

COMPLEXITY GROWTH IN GRADED READERS: A CROSS-LINGUISTIC CORPUS APPROACH

NÉSTOR DE ARMAS GUERRA

Universidad de Las Palmas de Gran Canaria

LAURA PADRÓN BRITO

Universidad de Las Palmas de Gran Canaria

1. INTRODUCTION

The role of extensive reading in the acquisition of a second language has been extensively acknowledged and valued within the field of language education. This recognition is supported by decades of research from Scholars (Bamford, 1984; Day, 1998; Krashen, 1989, 2004; Nation, 2009), who have consistently emphasized the importance of reading large amounts of comprehensible and engaging materials as a means to enhance fluency, vocabulary and overall language proficiency. In addition, extensive reading, often carried out through the use of graded readers, allows learners to build reading speed and confidence, develop general comprehension skills, and revisit grammatical structures and vocabulary in a meaningful and repetitive context (Albay, 2017). In particular, studies (Elley, 1991; Horst, 2005; Iwahori, 2008; Tsang, 1996) have demonstrated the wide-ranging benefits of extensive reading, including improvements in reading rates, vocabulary acquisition, motivation, and even writing and listening skills.

Graded readers, which are simplified text tailored to specific language proficiency levels, remain central to this process. These materials reduce the cognitive demands of reading by using controlled vocabulary, simplified grammar, and linear plot structures (Allen, 2009; Hardy, 2013; Rodrigo, 2016). While some are created based on intuitive methods, others are structurally simplified according to predefined vocabulary and grammar lists, though consistency among publishers remains

limited (Claridge, 2012). Despite these variations, all graded readers' main aim is to provide a scaffolded and accessible reading experience that facilitates gradual transitions to more complex and authentic texts.

Nevertheless, recent trends in classroom contexts reveal a troubling decline in reading engagement. Many learners view reading as a burdensome task rather than a pleasurable or enriching activity. This disinterest is particularly problematic at the upper levels of language instruction, where students are often unprepared to handle literature-based curricula due to limited prior reading experience (Chaves Tesser, 2000; Hardy, 2013). Rodrigo (2016) notes that students frequently drop out of advanced courses because they feel unequipped to transition from elementary to higher level reading demands. We believe that early exposure to graded readers within an extensive reading framework can alleviate this issue by offering learners the opportunity to succeed and find enjoyment in reading from the beginning.

Hence, this study seeks to reaffirm and highlight the pedagogical value of graded readers by examining the progression of linguistic complexity across different levels of these materials. By doing so, it builds on a robust foundation by previous similar research of scholarship advocating for extensive reading as an effective learner centered approach to language development (Cheetham et al., 2016; Claridge, 2012a; Kara, 2019; McQuillan, 2016; P. Nation & Wang Ming-tzu, 1999; Surmanov.S, 2024; Wodinsky & Nation, 1988).

2. OBJECTIVES

The main aim of this research is to investigate how difficulty increases across various levels of graded readers by compiling and analyzing two specific corpora created precisely for this study, one in English and the other in Spanish. Specifically, the research seeks to measure and compare the degree of linguistic progression that is present both within each language and across the two languages. The English-language corpus comprises approximately sixty graded readers drawn from the *Oxford Bookworms* series divided into six levels, of which approximately, the first and second comprise an A2 level, the third and fourth a B1, and

the fifth and sixth a B2, as established by the Common European Framework of Reference for Languages (CEFR). In addition, the Spanish-language corpus includes over thirty titles from the *Instituto Cervantes* collection and are divided into three levels: *Nivel 1*, as an A2 level, *Nivel 2* as a B1, and *Nivel 3* as a B2. That said, the study is particularly concerned with determining whether the increase in textual complexity across these proficiency levels is both gradual and consistent, as well as investigating the extent to which patterns of difficulty progression in English and Spanish may be considered comparable.

3. METHODOLOGY

This research employs a corpus-based approach, using computational linguistic tools to analyze the selected texts. The primary application utilized for this analysis is *Lancsbox X*, which enables the comparison of key lexical features that are central to understanding linguistic complexity. The study places special emphasis on the frequency and distribution of the core word classes, especially pronouns, nouns, adjectives, and verbs, across the texts from the corresponding different CEFR levels. That said, by systematically quantifying and analyzing these lexical elements, the study aims to capture variations in both complexity and lexical diversity that characterize the transition from one proficiency level to the next. The investigation includes both intra-language comparison, meaning within English and within Spanish respectively, and inter-language comparisons between English and Spanish, with the goal of identifying whether similar strategies and patterns are employed to manage linguistic progressing in graded readers across the two languages.

4. RESULTS

Beginning with the results, we will first look at the two tables below, in which we can see the exact numbers per part-of speech (POS) of each category and their increase in percentages. This detailed linguistic analysis of graded reader corpora in English and Spanish reveals significant and nuanced contrasts in how grammatical and lexical elements are distributed across proficiency levels. These distinctions highlight not only

inherent typological differences between the two languages, but also substantial variations in their pedagogical structuring to support second language acquisition. Examination of POS categories across proficiency levels reveals clear, distinct patterns in each language, bearing important implications for the learner's exposure to grammatical complexity, lexical development, and syntactic scaffolding.

4.1. UNNORMALIZED ENGLISH CORPUS RESULTS

TABLE 1. *Unnormalized English Corpus Graded Readers Levels 1 - 3*

Category	Level 1 Freq.	Level 2 Freq.	Level 3 Freq.	% Increase L1 to L2	% Increase L2 to L3
PRON	10,077	11,510	13,290	+14.22%	+15.48%
VERB	9,266	10,513	12,689	+13.47%	+20.67%
NOUN	9,226	10,130	12,386	+9.79%	+22.27%
ADP	5,662	6,646	8,052	+17.37%	+21.18%
DET	5,217	6,122	7,170	+17.33%	+17.13%
PROPN	5,136	5,284	5,063	+2.88%	-4.18%
AUX	3,853	4,782	5,922	+24.09%	+23.85%
ADV	3,541	4,335	5,269	+22.42%	+21.54%
ADJ	3,420	4,049	5,321	+18.39%	+31.43%
CCONJ	3,143	3,447	3,546	+9.67%	+2.87%
PART	2,085	2,289	2,898	+9.78%	+26.63%
NUM	924	958	869	+3.68%	-9.29%
SCONJ	606	1,266	1,685	+108.91%	+33.07%
INTJ	574	407	466	-29.09%	+14.50%

Source: self-production

TABLE 2. *Unnormalized English Corpus Graded Readers Levels 3 - 6*

Category	Level 3 Freq.	Level 4 Freq.	% Increase L3 to L4	Level 5 Freq.	% Increase L4 to L5	Level 6 Freq.	% Increase L5 to L6
PRON	13,290	27,564	+107.41%	30,373	+10.20%	36,207	+19.20%
VERB	12,689	25,543	+101.29%	28,353	+10.99%	35,075	+23.71%
NOUN	12,386	22,010	+77.67%	25,827	+17.34%	31,849	+23.32%
ADP	8,052	15,448	+91.88%	16,533	+7.02%	21,901	+32.47%
DET	7,170	13,438	+87.44%	14,663	+9.10%	18,003	+22.80%
AUX	5,922	12,469	+110.50%	14,040	+12.61%	18,008	+28.29%

ADV	5,269	10,513	+99.49%	11,091	+5.50%	13,841	+24.76%
ADJ	5,321	9,560	+79.63%	10,418	+8.97%	13,701	+31.48%
PROPN	5,063	9,378	+85.27%	11,693	+24.70%	15,216	+30.10%
CCONJ	3,546	6,669	+88.06%	6,592	-1.15%	8,269	+25.44%
PART	2,898	5,968	+106.00%	6,636	+11.19%	8,486	+27.85%
SCONJ	1,685	4,015	+138.27%	4,080	+1.62%	5,745	+40.69%
NUM	869	1,752	+101.57%	1,935	+10.45%	2,402	+24.17%
INTJ	466	737	+58.20%	949	+28.77%	1,090	+14.86%

Source: self-production

As can be seen above, in the English corpus, composed of Levels 1 through 6, a consistent and revealing pattern is the smooth, generally incremental progression in the frequency of major POS categories like pronouns, verbs, and nouns. This trajectory suggests that English graded readers are designed with a deliberate focus on pedagogical scaffolding, as each successive level systematically introduces increasingly complex grammatical forms, building steadily upon previously acquired knowledge. This continuity in development implies a methodologically grounded approach to text grading, where linguistic complexity is introduced in a sequenced and cumulative fashion. Such a smooth trajectory helps manage learners' cognitive load, facilitating the gradual internalization of sophisticated linguistic structures. For instance, pronouns consistently increase in frequency, reflecting the language's requirement for explicit subjects, thereby providing early and repeated exposure essential for understanding sentence structure. Similarly, auxiliary verbs show a steady rise, systematically introducing complex verbal constructions like perfect tenses and passive voice, aligning with the goal of equipping learners with a wider range of grammatical tools.

4.2. UNNORMALIZED SPANISH CORPUS RESULTS

TABLE 3. *Unnormalized Spanish Corpus Graded Readers "Niveles 1 – 3"*

Category	Nivel 1 Freq.	Nivel 2 Freq.	% Increase N1 to N2	Nivel 3 Freq.	% Increase N2 to N3
NOUN	4,332	2,926	-32.47%	4,446	+51.91%
VERB	3,650	2,307	-36.79%	3,831	+66.08%

DET	3,238	2,259	-30.22%	3,350	+48.34%
ADP	3,005	2,220	-26.10%	3,451	+55.59%
PRON	2,060	1,360	-33.98%	2,451	+80.22%
ADV	2,045	1,031	-49.60%	1,778	+72.51%
PROPN	1,523	746	-51.01%	1,473	+97.45%
AUX	1,290	732	-43.26%	1,268	+73.36%
ADJ	1,254	906	-27.74%	1,385	+52.87%
CCONJ	1,027	676	-34.16%	1,039	+53.69%
SCONJ	391	397	+1.53%	924	+132.49%
NUM	308	136	-55.84%	208	+52.94%
INTJ	192	48	-75.00%	179	+272.92%

Source: self-production

Conversely, the Spanish corpus, from *Nivel 1* through *Nivel 3* exhibits a markedly different and often erratic progression profile based on these raw and unnormalized counts. Instead of a steady increase, POS frequencies often show an initial decline from *Nivel 1* to *Nivel 2*, followed by a sharp and unexpected rise at *Nivel 3*. Crucially, this apparent jump is largely influenced by the fact that *Nivel 2* contains significantly fewer words overall compared to both *Nivel 1* and *Nivel 3*. This nonlinear appearance, even before normalization, suggests potential inconsistencies in text difficulty calibration. This lack of smooth, intermediate scaffolding, particularly between *Nivel 2* and *Nivel 3*, indicates learners might face a steeper and more challenging learning curve, abruptly confronting a higher volume of forms which could hinder comprehension. For instance, Spanish pronoun frequency is lower and more unpredictable, dipping in *Nivel 2*. Auxiliary verbs maintain a relatively flat and low frequency, consistent with the Spanish language's rich verb morphology.

These initial findings suggest a contrast between transitional, in the case of English, and abrupt developmental models of language input, in the case of Spanish. English graded readers exemplify a staged, supportive approach, allowing learners to consolidate knowledge. Spanish materials, however, seem to adopt a more abrupt model potentially prioritizing thematic or lexical control over consistent gradual grammatical gradation, leading to less manageable learning curves.

4.3. NORMALIZED ENGLISH CORPUS RESULTS

TABLE 4. *Normalized English Corpus Graded Readers Levels 1 - 3*

Category	Level 1 Norm.	Level 2 Norm.	Level 3 Norm.	% Increase L1 to L2	% Increase L2 to L3
PRON	159.95	159.86	156.35	-0.06%	-2.19%
VERB	147.08	146.01	149.28	-0.72%	+2.24%
NOUN	146.44	140.69	145.72	-3.93%	+3.57%
ADP	89.87	92.31	94.73	+2.71%	+2.63%
DET	82.81	85.03	84.35	+2.68%	-0.79%
PROPN	82.48	73.39	59.56	-11.02%	-18.84%
AUX	61.16	65.72	69.67	+7.46%	+6.01%
ADV	70.65	75.17	61.99	+6.39%	+17.53%
ADJ	54.29	56.24	62.60	+3.59%	+11.32%
CCONJ	49.89	47.88	41.72	-4.04%	-12.86%
PART	33.10	31.79	34.09	-3.94%	+7.24%
NUM	14.67	13.31	10.22	-9.28%	-23.16%
SCONJ	9.62	17.58	19.82	+82.80%	+12.74%
INTJ	9.11	5.65	5.48	-37.96%	-3.01%

Source: self-production

TABLE 5. *Normalized English Corpus Graded Readers Levels 3 - 6*

Category	Level 3 Norm.	Level 4 Norm.	% Increase L3 to L4	Level 5 Norm.	% Increase L4 to L5	Level 6 Norm.	% Increase L5 to L6
PRON	156.35	167.05	+6.84%	165.97	-0.65%	156.74	-5.56%
VERB	149.28	154.81	+3.70%	154.93	+0.08%	156.74	+1.17%
NOUN	145.72	133.39	-8.46%	141.13	+5.80%	138.48	-1.88%
ADP	94.73	93.62	-1.17%	90.34	-3.50%	94.81	+4.94%
DET	84.35	81.44	-3.45%	80.13	-1.62%	77.94	-2.73%
PROPN	59.56	56.85	-4.55%	63.87	+12.35%	65.87	+3.12%
AUX	69.67	75.57	+8.47%	76.72	+1.52%	77.96	+1.61%
ADV	61.99	63.72	+2.79%	60.61	-4.88%	59.92	-1.14%
ADJ	62.60	57.55	-8.06%	55.29	-3.93%	58.00	+4.89%
CCONJ	41.72	34.36	-17.64%	30.61	-10.90%	29.89	-2.35%
PART	34.09	36.17	+6.09%	36.26	+0.26%	36.74	+1.31%
NUM	10.22	10.62	+3.86%	10.57	-0.42%	10.40	-1.66%
SCONJ	19.82	24.33	+22.75%	22.30	-8.38%	23.55	+5.63%
INTJ	5.48	4.47	-18.53%	5.19	+16.10%	4.72	-9.01%

Continuing with the in-depth analysis of the results, we will now delve into the analysis of some of the POS usage across the six levels of the English graded reader corpus. Presented through normalized frequencies per 1,000 words as well as the percentage increase, this analysis reveals systematic, though nonlinear, changes in the frequency and distribution of word categories. The section reports both the overall patterns and more granular movements within each major POS category.

4.3.1. Pronouns (PRON)

The normalized frequency of pronouns starts at 159.95 tokens per 1000 words in Level 1 and shows a consistent, though very slight decrease through the proficiency scale, reaching 156.74 in Level 6. While the absolute use of pronouns increases from 10,077 to 36,207, this growth does not keep pace with the total corpus size at higher levels, reflecting a gradual deemphasis. This could suggest that graded readers progressively shift from personal or simple reference-heavy language towards more content-dense constructions.

4.3.2. Verbs (VERB)

Verb usage maintains remarkable stability across levels, increasing modestly from 147.08 at Level 1 to a peak of 156.74 at Level 6. The most notable changes occur in Level 4, with a 3.70% increase over Level 3, potentially reflecting the introduction of more complex verb forms such as passive voice, modals, or conditionals. However, the consistent frequency suggests verbs serve as a structural backbone throughout those stages of development, reflecting stable communicative needs.

4.3.3. Nouns (NOUN)

Nouns are one of the most frequent POS categories across all levels, starting at 146.44 and decreasing continuously until reaching 138.48 at Level 6. This suggests an early emphasis on content words as learners gain lexical resources, followed by a richer syntactic repertoire at higher levels that distributes emphasis across more categories such as adjectives and adverbs. Between all six levels, nouns show a significant decline, particularly between Levels 3 and 4, with a -8.46% value, supporting the idea

that high proficiency graded readers integrate more modifiers and structure-bearing elements rather than relying solely on lexical content.

4.3.4. Adpositions (ADP)

The normalized frequency of adpositions such as “in,” “on” or “off” starts at 89.87 and rises to 94.73 by Level 3, after which it maintains itself and then rises slightly to 94.81 at Level 6. This growth mirrors improved syntactic control such as prepositional phrases, while the decline at advanced levels may stem from an increased use of subordinate clauses and nominalizations that reduce dependence on simpler structures.

4.3.5. Determiners (DET)

Determiners show moderate fluctuation, starting at 82.81 and peaking at 85.03 in Level 2, then declining steadily to 77.94 in Level 6. These changes align with graded readers’ increasing grammatical sophistication, particularly in syntactic correctness. A 2.68% increase from Level 1 to Level 2 may reflect more precise noun phrase construction, while the -3.45% drop from Level 3 to 4 suggests these texts rely less on basic determiners as they incorporate complex noun modifiers like possessives or quantifiers not categorized as simple DETs.

4.3.6. Proper Nouns (PROPN)

Proper noun usage decreases steadily with continuous negatives percentages until a sudden rise at Level 5 with a 63.87, and continues to increase to 65.87 by Level 6. This may reflect initial reliance on named entities for clarity, which gives way to more abstract or generalized writing.

4.3.7. Auxiliary verbs (AUX)

Auxiliary verbs increased constantly and progressively from 61.16 in Level 1 to a maximum of 77.96 in Level 6. This suggests a peak in complexity involving compound tenses and passive constructions around intermediate and advanced levels, with a more efficient or varied approach at higher levels, perhaps incorporating more nonfinite or elliptical structures.

4.3.8. Adverbs (ADV)

Adverbs exhibit a distinctive trajectory. Starting at 70.65 at Level 1, they peak at 75.17 in Level 2, then decline steadily to 59.92 by Level 6. This sharp reduction from -4.88% from Level 4 to Level 5, suggests that beginner and intermediate graded readers may overuse adverbs for emphasis or clarity, such as *very*, *really* and *quickly*, while advanced texts shift to subtler or more structurally integrated means of modification.

4.3.9. Adjectives (ADJ)

Adjective usage sees an important rise from 54.29 at Level 1 to 62.60 at Level 3, followed by a decline to 55.29 at Level 5 and a slight increase to 58.00 by Level 6. This wavy pattern echoes trends seen in determiners and adverbs, with initial growth in descriptive capability, followed by leaner, perhaps more precise or nominalized expression at advanced levels. The -8.06% dropped from Level 3 to 4 and -3.93 from Level 4 to 5, may signal syntactic streamlining or lexical consolidation.

4.3.10. Coordinating Conjunctions (CCONJ)

CCONJ frequency decreases steadily across levels, from 49.89 in Level 1 to 29.89 in Level 6, with a net drop of over 17% from Level 3 to 4. This shift likely reflects syntactic evolution, as beginner graded readers may rely on simple coordination such as *and* or *but* while more advanced texts may adopt subordinations such as *although* or *because* as well as more integrated clause structures.

4.3.11. Numerals (NUM)

Numerals begin to rapidly decrease from 14.67 at Level 1 to 10.22 at Level 3, and then remain relatively stable, increasing slightly to 10.40 at Level 6. Their utility is likely task or topic dependent, and therefore their frequency is significantly unaffected by overall proficiency.

4.3.12. Subordinating Conjunctions (SCONJ)

Subordinators increase in normalized frequency from 9.62 to 23.55 despite absolute increases. Their percentage change is composed of many

dips and spikes, such as +82.80% from Level 1 to 2 but a -8.38% from Level 4 to 5, suggesting syntactic complexity may increase without requiring a proportional rise in explicit clause linking.

4.3.13. Interjections (INTJ)

Interjections show a steep decrease from 9.11 per 1000 words at Level 1, to 4.47 in Level 4 before increasing to 16.10% at Level 5 and decreasing again. Especially significant is the drop from Level 1 to 2 with a -37.96% difference. This trend suggests a separation from affective or informal language towards more academic, natural registers.

4.4. NORMALIZED SPANISH CORPUS RESULTS

TABLE 6. *Normalized Spanish Corpus Graded Readers “Niveles 1 – 3”*

Category	Nivel 1 Norm.	Nivel 2 Norm.	% Increase N1 to N2	Nivel 3 Norm.	% Increase N2 to N3
NOUN	180.50	182.88	+1,32%	171.00	-6.49%
VERB	152.08	144.19	-5.19%	147.35	+2.19%
DET	134.92	141.19	+4.65%	128.85	-8,74%
ADP	125.21	138.75	+10.82%	132.77	-4.31%
PRON	85.83	85.00	-0.97%	94.27	+10.90%
ADV	85.21	64.44	-24.38%	68.38	+6.13%
PROPN	63.46	46.62	-26.53%	56.65	+21.51%
AUX	53.75	45.75	-14.88%	48.77	+6.60%
ADJ	52.25	56.62	+8,37%	53.27	-5.93%
CCONJ	42.79	42.25	-1.27%	39.96	-5.42%
SCONJ	16.29	24.81	+52.30%	35.54	+43.23%
NUM	12.83	8.50	-33.77%	8.00	-5.88%
INTJ	8.00	3.00	-62.50%	6.88	+129.49%

Source: self-production

Continuing with the Spanish graded reader corpus, though smaller in scale and granularity than its English counterpart, it offers a compelling glimpse into the structuring (or lack thereof), of graded input across 3 proficiency levels, spanning a total of approximately 66,000 words, corresponding 24,000 of them to *Nivel 1*, 16,000 to *Nivel 2*, and 26000

to *Nivel 3*. The previously explained results were normalized to reveal more accurate trends in POS usage, and therefore this section delves into the normalized POS results across most categories and interprets the developmental implications of the observed patterns. Particular attention is paid to the degree of consistency, continuity, and pedagogical logic across all levels.

4.4.1. Nouns (NOUN)

The normalized frequency of nouns fluctuates irregularly, with 180.50 at *Nivel 1*, increasing to 2182.80 at *Nivel 2*, then dipping to 171.00 at *Nivel 3*. This u-shaped curve could reflect thematic variation in the text rather than level dependent complexity. For example, *Nivel 1* may include heavily noun-centered content like family members or concrete vocabulary, while *Nivel 3* might shift towards action-based stories or dialogues that emphasize verbs. The rise at *Nivel 2* could suggest a return to informational or descriptive content, possibly including more academic or abstract vocabulary. However, the inconsistent trajectory makes it difficult to draw strong conclusions about lexical scaffolding and nouns across levels.

4.4.2. Verbs (VERB)

Verb usage shows another unstable trend, beginning at 152.08 and then decreasing by 5.19% at *Nivel 2*, and rising to 147.35 at *Nivel 3*. This change is not too severe, which could indicate that learners are exposed to similar levels of verbal content regardless of the proficiency stage.

Unlike in English, where increased auxiliary usage reflects growing syntactic sophistication, the almost flat trajectory in Spanish suggests that verbal complexity may be managed through morphology rather than frequency. This highlights a typological contrast between the English language's analytic verbal system and Spanish language's synthetic one.

4.4.3 Determiners (DET)

Determiner usage mirrors other categories in its inconsistency, with 134.92 in *Nivel 1*, decreasing 4.65% in *Nivel 2*, and returning to 128.85

in *Nivel 3*. This may point to variation in noun phrase density, perhaps more driven by genre or topic than by developmental considerations. One possible explanation is the fluctuation in the types of nouns used, as a story centered around specific objects or people would naturally contain more definite and indefinite articles, while dialogues or action-heavy passages might rely more on pronouns or verbs, reducing determiner frequency.

4.4.4. Adpositions (ADP)

Adpositions such as *en*, *de* or *a* show a similar irregular pattern, with 125.21 per 1000 at *Nivel 1*, 138.75 at *Nivel 2* and 132.77 at *Nivel 3*. The dip at *Nivel 3* may again reflect a content shift rather than grammatical progression. Importantly, prepositions in Spanish hold substantial grammatical weight, such as verb-preposition combinations like *pensar en*, making their pedagogical handling especially relevant. Furthermore, the increase at *Nivel 2* could signal the introduction of more complex syntactic structures, particularly relative clauses, which are subordinate constructions that require more positional phrases. Still, the lack of a smooth upward trend suggests inconsistent scaffolding.

4.4.5. Pronouns (PRON)

Normalized frequencies for pronouns in Spanish are relatively high at all levels but show an inconsistent developmental trajectory. At *Nivel 1*, the frequency is approximately 86 per 1000 words, decreasing slightly to 85.00 per 1000 words in *Nivel 2*, then peaking at 94.27 in *Nivel 3*.

This nonlinear pattern suggests that the presence of pronouns in Spanish graded readers is not tightly controlled or scaffolded by level. In pedagogical terms, pronouns, especially subject pronouns, play a complex role in Spanish due to pro-drop tendencies, which may result in less overt pronoun usage even in more advanced texts. Furthermore, the relatively flat frequency range across the first two levels implies that pronoun exposure remains constant, perhaps due to their central function in discourse cohesion and narrative clarity, but they are then further implemented in the last level.

4.4.6. Adverbs (ADV)

Adverbs prove once again to be quite unstable, being at 85.21 per 1000 words at *Nivel 1*, decreasing by 24.38% to 64.44 at *Nivel 2*, and increasing to 68.38 at *Nivel 3*. This fluctuation suggests that adverbial usage is not significantly expanded or restricted by level because, like verbs, adverbs in Spanish are morphologically simple and semantically clear, making them accessible at all stages. Their use may thus be dictated by the narrative context, such as describing how or when actions occur than by grammatical complexity.

4.4.7. Proper Nouns (PROPN)

Proper nouns fluctuate severely, increasing and decreasing between levels almost exactly proportionately, as a decrease of 26.53% from *Nivel 1* to *Nivel 2* can be seen, which then increases by 21.51% to *Nivel 3*. The initial drop may reflect a move from highly personalized stories towards more general narratives, with the return to higher frequency at *Nivel 3* suggesting renewed use of named entities in more complex plots.

4.4.8. Auxiliary Verbs (AUX)

Auxiliary verb usage fluctuates slightly but lacks a coherent trend, from 53.75 in *Nivel 1* down to 45.75 in *Nivel 2* and back to 48.77 in *Nivel 3*. These figures are notably lower than in English, which reached over 70 per 1,000 words at advanced levels. The modest and uneven presence of auxiliaries likely reflects the limited reliance of the Spanish language on periphrastic tenses constructions such as *ha comido* or *está leyendo* compared to English. Additionally, the pedagogical approach in Spanish graded texts may prioritize simple conjunctive verbs like *comió* or *leyó* over compound forms in early stages, only introducing periphrases sporadically at higher levels.

4.4.9. Adjectives (ADJ)

Adjectives range narrowly from 52.25 at *Nivel 1* to 56.62 at *Nivel 2*, and then fall to 53.27 at *Nivel 3*. This final modest increase, in comparison

with the first level, aligns with expected developmental trends where learners gradually enrich their descriptive vocabulary. However, the flat slope again suggests minimal explicit scaffolding or lexical elaboration, which may indicate that writers of graded readers allow for flexible descriptive density at all levels, rather than strictly reserving adjectives for more advanced stages.

4.4.10. Coordinating Conjunctions (CCONJ)

Coordinating conjunctions maintain a downwards slope, as they decrease slightly from 42.76 per 1000 words at *Nivel 1*, 42.25 at *Nivel 2* and 39.96 at *Nivel 3*. These values could suggest that simple sentence chaining using *y* and *pero* is introduced early on and then not further emphasized.

4.4.11. Subordinating Conjunctions (SCONJ)

SCONJ usage is low but steadily incremental, with 16.29 at *Nivel 1*, 24.81 at *Nivel 2* and 35.54 at *Nivel 3*, with a 52.30% and 43.23% increase per level respectively. These figures suggest that complex subordinations such as *aunque*, *cuando* or *porque* are not heavily emphasized, even though more used at higher levels, or are instead handled through nominal clauses that don't rely on SCONJ forms.

4.4.12. Numerals (NUM)

NUM use follows a clear negative trajectory, as it decreases a steep 33.77% from *Nivel 1* to *Nivel 2*, and then a more moderate -5.88% from *Nivel 2* to *Nivel 3*. As in English, their utility is likely task or topic dependent, and therefore, their lack of employment may not necessarily be due to the increase in level but to the changes in narrative in the specific texts.

4.4.13. Interjections (INTJ)

Finally, INTJ usage is completely irregular, as it decreases 62.50% from *Nivel 1* to *Nivel 2*, and then shoots back up again by 129.29% at *Nivel 3*. The extremely clear lack of pattern reveals once again a lack of scaffolding at least regarding POS categories in the Spanish graded reader corpus.

5. DISCUSSION

5.1. ENGLISH CORPUS

The progression of POS usage across proficiency levels in the English graded reader corpus reflects the developmental trajectory of English acquisition, stylistic maturation and structural control. These shifts suggest that as levels advance, the texts not only expand their lexicon but also refine the balance and integration of grammatical structures in increasingly sophisticated ways.

Generally, early-stage graded readers rely heavily on functional vocabulary, pronouns, determiners, basic verbs and adverbs, which help construct simple, intelligible messages. This is evident in the high normalized values for PRON, DET and ADV at Level 1. These categories show a consistent decline across levels, reflecting a reduction in overt signaling and redundancy. For instance, adverbs decrease 4.88% from Level 4 to 5, signaling a transition from over-explanation towards more precise expression.

Simultaneously, the rise of content-bearing words such as nouns, adjectives and verbs through the intermediate levels suggests lexical growth, yet not without its limitations, as the plateau and eventual decline in NOUN and ADJ frequencies, 138.48 and 58.00 per 10000 words respectively at Level 6, imply a shift to structural complexity rather than lexical density, like noun phrase embedding, relative clauses or verb-based abstraction replacing adjectival descriptions.

In addition, the diverging trends in CCONJ and SCONJ are particularly telling. Coordinators drop significantly while subordinators flatten, suggesting a transition from parataxis to hypotaxis, a key to academic and formal writing. This aligns with increased AUX use up to Level 6 with 77.96 per 1000 words, likely reflecting complex verb tenses such as *has been writing*, or *will have completed*.

Furthermore, interjections and adverbs show a consistent decline. INTJ drops staggeringly from 9.11 in Level 1 to 4.72 in Level 6, while ADV usage is trimmed from 70.65 at the beginning to 59.92 in the last level. These declines suggest that graded readers want learners to gain genre

awareness and begin adopting the concise, neutral style expected in academic professional contexts. Similarly, the modest fall and rise of PROPEN indicate that reliance on abstraction gives way to specific referents. Finally, many categories like DET, NOUN, VERB, ADJ peak at intermediate levels before plateauing or declining. This may represent a transitional phase where graded readers portray sufficient vocabulary to express complexity but have not yet refined their structural economy. By Levels 5 and 6, graded readers seem to optimize, using fewer tokens to achieve more nuanced, syntactically embedded expression, suggesting an integration of grammatical function into more complex fluent constructions without needing high repetition.

5.2. COMPARISON WITH SPANISH CORPUS

The comparative analysis of English and Spanish graded readers corpora, both in absolute and normalized terms, reveals striking differences in linguistic progression, grammatical scaffolding, and potential pedagogical design. This section synthesizes the patterns uncovered in both corpora, moving from unnormalized (absolute) figures to normalized (per 1,000 word) frequencies. In doing so, we evaluate the extent to which each language corpus supports a structured linguistic progression for language learners and discuss what these findings imply about the design philosophies behind the graded materials.

First of all, before analyzing linguistic patterns, it is important to highlight a foundational difference; the number of levels and the scale of data differ significantly. The English corpus is spread across six levels (Levels 1-6), amounting to over 800,000 words, while the Spanish corpus spans only across three levels (*Niveles 1-3*) and contains almost 70,000 words in total. This is due to the fact that we were only able to find these specific collections of graded readers online, especially in the case of Spanish, as it appears that they are less common. This asymmetry in granularity of scale limits the ability of the Spanish corpus to present gradual development in learner exposure to grammar. The English data set offers a smoother continuum of language acquisition, with level progression reflected in both the quantity and complexity of grammatical categories. The Spanish data, in contrast, appear to offer more

abrupt transitions between levels, possibly due to its smaller scale or a different pedagogical orientation.

In addition, in absolute terms, English shows a clear, almost linear increase in frequency for most major POS categories, as the levels progress. Pronouns, auxiliary verbs, determiners, and nouns all display substantial growth from Level 1 through Level 6. For example, pronouns rise from 10,077 in Level 1 to 36,207 in Level 6. Auxiliary verbs more than triple, going from 3,853 to 18,008, and verbs as a whole, go from a little over 9000 instances to a little over 35,000. This reflects a robust pedagogical structure, because as learners progress, they are systematically exposed to more grammatical material.

The Spanish absolute figures, by contrast, reveal a sharp dip between *Nivel 1* and *Nivel 2*, followed by a resurgence at *Nivel 3*. For example, pronouns fall from 2,060 in *Nivel 1* to 1,360 in *Nivel 2* before climbing back to 2,451 in *Nivel 3*. Moreover, auxiliary verbs follow a similar path from 1,290 to 732 and then up to 1,473. Such inconsistencies could reflect semantic or stylistic changes in the text rather than learner progression. For example, *Nivel 2* texts may emphasize dialogue with fewer descriptive elements, resulting in temporarily reduced noun and determiner use. However, this does not suggest a deliberate pedagogical scaffold structure akin to that observed in English. However, it should also be kept in mind that *Nivel 2* has a much lower total number of words, causing this steep dip as well.

Continuing with the normalization of POS categories per 1,000 words, this provides a more accurate picture of linguistic density and progression. For English, normalized results confirm and strengthen the pattern seen in absolute frequencies. For example, pronouns grow from around 160 to around 175 per 1,000 words at its maximum, auxiliary verbs increase from around 61 to around 78 per 1,000 words, and verbs show a steady climb reflecting greater narrative and syntactic complexity, as well as adpositions, which follow a similarly incremental trajectory and a dip in determinants. These patterns suggest deliberate pedagogical scaffolding. The consistent increase in auxiliary verbs, for instance, signals an intentional progression from simple past or present tenses towards compound forms like present perfect or passive voice, which are essential

for upper intermediate proficiency. Moreover, subordinating conjunctions also increase, supporting more complex sentence structuring.

In the Spanish normalized data, many of the fluctuations persist despite corpus size adjustments. For example, auxiliaries show inconsistent use from 53.75 to 45.75, to 48.77, and determinants follow an increase and then a dip from 134.92 to 141.19, and to 128.85 per 1,000 words, while nouns and adpositions show similarly irregular patterns.

While these fluctuations are less extreme than their raw figures suggested, the absence of a consistent upwards trend implies a lack of graded grammatical progression, as even verbs, adverbs, and adjectives remain quite unstable across levels. Unlike in English, where grammar is clearly layered and built upon, Spanish seemed to expose learners to certain structures early on and later revisit them at differing frequencies without clear advancement or gradual progression.

These contrasts reflect not just differences in corpus design, but deeper typological and pedagogical principles. Firstly, English is a more analytic language, relying on auxiliaries, prepositions and determiners to signal tense, aspect and grammatical relations, which needs a gradual introduction of these forms from the beginning and as proficiency increases. However, Spanish is more synthetic, with tense and agreement information encoded morphologically, and therefore, increased grammatical complexity does not always require new lexical items, but may come from internal word changes, which cannot be captured in frequency counts of POS categories. Nevertheless, these typological features do not entirely explain the lack of structured growth in Spanish POS categories. The English corpus appears to be explicitly pedagogical in design, increasing grammatical density and variety in step with learner level, but Spanish in contrast, seems thematically or narratively driven, where grammatical features vary based on story type rather than learner proficiency. This suggests different assumptions about how language learning occurs, either via structured input or through contextual exposure.

Moreover, the English data hints towards a controlled pedagogical curriculum behind the graded readers, likely shared by the CEFR framework. The progression from simple to complex structures, including increased use of auxiliaries, pronouns, and conjunctions, seems deliberate and measurable. In contrast, the Spanish readers may prioritize

authenticity, thematic engagement or narrative coherence, over grammatical scaffolding. For instance, *Nivel 3* might favor action-driven stories with more verbs and fewer nouns, while *Nivel 2* could focus on expository or descriptive texts with denser nouns and determiner use, which to in order to confirm this, it would need of a detailed reading of all texts. If this were true nevertheless, this would raise important questions about how progression is defined in Spanish reading materials, possibly not strictly by grammar, but by discourse or vocabulary domains.

Finally, regarding the limitations and methodological considerations of the research, it must be said that the disparity in corpus size, specifically 800,000 words in English versus 66,000 words in Spanish, and the number of levels, pointedly, six levels in English versus three levels in Spanish limits direct comparability. The English corpus allows for finer grained distinctions, while the Spanish data compresses beginner, intermediate, and advanced stages into just three bands. Additionally, the analysis is limited to POS frequency and does not account for morphological complexity, lexical diversity, or syntactic structure. A learner exposed to the same number of verbs may still experience vastly different cognitive loads depending on verb conjugation complexity, such as *comer* versus *hubiéramos comido*. Hence, a more in-depth investigation would require a closer examination of all texts and very likely, finding a completely different Spanish graded reader corpus.

6. CONCLUSIONS

In sum, this comparative study reveals stark differences in the grammatical progression across English and Spanish graded reader corpora. The English texts demonstrate a consistent, pedagogically scaffolded structure, increasing the density and diversity of grammatical categories as learner levels progress. This likely reflects a curriculum-driven approach rooted in language acquisition theory and CEFR guidelines.

In contrast, the Spanish graded readers show irregular and often non-linear trends even after normalization. Although this may partly be explained by the typological features of Spanish, it also suggests a looser or more intuitive approach to level design, possibly prioritizing narrative content over structured grammatical instruction.

These findings have significant implications for language educators and material designers. For English, the current created reader design in the *Oxford Bookworms* collection appears to effectively support incremental grammatical acquisition, while for Spanish, however, the inconsistencies in the *Instituto Cervantes*'s collection may hinder learners' ability to build upon previously introduced grammatical structures, potentially leading to gaps or plateaus in acquisition. To address this, future Spanish graded readers could benefit from adopting a more deliberate progression of grammatical difficulty informed by empirical corpus analysis.

Finally, further research could expand this analysis to include morphological variation, syntactic complexity, and discourse-level features, providing a more holistic understanding of how grammatical input is structured across languages and graded materials. Moreover, qualitative investigations into reader comprehension and acquisition outcomes would complement frequency-based studies by offering deeper insights into the effectiveness of different design philosophies and second language reading pedagogy.

7. REFERENCES

- Albay, M. (2017). The Benefits of Graded Reading. *International Journal of Social Sciences & Educational Studies*, 3(4).
<https://doi.org/10.23918/ijsses.v3i4p177>
- Allen, D. (2009). A Study of the Role of Relative Clauses in the Simplification of News Texts for Learners of English. *System*, 37, 585–599.
- Bamford, J. (1984). Extensive Reading by Means of Graded Readers. *Reading in a Foreign Language*, 2(2), 218–260.
- Chaves Tesser, C. and D. R. Long. (2000). The Teaching of Spanish Literature: A Necessary Partnership between the Language and Literature Sections of Traditional Departments. *Foreign Language Annals*, 33, 605–613.
- Cheetham, C., Harper, A., Elliott, M., & Ito, M. (2016). Assessing Student Attitudes Toward Graded Readers, MReader and the MReader Challenge. *The Reading Matrix: An International Online Journal*, 16(2).
- Claridge, Gillian. (2012). Graded Readers: How the Publishers Make the Grade. *Reading in a Foreign Language*, 4(1), 106–119.

- Day, R. R., & B. J. (1998). Extensive Reading in the Second Language Classroom. *Cambridge University Press*.
- Elley, W. B. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning, 41*.
- Hardy, Jane. (2013). Getting Students to Enjoy Reading in a Foreign Language. *The Language Educator, 8*(6), 37–39.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review, 61*, 355–382.
- Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language, 20*(1), 70–91.
- Kara, B. (2019). The Effect of Graded Readers on Reading Comprehension Skill of EFL Students. *Language Teaching and Educational Research, 2*(2), 160–172. <https://doi.org/10.35207/late.554716>
- Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal, 73*(4), 440–464.
- Krashen, S. D. (2004). The power of reading: Insights from the research. *Heinemann*.
- McQuillan, J. (2016). *What can readers read after graded readers?* 28(1), 63–78. <http://nflrc.hawaii.edu/rfl>
- Nation, I. S. P. (2009). Teaching ESL/EFL Reading and Writing. *Routledge*.
- Nation, P., & Wang Ming-tzu, K. (1999). Graded Readers and Vocabulary. *Reading in a Foreign Language, 12*(2).
- Rodrigo, V. (2016). Graded readers: Validating reading levels across publishers. *Hispania, 99*(1), 66–86. <https://doi.org/10.1353/hpn.2016.0027>
- Surmanov, S. (2024). A COMPARATIVE ANALYSIS OF READING SKILL DEVELOPMENT THROUGH GRADED READERS. *Sustainable Agriculture, 2*(22).
- Tsang, W. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics, 17*(2), 627–642.
- Wodinsky, M., & Nation, P. (1988). Learning from Graded Readers. *Reading in a Foreign Language, 5*(1).