperamos:

(1) que el nivel educativo tenga un efecto en el comportamiento de los indicadores de productividad y en los aspectos léxicos: los estudiantes de mayor edad producirán textos más extensos, con mayor densidad y diversidad léxica y mayor número de nominalizaciones y de

adjetivos que los más jóvenes. Asimismo, esperamos encontrar un efecto del nivel educativo en el uso de los indicadores sintácticos y discursivos.

(2) que los indicadores examinados presenten valores más altos tras la implementación de las actividades diseñadas para despertar o generar conciencia en los participantes sobre las características de los textos analíticos.

A diferencia de lo esperado para los cambios macroevolutivos, y debido a la escasez de estudios empíricos sobre el efecto del trabajo pedagógico en el uso de indicadores lingüísticos específicos en los textos analíticos, nuestras expectativas sobre los cambios microevolutivos no están directamente motivadas por evidencia empírica y su justificación es esencialmente teórica.

Método

Participantes

En el estudio participaron un total de 212 estudiantes de tres niveles educativos distintos: —6º de primaria, 4º de la ESO y universidad —, bajo un diseño transversal de medidas repetidas. La muestra fue recogida en Castilla-León y Castilla-La Mancha, en dos centros de educación primaria y tres de secundaria, y en una facultad universitaria.

Tras un muestreo de conveniencia, la muestra estaba formada por 65 alumnos de primaria, 78 de secundaria y 69 estudiantes universitarios. En la Tabla 1 se muestran la media y el rango de edad y el género del participante por nivel educativo. No se excluyó a ningún participante por su bagaje sociolingüístico ni por presentar dificultades de aprendizaje.

<INSERT TABLE 1>

Materiales y procedimiento

Durante siete sesiones, los participantes realizaron un conjunto de actividades realizadas con su profesor habitual en sus clases. Los profesores habían recibido formación sobre la implementación de las actividades pedagógicas y se les orientó sobre la ejecución de las tareas respetando el diseño del estudio para garantizar un mismo contexto pedagógico para todos los grupos. Las actividades eran idénticas en los tres niveles educativos, pero las instrucciones verbales de los profesores se habían adaptado ligeramente para ofrecer el apoyo necesario a los participantes más jóvenes.

Estas actividades tenían por objeto incrementar el conocimiento de los participantes sobre las características de los textos analíticos e incluían lecturas sobre el tema tratado, esto es, la libertad de movimiento entre países, discusiones en clase y producción de textos analíticos. Asimismo, algunas sesiones se dedicaron a evaluar, a partir de una rúbrica, fragmentos de textos de otros compañeros y a reflexionar sobre ciertos aspectos clave de dichos textos.² Durante la recogida de datos, cada uno de los participantes produjo un total de cinco textos: Texto 1 sobre el tema A (*libertad de vestir*), producido antes del comienzo de las actividades pedagógicas; Textos 2, 3 y 4 sobre el tema B (*libertad de movimiento*), producidos antes, durante e inmediatamente después de las actividades, respectivamente, y Texto 5 sobre el tema C (*premios y castigos*), producido un mes después de finalizar las actividades. En este estudio se analizan los datos obtenidos a partir de los textos 2 y 4, ambos sobre el mismo tema, producidos antes del comienzo de las actividades e inmediatamente después de la instrucción, respectivamente.

Todos los textos obtenidos fueron transcritos y codificados mediante el formato CHAT (proyecto CHILDES; MacWhinney, 2000), y los datos se analizaron y codificaron utilizando programas CLAN. El corpus estaba formado por 424 textos segmentados en cláusulas, la unidad básica de análisis, siguiendo las indicaciones de Berman y Slobin (1994).

Medidas

Con las medidas obtenidas a partir del análisis de los textos se pretendía capturar aspectos lingüísticos de la calidad textual de los textos producidos por los participantes en cuatro dimensiones distintas: medidas de productividad, léxicas, sintácticas y discursivas (véase la sección previa Medidas de calidad textual en el desarrollo de la escritura analítica).

Medidas de productividad

Número de palabras. Se calculó el número total de palabras como medida de productividad a nivel de palabra. El conteo se realizó automáticamente mediante comandos CLAN en las transcripciones de los textos.

Número de cláusulas. Se obtuvo el número total de cláusulas como medida de la productividad a nivel de frase. El número de cláusulas se obtuvo automáticamente de las transcripciones de los textos mediante comandos CLAN . Dos asistentes de investigación con formación en lingüística segmentaron en cláusulas dos veces un 20% de los textos. Se comprobó la fiabilidad interevaluador mediante el coeficiente de correlación intraclase (CCI, bidireccional mixto de acuerdo absoluto), cuyo valor fue de .998 (95% IC: .997, .999).

Número de movimientos retóricos. Se identificaron manualmente los movimientos argumentativos (expresiones del posicionamiento del autor) y expositivos (expresiones en apoyo de dicho posicionamiento). Se realizaron pruebas de fiabilidad interevaluador en un 5% de los textos teniendo en cuenta el inicio y el cierre de los movimientos retóricos. El índice de acuerdo absoluto fue de .89. Todos los movimientos retóricos fueron codificados en las transcripciones, de modo que su cómputo también se realizó automáticamente mediante comandos disponibles en CLAN.

Medidas de nivel léxico

Longitud de palabra. Se calculó la media de letras de las palabras con contenido semántico, es decir, de las palabras etiquetadas como nombres (comunes), verbos (principales), adjetivos y adverbios. Los adverbios no gramaticalizados (es decir, los adverbios adjetivales) se consideraron palabras de contenido semántico. Para calcular esta medida, se utilizó el analizador morfológico de CLAN junto con una desambiguación manual.

Densidad léxica. Esta medida consistió en la proporción de palabras de contenido sobre el número total de palabras y se obtuvo automáticamente mediante el programa CLAN.

Diversidad léxica. Utilizamos el índice D disponible en CLAN bajo la denominación VOCD (MacWhinney, 2000). Este índice, a diferencia de otras medidas de diversidad léxica, no se ve afectado por la longitud del texto (Malvern et al., 2004).

Uso de adjetivos. Calculamos la proporción de adjetivos en el número total de palabras. Solo se tuvieron en cuenta los adjetivos calificativos.

Use de nominalizaciones. Para obtener esta medida se utilizó el analizador morfológico de CLAN junto con una desambiguación manual. A continuación se calculó la proporción de nominalizaciones sobre el número total de palabras. Tanto los nombres derivados de verbos como los derivados de adjetivos se consideraron nominalizaciones.

Medidas de nivel sintáctico

Densidad sintáctica. A partir del número total de palabras y de cláusulas, se calculó automáticamente la media de palabras por cláusula.

Uso de la subordinación. Se identificaron y codificaron manualmente las cláusulas subordinadas de los textos y se calculó la proporción de estas sobre el número total de cláusulas.

Se tuvieron en cuenta todos los tipos de subordinación, tanto finitas como no finitas. No obstante, en las medidas que siguen se distinguió entre cláusulas relativas y cláusulas adverbiales.

Cláusulas adverbiales. Identificamos todas las cláusulas adverbiales de los textos y calculamos la proporción de estas sobre el número total de cláusulas. En los textos analizados, este tipo de cláusula contiene información semántica relevante, como son las causas y las consecuencias. Estas cláusulas añaden información adicional y complementan la estructura argumentativa de las frases; es por ello que su uso resulta opcional y depende de las preferencias retóricas del autor.

Cláusulas relativas. Se identificaron todas las cláusulas relativas en los textos y se calculó la proporción de estas sobre el número total de cláusulas. Este tipo de cláusula actúa como complemento del nombre en la creación de sintagmas nominales complejos y suele aportar mayor riqueza y precisión a la información que contiene el nombre.

Medidas de nivel discursivo

Marcadores discursivos. Identificamos todos los marcadores discursivos (MD) de los textos según los criterios de Martín Zorraquino & Portolés, (1999) y calculamos la proporción de MD sobre el número total de cláusulas.

Densidad discursiva. Se calculó como la media de palabras por movimiento retórico. Esta medida se empleó como indicador de la cantidad de información contenida dentro de los límites de un fragmento discursivo.

Resultados

En la Tabla 2 se muestran las medias y las desviaciones típicas de todas las medidas por nivel educativo. Se incluyen los resultados de una serie de ANOVA mixtos de dos factores, con el nivel educativo como variable intersujeto (primaria-secundaria-universidad) y el texto como variable intrasujeto (Texto 1-Texto 2, en adelante T1 y T2). En general, los datos analizados exhibieron una distribución normal con valores de asimetría inferiores a 3 y de curtosis inferiores a 10 (valores absolutos; Kline, 2011). No obstante, identificamos algunos valores extremos ($> 2,5 DT$) que correspondían a participantes que habían producido un texto considerablemente más extenso (en palabras, cláusulas o movimientos retóricos) que sus pares del mismo nivel educativo. Además, en diversas variables no se cumplió el supuesto de homocedasticidad. Por ese motivo, se llevaron a cabo los análisis en R utilizando la función *bwtrim* del programa WSR2, robusto frente a desviaciones de la normalidad y a la heterocedasticidad y que facilita la estadística Q para cada uno de los efectos principales y para las interacciones entre estos. Para las interacciones significativas, se calculó el efecto simple del nivel educativo mediante ANOVA de un factor utilizando la función *t1way* de ese mismo programa. Asimismo, se calculó el efecto simple de Texto mediante la función *yuend*, que ofrece una prueba t robusta para muestras relacionadas (Mair & Wilcox, 2016). En estos casos, se realizaron comparaciones múltiples con la corrección de Bonferroni para evitar errores de tipo I.³

<INSERT TABLE 2>

Cambios micro y macroevolutivos en la productividad

Todas las medidas de productividad a nivel de palabra, frase o discurso mostraron influencia del nivel educativo del autor, de modo que los participantes más expertos generaban textos más extensos (Tabla 3). El aumento de la extensión del texto era la respuesta más habitual a la exposición pedagógica, puesto que se observaron efectos significativos de la variable Texto en

todas las medidas de productividad (Tabla 3). No obstante, estos efectos principales estaban moderados por una interacción significativa: $Q(2, 110.42) = 25.42, p < .001$, para el número de palabras; $Q(2, 109.97) = 20.88, p < .001$, para el número de cláusulas y $Q(2, 109.90) = 8.85, p < .001, \eta^2_p = .063$, para el número de movimientos retóricos. Un seguimiento de estas interacciones reveló que el efecto del nivel educativo era significativo tanto para el T1 como para el T2 ($p < .001$). Sin embargo, las pruebas t -test del efecto simple de Texto revelaron que, mientras los participantes más jóvenes incrementaban la extensión del texto generado en T2 en número de palabras o cláusulas ($ps < .001$), los estudiantes universitarios utilizaban un número significativamente menor ($ps \leq .017$). Por lo que respecta al número de movimientos retóricos los alumnos de primaria y los estudiantes universitarios no variaron significativamente el número de movimientos retóricos entre T1 y T2 ($ps > .05$), a diferencia de los estudiantes de secundaria ($p < .001$).

<INSERT TABLE 3>

<INSERT TABLE 4>

<INSERT FIGURE 1>

Cambios macro y micro evolutivos en las medidas léxicas

La mayor parte de las elecciones léxicas de los participantes (longitud de palabra, diversidad léxica, proporción de adjetivos y de nominalizaciones) mostraron cambios macroevolutivos significativos, puesto que se observó un efecto principal significativo del nivel educativo, excepto en el caso de la densidad léxica (Tabla 3). No obstante, solo la densidad y la diversidad léxicas cambiaban tras el trabajo pedagógico realizado con los participantes. Aunque la densidad léxica disminuyó en T2, la diversidad léxica aumentó (Tabla 4). Se observó una interacción

significativa Nivel educativo x Texto en la longitud de palabra, $Q(2, 104.02) = 5.69, p = 0.005$, y en la proporción de nominalizaciones, $Q(2, 100.51) = 5.29, p < .001$. Los ANOVA de seguimiento de la interacción para determinar el efecto simple del nivel educativo revelaron que el efecto era significativo para T1 y T2 en todas las medidas. Las pruebas t del efecto simple de Texto revelaron efectos distintos del trabajo pedagógico en función del nivel educativo de los participantes. Por lo que respecta a la longitud media de las palabras, los alumnos de primaria incrementaron significativamente la longitud media ($p = .006$), mientras que ni los estudiantes de secundaria ni los universitarios mostraron ningún cambio entre T1 y T2 ($ps > .05$). Por último, no se observaron diferencias significativas en la proporción de nominalizaciones entre T1 y T2 en ningún nivel educativo. Sin embargo, si observamos las medias por grupo y texto, vemos que los más jóvenes incluyen, de media, un mayor número de nominalizaciones en T2, mientras que los universitarios parecen invertir esa tenencia.

<INSERT FIGURE 2>

Cambios macro y microevolutivos en las medidas sintácticas

La mayoría de las medidas de sintaxis mostraron un aumento de complejidad ligado al nivel educativo (Tabla 3). No obstante, solo se observaron cambios microevolutivos en la proporción de cláusulas relativas (Tabla 4). Asimismo, la densidad sintáctica, es decir, la media de palabras por cláusula se vio afectada por una interacción significativa Nivel educativo x Texto, $Q(2, 109.17) = 4.86, p < .001$. Un seguimiento de la interacción reveló que el efecto del nivel educativo era significativo tanto en T1 como en T2 al comparar los participantes de primaria con los de mayor edad ($ps < .001$), mientras que los estudiantes de secundaria y los universitarios no se diferenciaron entre ellos ni en T1 ni en T2 ($ps > .016$). Por lo que respecta al efecto simple de Texto, solo los alumnos de primaria incluyeron un número significativamente mayor de palabras

por cláusula en T2 que en T1 ($p = .003$), mientras que los estudiantes de secundaria y los universitarios incluyeron una proporción de palabras por cláusula similar en ambos textos ($p > .05$).

<INSERT FIGURE 3>

Cambios macro y microevolutivos en las medidas discursivas

Ambas medidas del discurso, tanto la proporción de marcadores discursivos como la densidad discursiva (número de cláusulas por movimiento retórico), experimentaron un aumento general con el nivel educativo (Tabla 3). Sin embargo, la naturaleza de los cambios vinculados a la exposición al trabajo pedagógico resultó contraria a nuestras expectativas iniciales. En primer lugar, no se observó un efecto de Texto en la proporción de marcadores discursivos (Tabla 4), y esta medida tampoco se vio afectada por una interacción significativa $Q(2, 104.31) = 2.36, p = .099$. En segundo lugar, no se observó un efecto de Texto en la densidad de movimientos retóricos (Tabla 4). No obstante, esta medida sí fue moderada por el efecto de una interacción significativa, $Q(2, 104.91) = 4.70, p = .011$ (Figura 3). El ANOVA de un factor para determinar el efecto simple del nivel educativo indicó que el efecto era significativo tanto para T1 como para T2. En T1, los análisis post-hoc indicaron que los alumnos de primaria producían menos marcadores discursivos que los universitarios ($p < .001$). En T2 los alumnos de primaria y secundaria producían menos marcadores discursivos que los estudiantes universitarios ($ps < .016$). Además, las pruebas t del efecto simple del texto revelaron que solo los alumnos de primaria mostraban niveles más elevados de densidad discursiva en T2 que en T1 ($p = .015$), mientras que esta permanecía estable entre los estudiantes de secundaria ($p = .545$) y los universitarios ($p = .358$).

Discusión

Se analizaron una serie de aspectos léxicos, morfosintácticos y discursivos en los textos escritos de estudiantes de educación primaria, secundaria y universitaria, controlando el trabajo pedagógico al que los estudiantes estaban expuestos. Esto nos permitió identificar, por un lado, los aspectos textuales que se desarrollan con el nivel educativo, es decir, los cambios macroevolutivos en los indicadores de calidad de los textos analíticos. Por otro lado, nos permitió identificar los aspectos textuales que se desarrollan en el contexto de las actividades pedagógicas, es decir, los aspectos que experimentan cambios microevolutivos.

Nuestro estudio arroja dos resultados principales. Primero, la productividad, es decir, la cantidad de texto producida por los estudiantes, un aspecto que abarca todos los dominios del lenguaje, confirma su adecuación como medida del desarrollo de la calidad de la escritura. Segundo, no todos los aspectos del texto pertenecientes a un mismo dominio del lenguaje siguieron los mismos patrones evolutivos o mostraron una sensibilidad similar al trabajo pedagógico. En las próximas líneas comentamos los resultados y discutimos sus implicaciones en la evaluación de la calidad de la escritura y en las intervenciones pedagógicas.

Cambios macro y microevolutivos de la productividad textual

En línea con otros resultados sobre el desarrollo de la escritura en distintos géneros textuales, se observaron cambios macro y micro evolutivos en relación con la productividad (Berman & Nir-Sagiv, 2009; Crossley & McNamara, 2010; Salas et al., 2016), de modo que esta bien podría ser un buen indicador de la calidad textual global, puesto que es sensible a los cambios en varios dominios. Este resultado recurrente sugiere que la educación (y la experiencia con la lengua escrita) ayuda a los escritores a ampliar temas y argumentos, lo que facilita la producción de textos más largos (Crossley et al., 2011). Si otros factores (p. ej., edad, conocimientos sobre el contenido) están igualados iguales, el aumento de la extensión del texto podría implicar una

argumentación y un razonamiento más preciso. Del mismo modo, un incremento de la complejidad sintáctica requeriría también un mayor número de palabras y cláusulas (p. ej., cláusulas relativas; Scott, 2009).

Las actividades pedagógicas realizadas, que consistían en leer y discutir textos sobre un tema, influyeron en la productividad escrita pero principalmente durante los años de aprendizaje de la lengua escrita. Los estudiantes universitarios podrían necesitar un trabajo más intensivo o distinto tipo de actividades para alargar sus textos. Otra posible explicación es que los universitarios podrían efectuar cambios en los textos que son esencialmente distintos de los efectuados por los más jóvenes. Por ejemplo, mientras que los más jóvenes podrían incrementar la información que incluyen simplemente añadiendo más palabras o cláusulas, los estudiantes de mayor edad podrían conseguir que sus textos sean más informativos sin hacerlos necesariamente más largos. Dicho de otro modo, las elecciones retóricas de estos últimos podrían implicar una reorganización más sofisticada de sus recursos lingüísticos en lugar de la producción de una cantidad mayor de texto. Esta interpretación, que los estudiantes de nivel universitario se concentran en sus elecciones retóricas durante la composición de un texto, es coherente con los modelos de escritura competente (Alamargot & Chanquoy, 2001; Hayes, 2011). Futuras investigaciones deberían dedicarse a definir en mayor detalle las maneras concretas en que la productividad del texto es indicativa de su calidad.

Diferencias de sensibilidad en un mismo dominio lingüístico

El segundo resultado clave del presente estudio es que, exceptuando la productividad, diferentes medidas correspondientes a un mismo dominio (p.ej., léxico, sintaxis, discurso) mostraban distintas sensibilidades, tanto en el plano macroevolutivo como en el microevolutivo.

En el ámbito léxico, los resultados mostraron que los participantes más expertos utilizaban un vocabulario más sofisticado que a su vez era más diverso. Además, también tendían a utilizar palabras morfológicamente más complejas, tales como las nominalizaciones, y más elementos léxicos que contribuyen a incrementar la precisión de los enunciados e introducen matices en el discurso, como los adjetivos, lo que corrobora los resultados de investigaciones previas sobre el desarrollo de las decisiones léxicas en la escritura (Berman & Nir-Sagiv, 2009; Crossley et al., 2011; McNamara et al., 2010; Nippold, 2007). La única excepción fue la densidad léxica que, en la misma línea que Llauredó y Tolchinsky (2015), pero contrariamente a Cutillas et al. (2014), no reveló cambios evolutivos. La densidad léxica se ha utilizado como una medida del desarrollo léxico en distintas lenguas (principalmente en inglés, pero también en francés, hebreo, catalán y sueco) con resultados irregulares, lo que podría indicar tendencias específicas de cada lengua (Johansson, 2008). De hecho, se trata de una medida basada en el uso de palabras funcionales. Por ejemplo, el uso vacilante de las palabras funcionales en los niños (en algunas lenguas) podría hacer que sus textos parezcan más densos que los producidos por algunos autores de mayor edad y más competentes.

Algunas medidas léxicas fueron más sensibles al trabajo pedagógico realizado y revelaron cambios microevolutivos. Concretamente, los participantes parecían realizar distintas elecciones léxicas: la diversidad léxica incrementaba de T1 a T2 en todos los niveles educativos. Los participantes también utilizaron ítems léxicos más sofisticados, como evidencia el incremento en la longitud media de palabras de contenido utilizadas en el texto producido después de su exposición al trabajo pedagógico. No obstante, solo los alumnos de primaria mostraron este efecto. Podríamos especular que una base de conocimientos menor en términos del vocabulario por parte de los participantes jóvenes podría significar que también tienen la

posibilidad de obtener mayores beneficios en cuanto al aprendizaje de vocabulario nuevo o más complejo que los participantes de mayor edad.

Las distintas medidas sintácticas también mostraron distintos comportamientos. En general, de acuerdo con investigaciones previas sobre el tema (p. ej., Beers & Nagy, 2009), los escritores más experimentados tienden a ofrecer más información por cláusula y producen más cláusulas subordinadas, concretamente más cláusulas relativas. Las cláusulas relativas, al igual que los adjetivos, añaden complejidad a los sintagmas nominales, y al mismo tiempo facilitan la identificación y caracterización de las entidades en el discurso. Este aspecto de la estructura lingüística y del uso del lenguaje va acompañado de un incremento general en la complejidad clausal y se considera un criterio de referencia para medir el desarrollo sintáctico y discursivo (Ravid & Berman, 2010). Contra lo que esperábamos, el uso de cláusulas adverbiales disminuyó con el nivel educativo. Las cláusulas adverbiales son un recurso privilegiado para los escritores a la hora de presentar las razones y explicaciones que apoyan su argumentación. Sin embargo, se ha demostrado que esta categoría abarca un amplio abanico de formas y funciones distintas. Por ejemplo, en el estudio evolutivo realizado por Aparici (2010), solo se observaron efectos de la edad al diferenciar entre distintas subcategorías de las cláusulas adverbiales que transmiten información relevante en un texto argumentativo, como las cláusulas causales o consecutivas, de otras que sitúan los hechos descritos, por ejemplo, en el espacio o el tiempo.

Únicamente dos medidas sintácticas resultaron ser sensibles al trabajo pedagógico realizado. En primer lugar, los participantes utilizaron más cláusulas relativas en T2 en todos los grupos; es decir, después de varias clases dedicadas a incrementar sus conocimientos sobre las características específicas de los textos analíticos. En segundo lugar, los alumnos de primaria, pero no los de secundaria ni los universitarios, utilizaron de media más palabras por cláusula en

T2 que en T1. Hemos mencionado anteriormente que las cláusulas más complejas deberían contener más información, lo que podría constituir uno de los efectos más inmediatos del trabajo pedagógico realizado con los participantes más jóvenes. Por el contrario, los estudiantes de secundaria y universitarios podrían recurrir en mayor medida a otros recursos para incrementar la carga informativa de las cláusulas sin incrementar necesariamente el número de palabras contenidas en estas.

Los dos aspectos seleccionados para medir la calidad del texto a nivel discursivo muestran tendencias evolutivas similares, pero distinta sensibilidad al trabajo pedagógico. Tanto la densidad discursiva como el uso de MD incrementaron con el nivel educativo, lo que confirma que la capacidad de conectar hábilmente distintas partes del discurso se desarrolla con la edad (p.ej., Katzenberger & Cahana-Amitay, 2002). En términos de tendencias microevolutivas, y en contra de lo esperado, no se observaron cambios en el uso de MD entre T1 y T2. Este resultado podría explicarse, por un lado, porque no se tuvieron en cuenta los marcadores no convencionales que operan a nivel de frase (Cuenca, 2013) y, por otro lado, porque no se identificaron los usos no canónicos (semánticamente inadecuados) de dichos marcadores. Además, el uso de elementos explícitos como los MD es solo uno de los diversos mecanismos utilizados para incrementar la coherencia de un texto. En este sentido, coincidimos con trabajos anteriores que proponen que la coherencia de un texto puede no residir en la presencia de elementos de cohesión sino más bien en su ausencia (Crossley et al. 2011). Tanto los resultados relacionados con la densidad sintáctica y discursiva (cambios evolutivos evidentes y un efecto limitado del trabajo pedagógico) como con el uso de MD requieren un análisis más detallado de los distintos mecanismos que los escritores reclutan y utilizan para “expresar ideas y condensar contenidos semánticos” (Slobin, 1977, p.187, traducción libre). En este sentido, futuras

investigaciones deberían considerar, por ejemplo, el uso de medidas cualitativas relacionadas con la calidad de la estructura argumentativa y la adopción de una posición discursiva (Uccelli et al., 2019; Vilar et al., en revisión), el establecimiento de relaciones semánticas ilícitas o uso no canónico de mecanismos léxicos (p. ej., conectores intra e interoracionales)

Por último, por lo que respecta a los efectos macro y microevolutivos sobre la densidad del discurso, únicamente en el caso de los alumnos de primaria se observó un incremento en la densidad de movimientos retóricos entre T1 y T2. Como ya se ha indicado a nivel de cláusula, también en el nivel discursivo parece que los participantes recurrieron a otros recursos además de incrementar el número de unidades de la misma categoría (p. ej., un número mayor de palabras en las cláusulas o de cláusulas en los movimientos retóricos) para conseguir la máxima expresividad.

Para comprender mejor el proceso de desarrollo y dominio de un género discursivo (p. ej., textos analíticos) es crucial identificar qué aspectos textuales cambian con la edad y/o nivel educativo y cuáles cambian en un periodo de tiempo más reducido, en el contexto de un trabajo pedagógico explícito y supervisado. El uso de una secuencia de actividades idéntica nos permitió identificar cambios evolutivos en ciertos aspectos de la calidad textual. No obstante, la diversidad de comportamiento de los aspectos textuales que caracterizan a los distintos dominios lingüísticos sugiere que tanto la evaluación de la calidad textual como el trabajo pedagógico en este ámbito deben ser abordados a un alto nivel de granularidad (es decir, tal vez no se identificaron cambios en el dominio léxico en conjunto, pero sí se observaron en algunos mecanismos específicos de este dominio lingüístico). Detectar las funciones específicas que son más sensibles al trabajo pedagógico podría contribuir a un mejor conocimiento de los procesos que subyacen al desarrollo de la escritura analítica en los distintos niveles educativo. Pero

además, esta información contribuiría significativamente al diseño de futuros estudios de intervención y debería ser de utilidad para orientar la evaluación de los textos escritos.

Limitaciones del estudio

Como sucede con los estudios de este tipo, este trabajo presenta una serie de limitaciones. En primer lugar, el desarrollo de la escritura se caracteriza por una gran variación interindividual. A pesar de que nuestra muestra no era especialmente pequeña, una muestra de mayor tamaño probablemente nos hubiese permitido obtener unos resultados más sólidos e incrementar la potencia estadística en general, de modo que algunas tendencias probables pudiesen haber sido corroboradas o descartadas de manera más efectiva. En segundo lugar, aunque nuestro diseño experimental no pretendía ser una intervención, sería interesante comparar los resultados frente a un grupo de control que recibiese una formación más tradicional. Futuras investigaciones deberían tratar de reproducir el contexto educativo diseñado en este estudio y compararlo con una condición sin intervención pedagógica. Por último, una de las fortalezas del estudio, su enfoque centrado en un género del discurso, compromete la generalización de los resultados hacia otros géneros (p. ej., narrativo o descriptivo), que en este sentido no está garantizada.

Notas

¹ Se utiliza el término género para denominar un tipo de texto —oral o escrito— categorizado en función de diversos criterios externos tales como su audiencia, finalidad y tipo de actividad; es decir, hace referencia a un agrupamiento de textos convencional y culturalmente reconocido, basado en propiedades distintas de las léxicas o gramaticales.

² Para una exposición más detallada sobre las actividades pedagógicas, véase Vilar et al. (en revisión).

³ El umbral de significatividad, tras la corrección de Bonferroni, era $p < .016$.

Tabla 1

Media de edad, rango de edades y género de los participantes por nivel educativo

Nivel educativo	Media de edad	Rango de edad	Participantes		
			Hombre	Mujer	total
Ed. primaria	11.6	11.1 – 12.1	30	35	65
Ed. secundaria	15.8	15.0 – 18.3	33	45	78
Ed. universitaria	21	19.10 – 31.8	27	42	69

Tabla 2

Estadísticos descriptivos para todas las medidas en T1 y T2

Nivel Educ.	Medidas textuales	media (DT)	Mín.-Máx.	Media T1 (DT)	Media T2 (DT)
PRI	N.º palabras	207.38 (84.36)	84 – 388	85.89 (43.58)	121.49 (56.96)
SEC		467.09 (131.76)	179 – 936	201.36 (66.58)	265.73 (76.69)
UNI		702.18 (225.57)	312 – 1911	370.75 (148.10)	331.42 (95.10)
PRI	N.º cláusulas	38.11 (14.91)	14 – 74	16.58 (7.90)	21.52 (9.86)
SEC		74.37 (21.08)	18 – 138	31.88 (11.10)	42.49 (13.43)
UNI		105.49 (39.99)	45 – 312	55.81 (25.44)	49.68 (16.95)
PRI	N.º mov. retóricos	9.57 (3.93)	4 – 18	4.62 (2.44)	4.95 (2.26)
SEC		13.97 (3.04)	8 – 22	5.94 (2.32)	8.04 (2.17)
UNI		16.58 (4.94)	10 – 40	8.41 (4.26)	8.17 (2.95)
PRI	Longitud de palabra	4.32 (2.64)	3.80 – 5.09	4.38 (0.35)	4.24 (0.29)
SEC		4.58 (2.30)	4.10 – 5.12	4.56 (0.27)	4.59 (0.24)
UNI		4.71 (1.61)	4.38 – 5.24	4.69 (0.19)	4.73 (0.18)
PRI	Densidad léxica	39.85 (0.39)	31.01 – 56.34	40.79 (4.34)	38.90 (4.75)
SEC		39.83 (0.28)	34.43 – 45.86	40.11 (3.16)	39.56 (3.40)
UNI		40.40 (0.23)	36.00 – 47.10	40.64 (2.76)	40.17 (0.27)
PRI	Diversidad léxica	66.28 (25.27)	0 – 111.93	57.78 (33.16)	74.77 (28.37)
SEC		86.99 (15.31)	52.19 – 130.49	84.08 (17.55)	89.90 (18.67)
UNI		87.27 (12.11)	67.46 – 128.64	87.93 (16.28)	86.59 (12.93)
PRI	% adjetivos	4.18 (2.00)	0.45 – 10.26	3.69 (2.27)	3.57 (2.22)
SEC		5.13 (1.89)	0.86 – 9.24	5.19 (2.31)	5.60 (1.91)
UNI		5.96 (1.42)	3.14 – 9.48	6.25 (1.79)	6.08 (1.58)
PRI	% nominalizaciones	3.62 (1.73)	0 – 7.00	2.81 (2.30)	2.86 (1.97)
SEC		4.51 (2.05)	0 – 9.09	5.02 (2.05)	5.27 (1.94)
UNI		5.98 (1.33)	2.89 – 9.04	6.25 (1.79)	5.76 (1.45)
PRI	Densidad sintáctica	5.75 (.99)	3.75 – 9.94	5.20 (0.91)	5.71 (1.00)
SEC		6.18 (.88)	4.60 – 8.81	6.45 (1.02)	6.38 (0.94)
UNI		6.72 (.79)	5.04 – 9.19	6.89 (1.03)	6.82 (0.91)
PRI	% cláusulas subordin.	62.09 (9.99)	22.22 – 81.25	61.63 (12.14)	60.42 (13.45)
SEC		60.14 (8.73)	26.76 – 74.23	63.13 (11.23)	60.69 (10.18)
UNI		62.71 (6.67)	51.04 – 80.46	60.26 (12.17)	62.80 (7.97)
PRI	% cláusulas relativas	13.59 (6.02)	0 – 23.53	10.08 (7.61)	13.97 (8.61)
SEC		17.02 (6.44)	2.78 – 32.47	15.74 (8.33)	17.57 (7.07)

UNI		18.58 (5.84)	6.41 – 34.21	17.51 (6.16)	23.25 (7.74)
PRI		26.40 (9.01)	5.56 – 52.38	28.08 (13.59)	26.20 (11.31)
SEC	% cláusulas adverb.	26.34 (7.45)	7.04 – 43.40	27.92 (11.78)	26.75 (8.28)
UNI		18.58 (5.84)	6.41 – 34.21	23.90 (7.70)	23.16 (7.23)
PRI		1.49 (.86)	0 – 3.45	1.02 (1.03)	1.16 (1.00)
SEC	% marc. discursivos	1.52 (.75)	0 – 3.61	1.88 (1.01)	1.98 (0.88)
UNI		1.75 (.66)	0.39 – 3.75	1.76 (0.73)	1.53 (0.80)
PRI		5.34 (2.06)	2.00 – 12.80	5.05 (3.35)	6.37 (3.57)
SEC	Densidad discursiva	5.17 (2.20)	1.67 – 12.54	5.62 (2.58)	5.08 (2.46)
UNI		6.24 (3.01)	2.38 – 20.80	7.18 (4.23)	6.48 (3.12)

Nota. PRI – educación primaria, SEC – educación secundaria, UNI – educación universitaria (adultos)

Tabla 3

Resultados del ANOVA mixto de dos factores para el efecto principal de Nivel educativo

Medida textual	Q(2)	p	Pruebas post-hoc (Bonferroni)		
			PRI-SEC	PRI-UNI	SEC-UNI
No. palabras	229.94	< .001	$p < .001$	$p < .001$	$p < .001$
No. cláusulas	142.07	< .001	$p < .001$	$p < .001$	$p < .001$
No. mov. retóricos	49.01	< .011	$p < .001$	$p < .001$	$p < .001$
Longitud de pal.	59.48	< .001	$p < .001$	$p < .001$	$p < .001$
Densidad léxica	1.08	.342		n/a	
Diversidad léxica	16.63	< .001	$p < .001$	$p < .001$	$p = 1.00$
% adjetivos	43.50	< .001	$p < .001$	$p < .001$	$p < .001$
% nominalizaciones	88.23	< .001	$p < .001$	$p < .001$	$p < .001$
Densidad sintáctica	53.94	< .001	$p < .001$	$p < .001$	$p = .006$
% cláus. subordinadas	0.29	.746		n/a	
% cláus. relativas	33.53	< .001	$p < .001$	$p < .001$	$p < .001$
% cláus. adverbiales	5.94	.004	$p = 1.00$	$p = .017$	$p = .007$
% marc. discursivos	24.37	< .001	$p < .001$	$p < .001$	$p = .046$
Densidad discursiva	9.89	< .001	$p = .975$	$p < .001$	$p < .001$

Nota. PRI – educación primaria, SEC – educación secundaria, UNI – educación universitaria (adultos)

Tabla 4

Resultados del ANOVA mixto de dos factores para el efecto principal de Texto

Medida textual	Q(1)	<i>p</i>	Dirección del efecto
No. palabras	17.52	< .001	T1 < T2
No. cláusulas	12.77	< .001	T1 < T2
No. mov. retóricos	12.18	< .001	T1 < T2
Longitud de palabra	1.24	.267	n/a
Densidad léxica	10.25	.002	T1 > T2
Diversidad léxica	10.41	.002	T1 < T2
% adjetivos	< 0.01	.966	n/a
% nominalizaciones	0.02	.877	n/a
Densidad sintáctica	1.64	.203	n/a
% cláusulas subord.	1.62	.205	n/a
% cláusulas relativas	26.03	< .001	T1 < T2
% cláusulas adverb.	0.74	.392	n/a
% marc. discursivos	< 0.01	.992	n/a
Densidad discursiva	0.633	.428	n/a

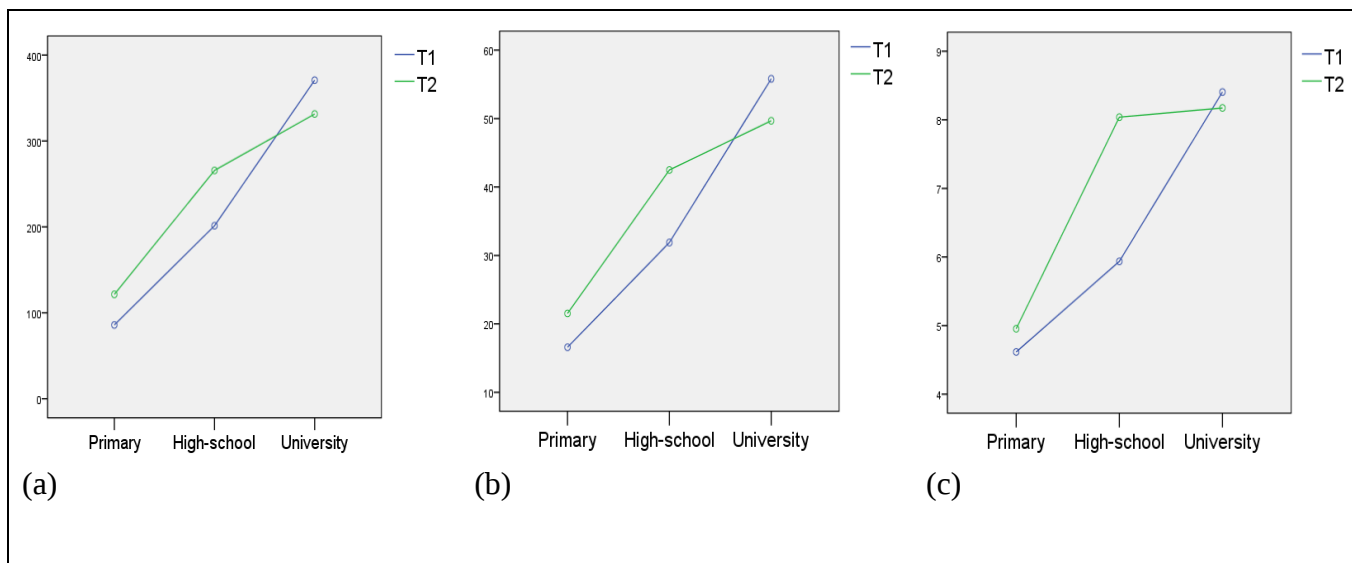


Figura 1. Representación gráfica de las interacciones significativas Nivel educativo x Texto para las medidas de productividad: (a) número de palabras; (b) número de cláusulas; (c) número de movimientos retóricos.

[eje horizontal: **Primaria** / **Secundaria** / **Universidad**]

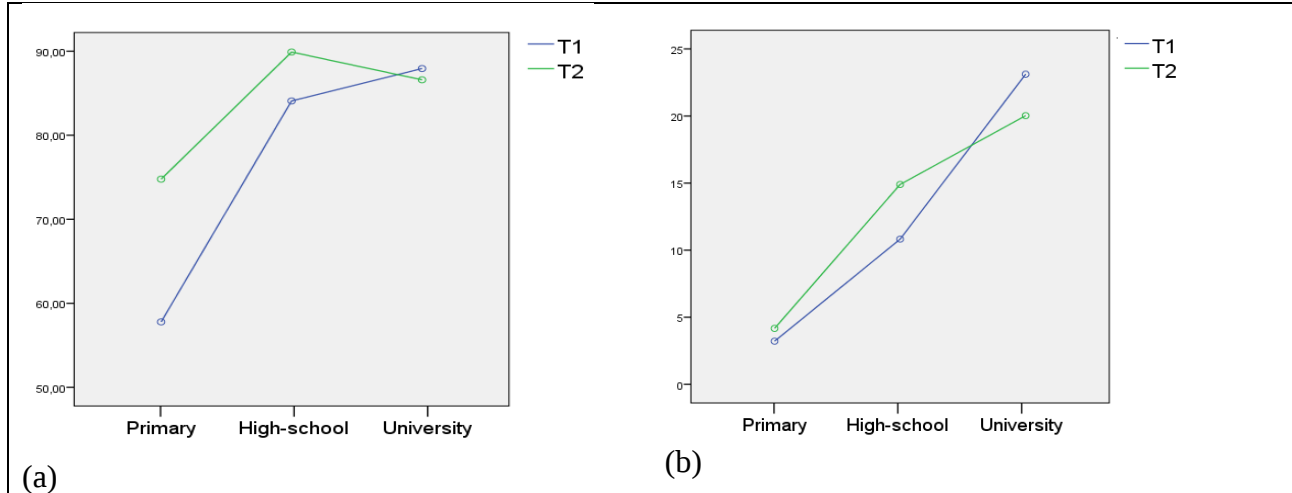


Figura 2. Representación gráfica de las interacciones significativas Nivel educativo x Texto para las medidas léxicas: (a) diversidad léxica; (b) longitud media de las palabras de contenido.

[eje horizontal: Primaria / Secundaria / Universidad]

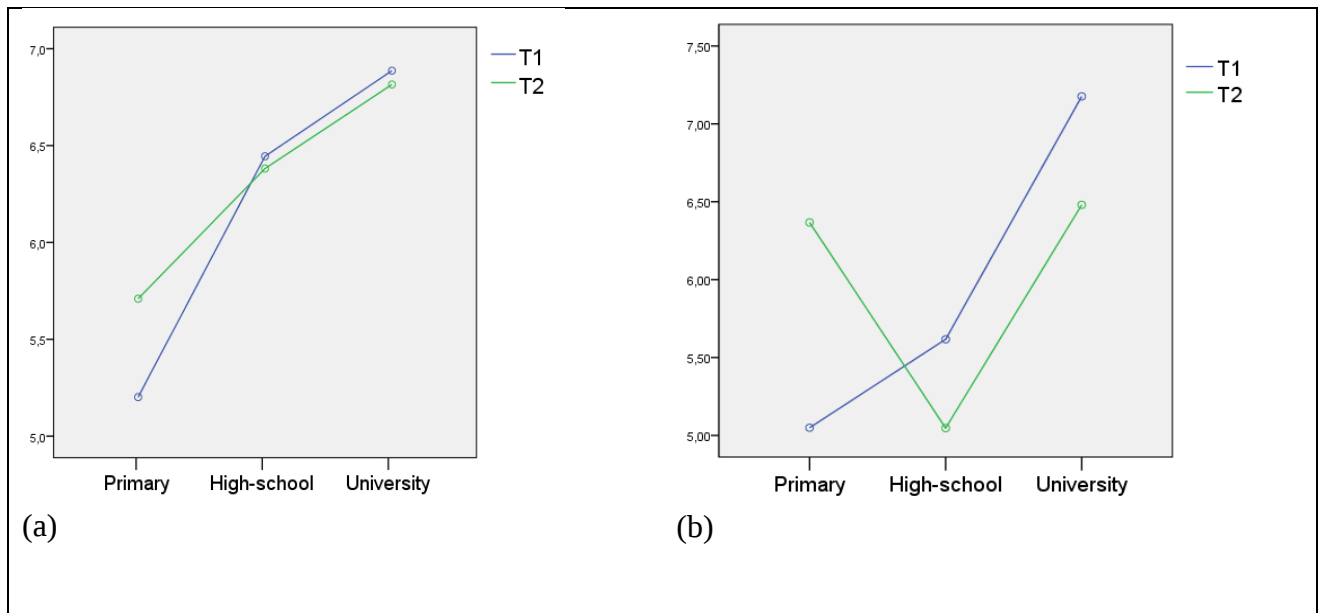


Figura 3. Representación gráfica de las interacciones significativas Nivel educativo x Texto para las medidas discursivas: (a) densidad sintáctica; (b) densidad discursiva.

[eje horizontal: Primaria / Secundaria / Universidad]

Acknowledgements / Agradecimientos

This work was supported by the Ministerio de Economía y Competitividad under Grant EDU2015-65980-R ('Escritura analítica y diversidad lingüística: cambios evolutivos y microevolutivos desde la educación primaria a la educación superior'). Rocío Cuberos was supported by a grant from the University of Barcelona (APIF). We would like to thank Liliana Tolchinsky and Hugo Vilar their valuable observations. / *Este trabajo ha recibido el apoyo del Ministerio de Economía y Competitividad a través del programa EDU2015-65980-R ('Escritura analítica y diversidad lingüística: cambios evolutivos y microevolutivos desde la educación primaria a la educación superior'). Rocío Cuberos recibió una beca de la Universidad de Barcelona (APIF). Queremos expresar nuestro agradecimiento a Liliana Tolchinsky y Hugo Vilar por sus valiosas observaciones.*

References / Referencias

- Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing: With commentaries by Ronald T. Kellogg and John R. Hayes*. Dordrecht: Kluwer Academic Publishers.
- Anglin, J. (1993). Vocabulary development: A morphological analysis. *Monographs of the society for research in child development*, 58(10), 1-165.
- Aparici, M. (2010). El desarrollo de la conectividad discursiva en diferentes géneros y modalidades de producción. Unpublished doctoral dissertation. Barcelona: University of Barcelona.
- Baratta, A. (2010). Nominalization development across an undergraduate academic degree program. *Journal of Pragmatics*, 42(4), 1017-1036.
- Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.
- Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing: An Interdisciplinary Journal*, 22(2), 185-200.

- Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale: Lawrence Erlbaum.
- Berman, R. A., & Nir-Sagiv, B. (2009). Cognitive and linguistic factors in evaluating expository text quality: Global versus local? In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 421-440). Amsterdam: John Benjamins.
- Berman, R. A., & Nir-Sagiv, B. (2010). The lexicon in writing-speech differentiation. Developmental perspectives. *Written Language & Literacy*, 13(2), 183-205.
- Berman, R. A., Ragnarsdóttir, H., & Strömqvist, S. (2002). Discourse stance. *Written Language & Literacy*, 5(2), 253-287.
- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. New York: Cambridge University Press.
- Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Charles, M. (2003). A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, 2(4), 313-26.
- Choi, I. (2007). How and when do children acquire the use of Discourse Markers? *CamLing 2007*, 40-47.
- Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 984-989). Austin, TX: Cognitive Science Society.

- 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.
- Crowhurst, M. (1983). Syntactic complexity and writing quality: a review. *Canadian Journal of Education*, 8(1), 1-16.
- Cuenca, M. J. (2013). The fuzzy boundaries between discourse marking and modal marking. In L. Degand, B. Cornillie, & P. Pietrandrea (Eds.), *Discourse markers and modal particles. Categorization and description* (pp. 191–216). Amsterdam: John Benjamins.
- Cutillas, L., Tolchinsky, L., Rosado, E., & Perera, J. (2014). Indicators of lexical growth throughout age, genre and modality for a Catalan L1 corpus. In A. Díaz Negrillo & F. Díaz Pérez (Eds.), *Specialisation and variation in language corpora* (pp. 159-186). Bern: Peter Lang.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287-312.
- Flowerdew, J. & Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific Purposes*, 29, 78-93.
- Gregori, C., & Clavel, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia - Social and Behavioral Sciences*, 198, 546-556.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman Publishing.
- Hayes, J.R. (2011). Kinds of knowledge-telling: Modeling early writing development. *Journal of Writing Research*, 3(2), 74-92.
- Hyland, K. (2005). *Metadiscourse: exploring interaction in writing*. London: Continuum.

- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers Lund University*, 53, 61-79.
- Katzenberger, I., & Cahana-Amitay, D. (2002). Segmentation marking in text production. *Linguistics*, 40(6), 1161-1184.
- Kim, Y. S., Al Otaiba, S., Puranik, C., Sidler, J. F., Greulich, L., & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study at the end of kindergarten. *Learning and Individual Differences*, 21(5), 517–525.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. The Guildford Press.
- Langacker, R. (1999). *Grammar and conceptualization*. Berlin & New York: Mouton de Gruyter.
- Llauradó, A., & Tolchinsky, L. (2013). Growth of text-embedded lexicon in Catalan: From childhood to adolescence. *First Language*, 33(6), 628-653.
- Mair, P., & Wilcox, R. (2016). Robust statistical methods in R using the WSR2 package. *Behavior Research Methods*, 1-25.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. New York: Palgrave MacMillan.
- Martín Zorraquino, M. A., & Portolés, J. (1999). Los marcadores del discurso. In I. Bosque & V. Demonte (Eds.), *Gramática Descriptiva de la Lengua Española* (Vol. 3, pp. 4051-4213). Madrid: Espasa Calpe.
- McCarthy, P. M., & Jarvis, S. (2010). MTLT, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42(2), 381-92.

- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86.
- Nippold, M. (2007). Peer conflict explanations in children, adolescents, and adults: examining the development of complex syntax. *American Journal of Speech-Language Pathology*, 16(2), 179-88.
- Penris, W., & Verspoor, M. (2017). Academic writing development: a complex, dynamic process. In S. Pfenniger & J. Navracsics (Eds.), *Future research directions for Applied Linguistics* (pp. 215-242). Bristol: Multilingual Matters.
- Ravid, D., & Levie, R. (2010). Hebrew adjectives in later language text production. *First Language*, 30(1), 27-55.
- Ravid, D., Rosado, E., Van Hell, A., & Zamora, A. (2002). Subject NP patterning in the development of text production: Speech and writing. *Written Language and Literacy*, 5(1), 69-93.
- Riedemann, H. (1996). Word length distribution in English press texts. *Journal of Quantitative Linguistics*, 3(3), 265-271.
- Salas, N., Llauradó, A., Castillo, C., Taulé, M., & Martí, M. A. (2016). Linguistic correlates of text quality from childhood to adulthood. In J. Perera, M. Aparici, E. Rosado & N. Salas (Eds.), *Written and Spoken Language Development across the Lifespan. Essays in Honour of Liliana Tolchinsky* (pp. 307-326). Cham: Springer.
- Sanders, T., & Noordman, L. G. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes* 29(1), 37-60.
- Schleppegrell, M. J. (2004). *The language of schooling. A Functional Linguistics Perspective*. Mahwah, New Jersey: Lawrence Erlbaum.

- Scott, C.M. (2009). A case for the sentence in reading comprehension. *Language, Speech, and Hearing Services in Schools, 40*, 184-191.
- Slobin, D. I. (1977). Language change in childhood and in history. In J. McNamara, *Language Learning and Thought* (pp. 185-214). New York: Academic Press.
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112-133). New York: Cambridge University Press.
- Tolchinsky, L., Aparici, M., & Rosado, E. (2017). Escribir para pensar y persuadir. *Textos de didáctica de la lengua y la literatura, 76*, 14-21.
- Toulmin, S. (2002). *The uses of argument*. New York: Cambridge University Press.
- Uccelli, P., Deng, Z., Phillips Galloway, E., & Qin, W. (2019). The role of language skills in mid-adolescents' science summaries. *Journal of Literacy Research, 51*(3), 357-380.
- Uccelli, P., Dobbs C., & Scott, J. (2012). Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication, 30*(1), 36-62.
- Upton, T., & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies, 11*(5), 585-605.
- Vilar, H., Tolchinsky, L., & Aparici, M. (submitted). Developmental and microdevelopmental changes in bilinguals analytical text writing from primary to higher education.
- Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Tschinkel, E., & Kantor, P. T. (2011). Modelling the development of written language. *Reading and Writing: An Interdisciplinary Journal, 24*(2), 203-220.
- Zipf, G. K. (1932). *Selected studies of the principle of relative frequency in language*. Cambridge, MA: Harvard University Press.

