ITL - International Journal of Applied Linguistics

Morphological Complexity and Rated Writing Proficiency: The Case of Verbal Inflectional Diversity in L2 Spanish

--Manuscript Draft--

Manuscript Number:	ITL-20009R2
Full Title:	Morphological Complexity and Rated Writing Proficiency: The Case of Verbal Inflectional Diversity in L2 Spanish
Short Title:	
Article Type:	Article
First Author:	Nuria de la Torre García
Other Authors:	María Cecilia Ainciburu
	Kris Buyse
Corresponding Author:	Nuria de la Torre García Universidad Nebrija Madrid, Madrid SPAIN
Funding Information:	
Section/Category:	Article
Keywords:	linguistic complexity; inflectional diversity; Morphological Complexity Index; L2 Spanish writing; rated proficiency
Abstract:	Linguistic complexity measures are used to describe second language (L2) performance and assess levels of proficiency and development. Although morphology is considered crucial in L2 acquisition, morphological complexity has been relatively neglected, hindering comprehensive views of grammatical complexity in L2. This article presents an application of a recently proposed metric of morphological diversity, the Morphological Complexity Index (MCI), in an L2 Spanish corpus of 113 essays classified into four proficiency levels by expert evaluators. The aim of the study is to investigate the relationships of MCI with subjectively rated proficiency and with other four measures of L2 complexity. Results indicate that morphological complexity, as measured by MCI, does not vary significantly across proficiency levels in this corpus. The MCI shows significant correlations with lexical but not with syntactic complexity measures. Findings are interpreted in the light of the characteristics of the corpus and the acquisition of the Spanish verbal system.
Author Comments:	
Order of Authors Secondary Information:	

Morphological complexity and rated writing proficiency: the case of verbal inflectional diversity in L2 Spanish

In the field of writing research and within the Complexity Accuracy and Fluency approach, linguistic complexity has been operationalized with a considerable number of measures that have been utilized to gauge writing proficiency, quality and development (see, Crossley, 2020; Lan, Liu & Staples, 2019, for recent reviews). Linguistic complexity is considered a multidimensional construct that can be assessed in different domains of the language system (for a taxonomy of the construct, see, Bulté & Housen, 2012). Traditionally, syntactic and lexical complexity have been the most studied sub-constructs of linguistic complexity, and morphological complexity has been relatively neglected in SLA research. The low frequency of studies on morphological complexity contrasts with the high importance given to the morphological component in SLA, as it has been claimed that the large amount of syntactic-semantic information encoded in inflectional morphemes could explain many aspects of interlanguages (e.g. Slabakova, 2014). Furthermore, as De Clercq & Housen (2019) indicate, the trade-offs between different linguistic subsystems found in L2 development research suggest that the measurement of linguistic complexity should not be reduced to the lexical and syntactic dimensions, especially when the target language is considered morphologically rich.

Despite the relative prominence of complexity in the literature, there is no agreement on the definition of the construct nor, therefore, on its consistent operationalization. It has been defined in at least two ways, as relative complexity or difficulty and as structural or absolute complexity.

The notion of relative complexity is cognitive in nature and refers to the effort and resources that language users have to invest to make use of a linguistic element. In the

field of typological linguistics, where morphological complexity has received much attention, the relative approach to complexity has given rise to proposals that explain language change and evolution by relating language contact, especially if extensive L2 learning is involved, to morphological properties that cause difficulties to adult L2 learners (Kusters, 2008)

By contrast, the notion of absolute or structural complexity is quantitative in nature and user-independent. It is one of the objective factors, along with others such as frequency or saliency, that contribute to the greater or lesser difficulty with which certain linguistic features are learned and used (Bulté & Housen, 2012). In the structural approach, morphological complexity can be assessed taking into account both grammatical meaning (e.g. number of functional distinctions that are grammaticalized) and its encoding (e.g. number of morphemes, clarity on form-meaning correspondences). In information-theoretical approaches, complexity is often measured following criteria of length of description e.g. if multiple meanings correspond to a single morpheme in a given system, its description will be longer because additional specifications are required to account for a form-meaning relationship that is not one-to-one (Miestamo, 2008).

In SLA research, traditionally less interested in morphological complexity than typological linguistics, the construct has been measured by means of metrics such as frequency of tensed forms, frequency of modals, number of different verb forms and variety of past tense forms (Bulté & Housen, 2012). Pallotti (2015b) points out the questionable validity of such measures which rely on form–function relationships in unstable interlanguages or conflate structural complexity and difficulty.

In response to the need for an objective metric of complexity development in L2 inflectional systems, Pallotti (2015b), followed by Brezina and Pallotti (2019), proposes the Morphological Complexity Index (MCI). The MCI operationalizes morphological complexity in terms of inflectional diversity. A text is more complex if it contains a broader variety of inflectional types in a given word class. It is irrelevant whether certain forms could be considered more difficult or more advanced from the point of view of interlanguage development (Brezina & Pallotti, 2019, p. 100). Thus, a series of three Spanish verb forms, such as *com-e-s* (eat:PRESENT.INDICATIVE.2-SG), *com-e-n* (eat:PRESENT.INDICATIVE.3-PL), *com-e-mos* (eat:PRESENT.INDICATIVE.1-PL) is more complex than a series like *com-ie-ran* (eat:IMPERFECT.SUBJUNCTIVE.3-PL), *beb-ie-ran* (drink:IMPERFECT.SUBJUNCTIVE.3-PL), *corr-ie-ran* (run:IMPERFECT.SUBJUNCTIVE.3-PL). In the first example, the diversity of inflectional types is wider: three tokens and three inflectional types *(-s, -n, -mos*). In the second example, there are three tokens, but just one inflectional type (*-ran*). Complexity is higher in the first series, even though the suffix for the imperfect subjunctive in the second one is usually considered of later acquisition in interlanguage development or structurally more complex, because of the stem vowel alterations (*com-E-r* 'to eat' → *com-IE-ran*).

To compute the MCI for a given word class in a text, all exponences, i.e. inflectional processes that modify lexical bases, must be identified and extracted (see, Brezina & Pallotti, 2019, p. 105, for details about the notions of exponence and lexical base). Then, the diversity of those exponences must be calculated. To avoid the text length effects of a type-based measure such as MCI, scores are calculated and averaged within and across a number *n* (e.g. 100) of randomly selected subsamples of *k* (e.g. 5) exponences (see, Brezina & Pallotti, 2019, pp. 106–108, for more details). This mathematical

procedure can be carried out online using a tool created for this purpose by Brezina and Pallotti (2015).

This article studies the behaviour of the MCI in the verbal system of written texts in L2 Spanish. The main objective is to examine whether MCI values discriminate between different levels of writing proficiency, as judged by expert evaluators in an international examination. We also examine the sources that contribute to inflectional diversification across the proficiency levels, and the relationship between the MCI and various measures of syntactic and lexical complexity. We intend to contribute data in two directions. On the one hand, morphological complexity, which is seldom present in complexity studies, has the potential to add valuable information for a better understanding of the multidimensional construct of linguistic complexity. In particular, verbal inflectional diversity might be an informative indicator of proficiency and development in a language with a relatively rich inflectional system such as Spanish. On the other hand, and as far as we know, the MCI has not yet been confronted with written language samples classified into proficiency levels by criteria of communicative success. In this regard, complexity measures should be interpreted in the light of their communicative adequacy, because a more complex discourse is not always more effective (Pallotti, 2009).

Before stating our research questions, we summarize previous research that includes MCI as a variable and briefly describe the Spanish verbal inflectional system and its acquisition in L2 contexts. Subsequently, we address the study's methodology and conclude with the presentation and discussion of our results.

The Morphological Complexity Index in Previous Research

Despite the recent nature of the proposal, the MCI has already been applied to study verbal inflectional diversity at different proficiency levels and for different languages, genres and discursive modalities. Brezina and Pallotti (2019) presented two studies, one on Italian and the other on English, both based on written argumentative texts. Participants were university students, Italian-speakers in the study on L2 English and Dutch speakers in the study on L2 Italian. Dutch speakers from secondary schools in Flanders were also the participants in De Clercq and Housen (2019), who worked with a multilingual corpus (French and English) of oral narratives. Bulté and Roothooft (2020) used oral descriptive monologues from the International English Language Testing System (IELTS) speaking test, with participants from different L1 backgrounds. In Brezina and Pallotti's (2019) and De Clercq and Housen's (2019) studies, corpora consisted of texts by native speakers (NSs) and non-native speakers (NNSs) classified into proficiency levels. The levels were determined employing a cloze test in Brezina and Pallotti (2019), and a combination of instruction time and accuracy measures in De Clercq and Housen (2019). By contrast, Bulté and Roothooft (2020) did not had a benchmark group of NSs and their dataset belonged to five adjacent IELTS bands determined by subjective ratings. Brezina and Pallotti (2019) state that the level of the participants in their study on L2 Italian ranged from A2 to B2 of the Common European Framework of Reference (CEFR; Council of Europe, 2001), and from CEFR level B1 to C1 in the study on L2 English. De Clercq and Housen (2019) did not match the level of their participants to any reference. Lastly, the IELTS bands investigated by Bulté and Roothooft (2020) are related to CEFR levels A2/B1 to C1.

The results of all these studies somehow lead to the same conclusion: morphological complexity is a function of proficiency and the degree of inflectional

diversity of the target language (Brezina & Pallotti, 2019, p. 99). The findings indicate a significant growth of inflectional diversity as proficiency increases. This growth is especially visible at lower proficiency levels and, when a baseline of NS production is provided, it is observed that learners reach a certain threshold level, after which mean MCI values are similar to those of NSs and diversity remains constant. That morphological threshold is reached at higher proficiency levels in languages with a relatively complex inflectional system, e.g. French or Italian in contrast to English.

Furthermore, De Clercq and Housen (2019) report crosslinguistic differences regarding the sources of diversification of inflectional morphology. In English, the increase in diversity is due primarily to the presence of irregular forms and past tenses. In the case of French, it can also be explained by "a greater range of regular forms at more advanced levels of development" (De Clercq & Housen, 2019, p. 90).

Referring to relationships with other measures of linguistic complexity, the MCI values correlated with measures of lexical and syntactic complexity in Brezina and Pallotti's (2019) study on Italian, but not in the one on English. Lexical complexity measures and two of the five syntactic complexity measures computed in Bulté and Roothooft's (2020) study, the mean length of clause and the mean length of noun phrase, also correlated with the MCI scores.

The results in previous research point to the fact that the MCI can account for the development of inflectional diversity at different L2 proficiency levels, most probably until NS values are reached. They also demonstrate how inflectional diversity manifests itself in different ways, depending on the degree of morphological complexity of the target languages. In this paper we provide data on the behaviour of the MCI in L2 Spanish at four proficiency levels determined by judgments of writing quality.

The Spanish Verbal Inflectional System

In Spanish, verbs are inflected for person, number, tense, aspect and mood. Inflectional verbal affixes in Spanish are attached to themes or stems, which in turn are combinations of roots and theme or stem vowels. Thus, roots, which carry lexical meanings, can be followed by three morphemes: (1) the semantically empty, stem vowel (SV), (2) the semantically cumulative TAM morpheme that expresses Tense (past, present, future, conditional), Aspect (imperfective or perfective) and Mood (indicative or subjunctive), and (3) the cumulative PN or agreement morpheme that expresses Person (1^{st}, 2^{nd} and 3^{rd}) and Number (singular or plural). For example, the form *cantábamos* (sing:IMPERFECT.INDICATIVE.1PL) can be analysed as follows: ($cant_{ROOT}$-a_{SV})$_{STEM}$+ba_{TAM}+mos_{PN}, where the suffix *-ba-* denotes past tense, imperfective aspect and indicative mood, and the morpheme *-mos* implies a first-person plural subject (Roca, 2010).

The meanings of TAM and PN are external to the verbal form: they are syntactically determined according to the conditions of time, modality, aspect, number, and person of the utterance. Therefore, these two constituents are considered strictly inflectional, in contrast with the SV, determined internally by the root and also found inside derivational affixes, e.g. deverbal nouns such as *rec-I-bidor* 'entrance hall' from *recib-I-r* 'to welcome' or, *com-E-dor* 'dining room' from *com-E-r*, 'to eat' (Alcoba, 1999, p. 4918; Bermúdez-Otero, 2013, p. 28).

Spanish verbs are traditionally distributed into three conjugations determined by the SV appearing in the infinitive *(-a, -e, -i*). The paradigm of each conjugation contains 59 cells corresponding to 10 tenses with 6 PN forms (except the defective imperative, with just two) and three non-finite forms, namely infinitive, gerund, and past participle

(Roca, 2010, p. 409). In addition, there are 8 compound tenses and 2 compound non-finite forms, that is, 50 more cells per conjugation. Besides, more than 1,000 verbs are considered irregular (Busquets & Bonzi, 1993 in Alcoba, 1999, p. 4936). Irregular verbs can be divided into the idiosyncratic or suppletive ones (e.g. *ser* 'to be' → *era* 'I was') and those that have one or more variants with alterations in the root or the suffixes (e.g. *tener* 'to have' → *tengo* 'I have', *tienes* 'you have', *tuve* 'I had'). Therefore, verbal morphological complexity is relatively high and often challenging for L2 Spanish learners.

The Acquisition of Spanish Verbal Inflectional System

L2 Spanish learners, like those of other L2s, show difficulties in processing and producing verbal morphology. At earlier stages, learners rely on lexical cues such as temporal or modal adverbs and overt subject pronouns (not required in Spanish as it is a null-subject language). At later stages, they attend to verbal morphology, but even at advanced stages of development, it is possible to detect uninflected and incorrect forms (Presson et al., 2013). However, according to empirical data and because of certain morphosyntactic characteristics of the Spanish language, Spanish-speaking children and L2 Spanish learners of elementary levels use finite forms and resort less frequently to default infinitives than has been found to be the case in other languages such as French or German (Montrul, 2004).

With regard to the learning of regular and irregular patterns involved in Spanish conjugation, Presson et al. (2013, p. 825) provide data which suggest that learners treat "conjugation as a compositional task rather than a retrieval task for individual inflected forms". This compositional strategy applies both to regular and (not fully suppletive)

irregular verbs and requires more effort and resources at earlier stages of development. At later stages, with more exposure and practice, composition is gradually proceduralised and becomes increasingly automated.

Much research on L2 Spanish focuses on TAM morphology, which is learned before its correct use and semantic function (Collentine, 2010; Montrul, 2004). Spanish TAM morphemes are considered to emerge systematically: "present before past, preterite before imperfect, indicative before subjunctive" (Montrul, 2004, p. 170). The grammaticalization of aspectual and modality distinctions gives rise to difficulties for many learners. Consequently, the acquisition of past tenses and subjunctive mood may take a long time, with a number of developmental stages.

In sum, research on the acquisition of the Spanish verbal system reveals that, overall, it is a challenging domain for learners, which can be affected by attrition and fossilisation. From a perspective of morphological complexity, this study investigates whether the intermediate stages in which verbal morphology is acquired in L2 Spanish are reflected in the learners' production data. More proficient learners are supposed to need less of the resources required to deal with compositional strategies in conjugation and to be more prepared to use the subtleties of the tense-aspect system and the subjunctive/indicative distinction. If this is reflected on writing, greater verbal diversity is expected to characterize texts written at higher levels of proficiency. However, other morphosyntactic characteristics of Spanish may hamper the sensibility of inflectional diversity to account for proficiency. As finite forms emerge relatively early in production, it is possible that morphology at lower levels of proficiency would be already complex enough to be comparable to that of more advanced levels. This study therefore attempts to ascertain whether an objective metric of morphological diversity such as MCI can

effectively account for the complexity of inflectional paradigms deployed at different levels of proficiency in a language with the morphosyntactic characteristics of Spanish.

Method

This section is divided into four subsections. In the first one, our research questions and working hypotheses are specified. The corpus and the general procedure are described in the second and third subsections; the last subsection sets out the linguistic analysis underlying the calculation of the MCI.

Research Questions

This study explores the behaviour of the MCI as a measure of verbal inflectional diversity, a dimension of morphological complexity. The MCI is calculated on L2 Spanish written texts classified into four proficiency levels by expert raters in an international examination. Specifically, we seek to answer the following research questions:

RQ1. Can a measure of morphological complexity such as MCI discriminate between different writing proficiency levels established primarily by criteria of communicative adequacy?

RQ2. What are the sources of morphological diversification as the proficiency level progresses?

RQ3. Does the MCI exhibit similar behaviour to that of other measures of syntactic and lexical complexity traditionally used to assess and predict writing proficiency levels?

On RQ1 we hypothesize that the slow acquisition of the relatively complex Spanish verbal system will be reflected in production data and therefore morphological diversity will increase significantly along proficiency levels in our corpus (H1). This is

in line with the findings of Brezina and Pallotti (2019) on Italian and De Clercq and Housen (2019) on French. Results for English in De Clercq and Housen (2019) and Bulté and Roothooft (2020) also support H1.

Regarding RQ2, our hypothesis is that Spanish will have a similar development to that observed in French by De Clercq and Housen (2019): inflectional diversity will increase primarily as a result of the contribution of a progressively more extensive range of regular forms. Specifically, according to research on acquisition, we hypothesize that the gradual incorporation of past and subjunctive forms will significantly enrich the set of morphemes used at more advanced levels (H2).

Our hypothesis on RQ3 is that MCI will positively correlate with syntactic and lexical complexity measures. On one hand, better lexical access could free up resources to use compositional strategies in conjugation and make available a broader repertoire of inflectional forms. On the other hand, the use of complex syntax (e.g. dependent clauses) could also influence, although in a more limited way, the degree of inflectional diversity, since the use of subjunctive in Spanish is primarily restricted to subordinate clauses (Collentine, 1995). Results for Italian in Brezina and Pallotti (2019) and for English in Bulté and Roothooft (2020) support this hypothesis (H3).

Corpus Description

This study is based on argumentative texts from the written section of the Certificate of Spanish: Language and Use (Spanish acronym CELU) – a high-stakes Spanish proficiency test, developed by the Interuniversity Consortium for the Teaching and Testing of Spanish as a Second and Foreign Language (Spanish acronym CELSE) and officially acknowledged by the Argentine Ministries of Education and of Foreign Affairs. The levels that can be achieved are Basic, Intermediate and Advanced, although only the

last two are certified. According to CELSE (n.d.), the Intermediate level is comparable to CEFR level B2, and to the Advanced Low level of the American Council on the Teaching of Foreign Languages (ACTFL; 2012). In turn, the Advanced level corresponds to CEFR C1 and ACTFL Superior.

The written section of the test consists of four tasks with non-specified text length, but to be completed in a maximum of three hours[1]. Each text is scored independently by two specifically trained raters. One rater uses an analytical rating scale made up of four categories (contextual adequacy, discursive adequacy, morphosyntax, and vocabulary) and the other, a holistic scale focused on successful completion of the task[2]. In cases of discrepancy, a third rater adjudicates. After the grading process, the texts are classified into five groups: advanced, upper-intermediate, intermediate, basic, and non-qualifying. We analysed texts belonging to the first four groups, due to the scarcity of texts from the last group.

Our corpus contains 113 texts written by 113 learners. The texts come from task 4 of two different CELU exam sessions. Task 4 is a writing task integrated with a reading task. In both sessions, it consisted of producing an opinion article for a newspaper from briefing notes and surveys. In one of them, the topic was social networks and their influence on school achievement (n = 57), and in the other, helmet use and road safety (n = 56). In addition, a benchmark group of 25 samples produced by NSs was also analysed. They consisted of opinion articles about social networks and road safety from the media in various Spanish-speaking countries and were randomly collected from the *Corpus de Referencia del Español Actual* (Real Academia Española, 2008).

The overall corpus contains 34,295 words and 5,464 verbal forms. Table 1 presents an overview of the groups. It already reveals that in the learner groups, text length

significantly increases across proficiency, and that NS texts included, on average, fewer verbs than the L2 texts.

Table 1

General Data on the Composition of the Groups

Group	Level	N-texts	N-words	N-verbs	Words/text		Verbs/text	
					M	*SD*	*M*	*SD*
1	Basic	33	6434	1160	195	67.1	35.2	15.7
2	Intermediate	28	6277	1074	224	41.2	38.3	8.00
3	Upper-intermediate	25	6573	1053	263	74.1	42.1	12.1
4	Advanced	27	8391	1295	311	102.0	48.0	17.8
5	Native speakers	25	6620	882	265	31.4	35.2	8.81

Note. N-texts: number of texts, N-words: number of words, N-verbs number of verbal forms, Words/text: number of words per text, Verbs/text: number of verbs per text.

Only 85% of the participants in this study declared their first language (L1) at the time of the exam. According to the available data, they were speakers of 14 different L1s, the most common being German (35%), English (26%) and French (11%).

General Procedure

To compute the MCI, verb forms were identified, and inflectional exponences were manually extracted and coded (see next subsection for details about the linguistic analysis). Codes were inputted to the Brezina and Pallotti's (2015) online tool in order to calculate the MCI scores. The tool parameters were configured to perform 100 random trials on subsamples of 5 elements. This configuration allows for stable calculations of MCI in short texts (Pallotti, 2015a).

In addition to the MCI and to answer RQ3, another four complexity measures were computed. Two represent syntactic complexity: the mean length of finite clause (MLC), i.e. a structure that consists of a verb with an agreement morpheme and does not necessarily include an additional clause element, for example an explicit subject or an

object, and the average number of finite clauses per T-unit, i.e. a main clause and all associated dependent clauses (C/TU). Both measures are widely used in SLA research (e.g. Norris & Ortega, 2009). The remaining two measures represent lexical complexity. Lexical diversity was calculated in terms of D, a measure of the proportion of unique lexical items (types), which is considered independent of text length (Malvern et al., 2004). D was calculated using the VOCD program in CLAN (MacWhinney, 2000), after lemmatizing the texts with the MOR program, also available in CLAN. VOCD was applied to all words (content and function) excluding proper nouns and non-target language items[3]. Lexical sophistication was measured as the percentage of advanced lemma types in the texts. Like in the index known as "Beyond-2,000" (Daller & Xue, 2007), the types not belonging to the list of the 2,000 most frequent lemmas in *A Frequency Dictionary of Spanish* (Davies & Hayward Davies, 2018) were considered advanced[4].

The statistical analyses were conducted with Jamovi 1.2 (The jamovi project, 2020).

Identification and Coding of Inflectional Exponences

The coding of the verbal forms was performed following the criteria and examples of previous MCI implementations in Brezina and Pallotti (2019) and Pallotti (2015a, 2015b) for Italian, and De Clercq and Housen (2019) for French. Allomorphs were treated as separate exponences, unlike the cases of homonymy. Non-existent forms in the target language, when intelligible and analysable, were coded following the recognizable patterns. We only desisted from coding common patterns in the case of highly irregular forms. Additionally, following Olbertz (1998, p. 34), we considered that there are only two true auxiliaries in Spanish, namely *haber* 'to have' (auxiliary for the formation of

compound tenses) and *ser* ‘to be’ (copula that serves as a passive auxiliary). Therefore, those are the only analytical constructions counted as one single morphological operation in our analysis.

There is one aspect in which our coding differs from previous research. While stem vowels (SVs) were treated as part of the inflectional endings in Brezina and Pallotti’s (2019) and De Clercq and Housen’s (2019) analyses, we considered stems – made up of roots and SVs – as the verb lexical bases (LBs) to which suffixes could be attached or which could be modified. A notable consequence of this decision is that both global inflectional diversity and certain possible effects of lexical diversity are reduced. For example, in an analysis that considers roots as LBs, in a series of infinitives such as *hablar, beber, vivir* ‘to speak, to drink, to live’, three inflectional types must be identified: *-ar, -er, -ir*. That series has higher inflectional diversity than *hablar, cantar, bailar* ‘to speak, to sing, to dance’, with just one type, *-ar*. Alternatively, in the analysis that considers the SV inside the LB, the inflectional diversity of both series is the same and is minimal (just the inflectional type, *-r*). The consideration of the stem as the LB has a theoretical basis because the realization of the SV depends on the properties of the root. However, it should also be kept in mind that in Spanish verbs, there is a covariation of the SV depending on the conjugation of the verb and the suffixes attached to the stem (e.g. *com-e-r* ‘to eat’ but *com-ie-ron* ‘they ate’). We treat this type of allomorphy as a stem alteration that produces some additional diversity in all verbs, including the regular ones. This decision allows for a better control of the lexical diversity effects on morphological diversity, but it is also evident, as De Clercq and Housen (2019, p. 93) acknowledge, that “the potential level of morphological diversity is decided beforehand

by the researcher and the morphological analysis may not be fully representative of the learner's interlanguage".

We followed the morphological analysis by Alcoba (1999) for the identification of suffix and stem alterations. The coding of the stems was inspired by De Clercq and Housen's (2019, pp. 82–83) procedure for French. The stems were assigned a numbered code ($T_{(heme)}1$, T2…) to reflect the possible use of two or more different stems for the same verb in a text. The most frequent stem of a given verb in the entire corpus was regarded as the default LB and labelled as T1 (Brezina & Pallotti, 2019, p. 116). The stem alterations only contribute to overall diversity if two different stems of the same verb are found in a text, without identifying them by reference to the target grammar. The calculation of the MCI values was based on codes such as those exemplified in Table 2.

Table 2

Examples of Codes Assigned to Verb Forms to Compute the MCI

Verb form	Code
recordar (remember:infinitive)	T_n-r
recuerda (remember:present.indicative.3-sg / imperative.2-sg)	T_n -Ø
va (go:present.indicative.3-sg)	va
recuerdan (remember:present.indicative.3-pl)	T_n-n
son (be:present.indicative.3-pl)	son
recordaba (remember:imperfect.indicative.3-sg)	T_n-ba
recordó (remember:preterite.3-sg)	T_n-ó
quiso (want:preterite.3-sg)	T_n-o
recordará (remember:future.3-sg)	T_n-rá
recordaría (remember:conditional.3-sg)	T_n-ría
recuerde (remember:present.subjunctive.3-sg)	T_n-e
recordara (remember:imperfect.subjunctive.3-sg)	T_n-ra
recordase (remember:imperfect.subjunctive.3-sg)	T_n-se
recordando (remember:gerund)	T_n-ndo
ha recordado (remember:present.perfect.indicative.3-sg)	ha-T_n -do
había recordado (remember:past.perfect.indicative.3-sg)	había-T_n -do
haya recordado (remember:present.perfect.subjuctive.3-sg)	haya-T_n -do
hubiera recordado (remember:past.perfect.subjuctive.3-sg)	hubiera-T_n -do
hubiese recordado (remember:past.perfect.subjuctive.3-sg)	hubiese-T_n -do
habría recordado (remember:conditional.perfect.3-sg)	habría-T_n -do
fue recordada (remember:preterite.passive.3-sg)	fue-T_n -da

Results

In this section, the results of the statistical analyses are presented for our three RQs. First, we report on the relationship between MCI scores and proficiency levels. Second, we outline the results about the sources of inflectional diversification as proficiency level progresses. Third, we examine the correlations between MCI and measures of syntactic and lexical complexity and compare their contribution to the prediction of proficiency levels.

The Morphological Complexity Index and proficiency levels

Overall, MCI values indicate that, on average, verbal inflectional diversity increases as the texts are considered of higher quality by the judges, although the trend is not linear because of the decrease between the means of the second lowest and the second highest L2 groups. Variance decreases as the level increases, but higher level L2 texts are generally more homogeneous than those of NSs (for a possible explanation of this finding, see Brezina & Pallotti, 2019, p. 117).

Table 3

Descriptive statistics for MCI

Group	*n*	*M*	*SD*
1	33	5.00	0.875
2	28	5.11	0.840
3	25	5.08	0.695
4	27	5.40	0.529
NS	25	5.91	0.779

A one-way ANOVA test reveals significant differences between the groups: $F(4,133) = 6.36$, $p < 0.001$, $\eta^2_p = 0.160$ and post-hoc Tukey tests (Table 4) indicate differences between the native speakers and the learners' groups, except for the highest-level one (group 4). The L2 groups do not differ significantly from one another.

Table 4

Post-Hoc Tukey Tests

(I) Group	(J) Group	Mean difference (I-J)	*t*	Cohen's d	p_{tukey}
1	2	-0.110	-0.561	-0.128	0.980
	3	-0.081	-0.404	-0.101	0.994
	4	-0.400	-2.028	-0.541	0.258
	NS	-0.911***	-4.514	-1.090	< .001
2	3	0.028	0.135	0.036	1.000
	4	-0.291	-1.417	-0.412	0.618
	NS	-0.801**	-3.826	-0.987	0.002
3	4	-0.319	-1.510	-0.519	0.558
	NS	-0.829**	-3.853	-1.123	0.002
4	NS	-0.510	-2.416	-0.772	0.117

Notes. NS = native speakers; Cohen's d does not correct for multiple comparisons.
** p < .01, *** p < .001

An example of a text belonging to the lowest-level L2 group, but exhibiting a considerable verbal inflectional diversity, can illustrate these data. The verbs in the text are followed by the codes used to compute its MCI, which is above the entire corpus average. Verbal forms with markers signalling aspect (preterite/imperfect) and mood (subjunctive), traditionally considered of late acquisition, are present in the text and are used with relative adequacy. The C/TU and D values of the text are also above average, unlike its MLC and Beyond-2,000 scores, which are below the mean values of the L2 groups. The text turns out to be difficult to understand and not effective from the communicative point of view, probably due to imprecision and error frequency (for interactions between accuracy and complexity, see e.g. Polio & Shea, 2014).

> *Estos días tiene* (T1-Ø:3-SG-PRESENT-INDICATIVE) *un tema conflicto que está* (T1-á:3-SG-PRESENT-INDICATIVE) *discutiendo* (T1-ndo:GERUND) *mucho en la ciudad de San Nicolás.*
> *Moto es* (es:3-SG-PRESENT-INDICATIVE) *muy popular en esta época. Nos da* (T1-Ø:3-SG-PRESENT-INDICATIVE) *facilidad en nuestra vida. Pero tambien por otro lado sucede* (T1-Ø:3-SG-PRESENT-INDICATIVE) *algunos problemas ejemplo el tema sobre casco.*
> *Por un lado un grupo de la persona les parece* (T1-Ø:3-SG-PRESENT-INDICATIVE) *bien que use* (T2-e:3-SG-PRESENT-SUBJUNCTIVE) *casco cuando está* (T1-á:3-SG-PRESENT-INDICATIVE) *coducido* (T1-do:PARTICIPLE). *Tiene algunos numeros que pueden* (T1-n:3-PL-PRESENT-INDICATIVE) *expresar* (T1-r:INFINITIVE) *ventaje de*

que use (T2-e:3-SG-PRESENT-SUBJUNCTIVE) *casco. Cuando ocurría* (T1-a:3-SG-IMPERFECT-INDICATIVE) *accidente la persona quién estaba* (T2-ba:3-SG-IMPERFECT-INDICATIVE) *conduciendo* (T3-ndo:GERUND) *usó* (T2-ó:3-SG-PRETERITE) *casco dentro de ellos 73% menos de mortalidad, 85 menos de lesiones graves, 76% más de efectividad en la prevención de lesiones cerebrales.*
Por otro lado, otro gurpo de la persona a ellos no los gusta (T1-Ø:3-SG-PRESENT-INDICATIVE) *usar* (T1-r:INFINITIVE) *casco cuando está* (T1-á:3-SG-PRESENT-INDICATIVE) *manejando* (T1-ndo:GERUND) *el moto. Creen* (T1-n:3-PL-PRESENT-INDICATIVE) *que el casco afecta* (T1-Ø:3-SG-PRESENT-INDICATIVE) *sus visiones, no pueden* (T1-n:3-PL-PRESENT-INDICATIVE) *ver* (T1-r:INFINITIVE) *bien de costado. Otro mito es* (es:3-SG-PRESENT-INDICATIVE) *lo que con el casco ellos no pueden* (T1-n:3-PL-PRESENT-INDICATIVE) *oir* (T1-r:INFINITIVE) *bien los sonidos de la calle. El último motivo es* (es:3-SG-PRESENT-INDICATIVE) *lo que en caso de accidente, los cascos provocan* (T1-n:3-SG-PRESENT-INDICATIVE) *heridas en el cuello o médula espinal.*
A mi opinión, es (es:3-SG-PRESENT-INDICATIVE) *mejor lo que usamos* (T1-mos:1-PL-PRESENT-INDICATIVE) *casco cuando condicimos* (T1-mos:1-PL-PRESENT-INDICATIVE). *Si sucediera* (T2-ra:3-SG-IMPERFECT-SUBJUNCTIVE) *accidente, podríamos* (T2-ríamos:1-PL-CONDITIONAL) *tener* (T2-r:INFINITIVE) *menos heridas por que nuestros cabezos fueron protegido* (fueron-T1-do:3-PL-PRETERITE-PASSIVE) *por el casco.*
Cuida (T1-Ø:2-SG-IMPERATIVE) *la vida cuya. Usa* (T1-Ø:2-SG-IMPERATIVE) *casco cundo conduzca* (T3-a:3-SG-PRESENT-SUBJUNCTIVE) *moto en el futuro.*

The Sources of Inflectional Diversification

We used the same procedure as De Clercq and Housen (2019) to find out whether the diversity computed by the MCI had different sources across the proficiency levels. We recoded suppletive (e.g. *es* be:3-sg-present-indicative: *ser*) and irregular forms (e.g. *conduzca* drive:3-sg-present-subjunctive: *conducir*) into a single category and calculated a second version of the MCI. This version was computed from all the different codes corresponding to regular exponences (e.g. T1-Ø, T1-n) and a single code for all irregular and suppletive forms ("irregular"). Table 5 compares the means from both analyses.

Table 5

Mean and Standard Deviation for MCI with collapsed suppletive and irregular forms and Difference between Mean Scores for Both MCI Versions

Group	*n*	*M*	*SD*	*Difference*
1	33	3.89	0.770	1.11

2	28	4.03	0.717	1,08
3	25	4.25	0.772	0.831
4	27	4.47	0.484	0.931
NS	25	5.43	0.894	0.484

Descriptive statistics for this MCI version indicate a linear growing of morphological diversity along the proficiency levels. The difference between scores on both versions tends to be smaller in the more advanced groups, and much smaller in the L1 group. Although strong correlations are found between both MCI versions (Pearson's r = 0.812, p<.001, n = 138), the ANOVA for this second version reveals larger differences across the groups: the overall effect size is larger (F(4,133) = 18.05, p < 0.001, η^2_p = 0.352) and Tukey tests demonstrate significant differences not found in the first version between the lowest and the highest-level L2 groups and between the highest-level L2 and the L1 groups (Table 6).

Table 6

Post-Hoc Comparisons for MCI with collapsed suppletive and irregular forms

(I) Group	(J) Group	Mean difference (I-J)	*t*	Cohen's d	p_{tukey}
1	2	-0.135	-0.713	-0.181	0.953
	3	-0.358	-1.827	-0.464	0.363
	4	-0.577*	-3.010	-0.877	0.026
	NS	-1.534***	-7.839	-1.858	< .001
2	3	0.222	-1.095	-0.299	0.809
	4	-0.441	-2.217	-0.719	0.180
	NS	-1.399***	-6.888	-1.737	< .001
3	4	-0.219	-1.069	-0.343	0.822
	NS	-1.177***	-5.636	-1.408	< .001
4	NS	-0.958***	-4.674	-1.347	< .001

Notes. Cohen's d does not correct for multiple comparisons.
* p < .05, *** p < .001

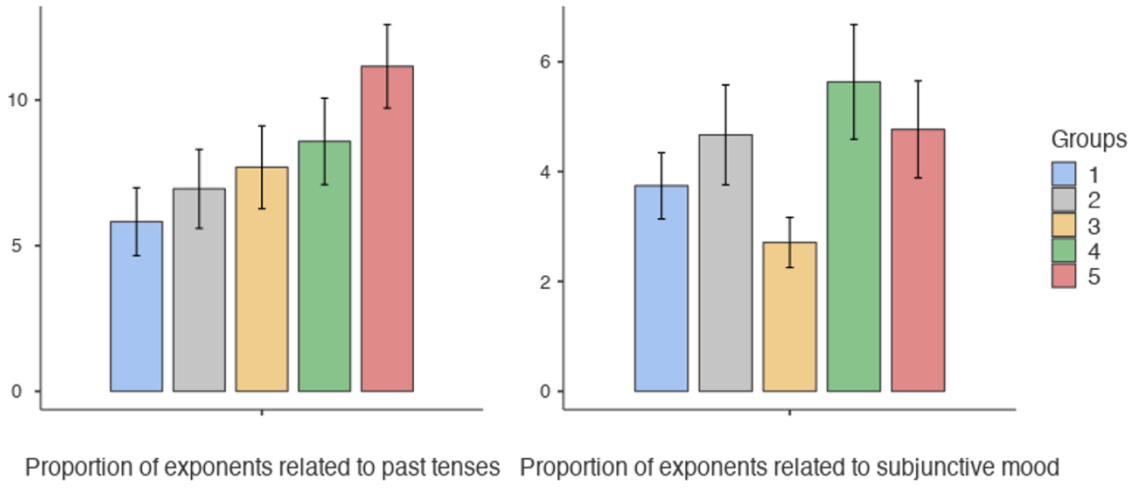
These data suggest that the contribution of irregular and suppletive forms to morphological diversity decreases as proficiency increases. Therefore, the results would

imply that the observed increase in inflectional diversity along the proficiency levels is primarily due to the use of an evolving range of regular exponents.

In an attempt to better characterize the morphological diversification in our corpus beyond the regular/irregular nature of the forms, we carried out a tentative analysis involving past tenses and subjunctive mood, which, according to research on L2 Spanish, could enrich learners' morphology at advanced stages. For this analysis we examined the codes used to compute the MCI in the NS' texts and isolated those corresponding to past tenses (e.g. T_n-ió as in *repartió* (deliver:3-sg-preterite), T_n-ba as in *atacaba* (attack:3-sg-imperfect-indicative), ha-T_n-do as in *ha creado* (create:3-sg-present perfect-indicative) or había-T_n-do as in *había vivido* (live:3-sg-past perfect-indicative)) and subjunctive mood (e.g. Tn-e as in *compre* (buy:3-sg-present-subjunctive), Tn-ra as in *hablara* (speak:3-sg-imperfect-subjunctive), haya-Tn-do as in *haya mirado* (look:3-sg-present perfect-subjunctive) or hubiera-Tn-do as in *hubiera sentido* (feel:3-sg-past perfect-subjunctive)). Then we calculated the proportion of those exponents in every text of the entire corpus[5]. Bar graphs in Figure 1 show the percentage that these forms represent with respect to the total number of verb forms used in the texts for each group. The mean for the past tense forms ranged from 5.82% in the lowest-level group (group 1) to 11.2% in the NS' texts (group 5). Subjunctive forms are scarcer, present a non-linear pattern and ranged from 2.71% in the second highest L2 group (group 3) to 5.63% in the highest L2 level group (group 4).

Figure 1

Proportion of Exponents Related to Past Tenses and Subjunctive Mood



Despite the great variation found inside the groups, the proportion of past tense forms in the texts correlates with the MCI (Spearman's *ro* = 0.555, *p*<.001, n = 138) and with the level (Spearman's *ro* = 0.296, *p*<.001, n = 138). On the contrary, forms related to the subjunctive mood do not correlate with the level (Spearman's *ro* = 0.076, p = 0.378, n = 138), although they seem to contribute to increasing the MCI scores (Spearman's *ro* = 0.387, p<.001, n = 138).

These data suggest that, as learners become more proficient, they are able to use more past tense forms and, consequently, their texts become more complex from the morphological point of view. However, the use of subjunctive forms does not seem to be related to proficiency. Individual stylistic choices could partly explain the data, including the high variation found inside the groups. A comparison of more complex and simpler texts indicates that the inclusion of past tense narrative moves in the argumentations leads to higher MCI scores in all groups. Regarding the subjunctive forms, texts that express emotional reactions and moral judgments also tend to have higher MCI scores than texts that mainly use the exposition of facts and beliefs. The argumentation can be equally effective in both cases, but, in the first case, subjunctive forms are more frequent and MCI values tend to be higher.

The Morphological Complexity Index and Syntactic and Lexical Complexity

Descriptive statistics for the syntactic and lexical complexity measures in the L2 groups (Table 6) indicate that syntactic complexity (MLC and C/TU) and lexical sophistication (Beyond-2,000) values increase linearly along the four proficiency levels (except for the small decrease of MLC between groups 3 and 4). Conversely, lexical diversity (D) exhibits a more irregular behaviour.

Table 6

Descriptive Statistics for Syntactic and Lexical Complexity Measures

	MLC		*C/TU*		*D*		*Beyond-2,000*	
L2 Groups	M	SD	M	SD	M	SD	M	SD
1	7.626	1.541	1.65	0.362	46.304	9.451	13.5	3.66
2	8.189	1,112	1.71	0.325	43.508	10.771	13.7	3.98
3	8.956	1.110	1.88	0.357	43.487	9.645	15.5	3.74
4	8.941	0.922	1.99	0.421	45.743	10.469	18.0	3.04

Regarding the relationship between the complexity measures, MCI does not significantly correlate with either measure of syntactic complexity. By contrast, weak positive correlations were found between MCI and the measures of lexical complexity (Table 7).

Table 7

Correlations Between Complexity Measures

	MCI	MLC	C/TU	D
MLC	-0.157			
C/TU	-0.026	-0.017		
D	0.302**	-0.190*	-0.148	
Beyond-2,000	0.234*	0.314***	0.086	0.071

Notes. N = 113; * p < .05, ** p < .01, *** p < .001

To examine the predictive contribution of MCI to the proficiency judgements in comparison to the syntactic and lexical complexity measures, we fit an ordinal logistic regression model with the proficiency level of the L2 groups as dependent variable and the five complexity measures as covariates. According to the results in Table 8, syntactic complexity (MLC and C/TU) and lexical sophistication (Beyond-2,000) measures are significant predictors of L2 proficiency in this context. Morphological diversity (MCI) is also positively associated with proficiency level, but it is a weaker predictor than syntactic complexity and lexical sophistication. Finally, lexical diversity (D) can be excluded from the model as its contribution is not significant.

Table 8

Ordinal Logistic Regression Model Coefficients

			95% CI						95% CI	
Threshold	Predictor	Estimate	Lower	Upper	SE	Z	*p*	Odds ratio	Lower	Upper
1 | 2		14.2			2.6	5.46	< .001			
2 | 3		15.7			2.68	5.88	< .001			
3 | 4		17.3			2.76	6.25	< .001			
	MLC	0.77484	0.457	1.111	0.1659	4.67	< .001	2.17	1.58	3.04
	C/TU	2.34555	1.311	3.455	0.5443	4.309	< .001	10.44	3.711	31.65
	Beyond-2,000	0.13437	0.037	0.236	0.0503	2.669	0.008	1.14	1.038	1.27
	D	0.00239	-0.04	0.041	0.0195	0.123	0.902	1	0.965	1.04
	MCI	0.52531	0.024	1.046	0.2596	2.023	0.043	1.69	1.024	2.85

Model fit measures: Deviance = 251, Nagelkerke's $R^2 = 0.252$, $\chi^2 = 60.7$, df = 5, $p < .001$

The results indicate that four of the five complexity measures used as predictor variables, MLC, C/UT, Beyond-2,000 and MCI, can explain an estimated 25% of the variance in proficiency level.

Discussion

In this section we summarize and discuss our results. First, we will deal with the results of the MCI and their relationship with proficiency and the sources of morphological diversification. Next, we will refer to the interrelation between the different complexity variables used in this study.

Morphological Diversity, Proficiency and the Sources of Inflectional Diversification

The results on RQ1 indicate that, in our corpus, the MCI scores are fairly comparable among the L2 groups, but significantly different from those of the native speakers, except for the most advanced group. These results do not confirm our H1 and suggest that there are no significant differences in the diversity of the morphological repertoire that the learners activated in their writing across the proficiency levels. The findings contrast with those in previous research on other languages reporting that MCI was able to distinguish between the proficiency level of learners until they reached a certain threshold and MCI scores became similar to those exhibited by NS (Brezina & Pallotti, 2019; De Clercq & Housen, 2019)[6]. In our data, such attainment of a native-like degree of inflectional diversity cannot explain the lack of development observed in the L2 groups, since they do not differ from each other and only the scores of the most advanced group are similar to those of the NS. The results are somewhat surprising, given the relative complexity of the Spanish verbal system and its slow and gradual acquisition. To a certain extent, the lack of inflectional development observed in our L2 groups can be due to a possible higher range of proficiency levels in our corpus in comparison to that of previous studies[7]. Nevertheless, there are other factors that can contribute to explain

the data, such as the characteristics of the corpus, the means used to determine the proficiency levels and the acquisition of the verbal inflectional system in Spanish.

First, we refer to the characteristics of our participants, who form a heterogeneous group: they do not come from the same school or university nor do they all have the same L1, as was the case in previous research on Italian and French. Only the issue relating to the L1 could, by itself, have sufficient potential to produce diverse and mixed linguistic profiles within the same communicative proficiency levels. Lu & Ai (2015) found significant differences in syntactic complexity among writers with different L1 backgrounds. It could be the same for morphological complexity, as much research supports the influence of the L1 on the acquisition of morphology in L2[8].

Second, the manner of determining the proficiency levels in our corpus – by means of human judgements – can also help to explain the results. In lexical diversity literature, there is some agreement on the difficulty of examiners to perceive diversity in comparison to other constructs such as sophistication (e.g. Malvern et al., 2004, p. 103). The MCI, based on a strictly structural definition of complexity, computes the variability of inflectional exponences, which might not be a highly perceptible construct for human evaluators. In the case of Brezina and Pallotti's (2019) and De Clercq and Housen's (2019) studies, the proficiency levels were established by means of cloze tests, classroom exposure and accuracy measures. Those instruments could group the data in a way that facilitate more than human judgements the inflectional diversity sensitivity to distinguish across groups.

However, the factors outlined so far cannot explain the divergence between our results and those of Bulté and Roothooft (2020), since their participants had heterogenous backgrounds and the levels of proficiency were subjectively determined. It is plausible

that task type and modality also modulate morphological diversity. On the one hand, the written mode could allow our lower-level participants to use compositional strategies in conjugation and, consequently, to activate a range of inflectional forms unavailable in speech for Bulté and Roothooft's (2020) participants. On the other hand, in our study, unlike in previous research, the elicitation task was integrated (i.e. test-takers integrated information from source texts). Kyle and Crossley (2016) found that measures of lexical complexity were stronger predictors of holistic scores of writing proficiency in independent than in integrated essays. Morphological complexity could also have a weak relationship with proficiency in source-based texts, as the availability of vocabulary and content from the reading texts could free up learner's resources to engage in more complex writing from the grammatical point of view.

Factors related to the acquisition of the Spanish verbal system, characterised by early manifestation of functional categories in production, could also help to explain the degree of complexity found in the less advanced groups and the low discriminatory power of the MCI in comparison to studies on other languages such as French. Cases of missing inflection or default infinitives are absent from our corpus, in contrast to the interesting interactions between default and inflected forms reported by De Clercq & Housen (2019, p. 88). Moreover, the lack of development signalled by the MCI does not necessarily mean that morphological development has come to an end in any of the proficiency levels (see also, De Clercq & Housen, 2019, p. 91). On the contrary, according to research on the acquisition of Spanish, at relatively high proficiency levels it is not inflectional diversity that evolves, but the range of meanings of the inflectional forms in a sort of dissociation between morphosyntactic and semantic knowledge. The meanings expressed with the inflectional morphemes and their multiple variations in discursive contexts are

aspects beyond the scope of a measure such as the MCI, intentionally restricted to compute the variability of the forms.

Our RQ2 inquired into the sources of morphological diversification. To ascertain the role played by irregular and suppletive forms, we computed an alternative version of MCI with collapsed irregular forms. The fact that the differences between the scores of both MCI versions decreases as the level increases suggests that a growing range of regular verbal paradigms progressively contributes to diversity along the proficiency levels. This confirms our H2 and is in line with De Clercq and Housen's (2019) findings on French. Strikingly, while in De Clercq and Housen's (2019) study the differences between scores of both measures tend to be larger at higher levels, in our corpus they tend to be smaller. This would imply that in Spanish the contribution of regular morphology is even more decisive than in French in determining the degree of inflectional diversity at more advanced levels. In fact, our results indicate that disregarding irregularities leads to MCI scores with higher discriminatory power across proficiency levels. One possible reason for this divergence is that in Spanish all the forms of the different paradigms are practically distinguishable from each other, while in oral French some of them are formally identical (De Clercq and Housen, 2019, p.77). If the available range of productive morphological markers in Spanish is wider than in French, and advanced learners use them more fluently, the MCI could behave differently in both languages when different levels of proficiency are compared. The issue needs further investigation because a more focused analysis on regular morphology, specifically on past and subjunctive forms, also suggests that in our corpus MCI scores could be, in many cases, more related to individual stylistic choices than to proficiency. Despite these limitations, the regression analysis set up to answer RQ3 showed that MCI is a significant predictor

of L2 Spanish proficiency and that is worthwhile to use inflectional diversity together with other dimensions of complexity to asses proficiency.

The Morphological Complexity Index and Syntactic and Lexical Complexity

Our H3 is only partially confirmed by the results on RQ3. They coincide, to a certain extent, with previous research reporting correlations between morphological and lexical complexity measures (Brezina & Pallotti, 2019; Bulté & Roothooft, 2020), but differ on the absence of correlation with syntactic complexity. The last finding is unexpected, since one of the syntactic measures, C/TU, is theoretically related to the acquisition of the subjunctive and, therefore, to inflectional diversification. However, there are texts in the corpus with very high C/TU scores and complete absence of subjunctive forms. In fact, the percentage of subjunctive forms used in the texts does not correlate with C/TU. The subjunctive appears primarily in subordinate clauses, but many types of subordinate clauses do not need the subjunctive to be correct in Spanish. The use of subjunctives in this corpus not only seems to be independent of proficiency but also of complex syntax.

With respect to the correlation between lexical and morphological complexity, it is worth bearing in mind that exponences were extracted in a way that reduced lexical diversity effects on the MCI calculation i.e. SVs were considered as part of the LBs (see, De Clercq & Housen, 2019, p. 92, for the effects of considering SVs as part of the inflectional exponents). Moreover, lexical measures were computed from lemmas, thus they were not affected by morphological diversity. The correlation could be partly explained by the existence of a link between lexical knowledge and emergence of grammatical forms since better lexical access, also facilitated by the task type, could free

up resources to make available a broader repertoire of inflectional forms. Trade-offs between the linguistic subsystems should be more carefully examined in order to explain these relationships between different complexity metrics and subconstructs.

Finally, as in Bulté & Roothooft (2020), MCI makes a significant contribution to explaining the variance in proficiency, although it is a weaker predictor in comparison to syntactic complexity and lexical sophistication. Conversely, lexical diversity is not associated with proficiency in these texts. Thus, syntactic complexity and lexical sophistication correspond more closely to proficiency than lexical and morphological diversity, suggesting again that structural diversity, despite of its methodological advantages (Pallotti, 2015b), could be a construct that covers proficiency progress in this context less effectively than other dimensions of complexity.

Limitations and Future Perspectives

Several limitations of the present study must be acknowledged. First, the NS samples used as a baseline were not collected under examination conditions, as the L2 texts were. Time pressure and other typical circumstances of test situations may have conditioned the L2 texts and not those of the NS. Second, the test-takers' texts were elicited from two different exam forms. Although the text type was kept constant, there might be topic effects that have gone unnoticed. Lastly, the relatively high proficiency level of our corpus, combined with the integrated nature of the elicitation task and the modality, could undermine the sensitivity of MCI to detect differences among proficiency levels. Examining the development of inflectional diversity in lower proficiency levels, in speech and in independent tasks may prove informative for future research in L2 Spanish.

Conclusions

This study examined inflectional diversity in L2 Spanish written texts and its relationship with proficiency and other dimensions of linguistic complexity. As the proficiency levels were determined by subjective ratings of writing quality, our results contribute to the interpretation of morphological complexity measures in the light of communicative adequacy. Inflectional diversity was measured by means of the MCI, which along with one measure of lexical sophistication (Beyond-2,000) and two measures of syntactic complexity (MLC and C/TU) explained an estimated 25% of the variance in proficiency level. However, MCI scores indicate that lower-level written discourses in Spanish can exhibit wide repertoires of inflectional resources, comparable to those of texts rated at higher levels of proficiency. This finding contrasts with previous studies on other languages and with the fact that the Spanish verbal system is considered complex and challenging for L2 learners, who acquired it slowly, gradually and with difficulty. The finding can be better understood by taking into account a constellation of factors such as the overall relatively high proficiency level of the corpus, the heterogenous background of the participants and their individual stylistic choices or the written and integrated nature of the elicitation tasks. Moreover, the structural diversity construct measured by MCI can be a construct with a weak relationship with rated proficiency. It might be hardly perceived by examiners, who could pay more attention to the adequacy and complexity of the meanings expressed by the morphological markers.

Notes

1. Original samples of CELU examinations can be found at http://www.celu.edu.ar/?q=es/node/25

2. It would be interesting to analyze the analytical ratings in order to observe their relationship with the holistic scale and the MCI data, especially in the case of the morphosyntactic scale. Unfortunately, this information is not available as we only received the texts anonymized and classified into levels of proficiency.

3. The command used to calculate D was *vocd +sm;*,|*,#+*,o% -sm|neo,|n:prop*. The switch *+sm* includes in the calculation any lexical root (*;**), word class information (*|**) and any derivational suffix if present (*#+**), and ignores any other information (*o%*). The switch *-sm* excludes non target language items (*|neo*) and proper names (*|n:prop*). As MOR grammar annotates Spanish verbs as if they belonged to four different word classes (*v|*, *inf|*, *imp|*, *part|*), before calculating D, we used the *chstring* command on the MOR line to change all of them to *v|*.

4. According to Davies & Hayward Davies (2018), the frequency dictionary is based on two different corpora, both browsable in www.corpusdelespanol.org. One corpus contains 20 million words from the 1900s portion of the genre-oriented *Corpus del Español*. Two thirds of the corpus come from the written register and one third from spoken Spanish. The second corpus is a 20-million-word corpus taken from web pages in 2013-2014.

5. We are aware that in the learners' texts subjunctive and past tense forms could be only partial and unstable representations of the target grammar paradigms, but, even so, they contribute to diversity in the way it is computed by the MCI.

6. In order to verify whether our decision of analyzing the SV as part of the lexical base was not a confounding factor when comparing our results with those of previous studies, we carried out an analysis of the learners' texts using the same procedure as in previous studies i.e. analyzing the SV as part of the inflectional

exponent. As expected, global diversity was higher, but overall results were similar to those presented in this article (MCI mean (SD): group 1: 7.203 (0.676), group 2: 7.361 (0.562), group 3: 7.310 (0.536) and group 4: 7.471 (0.492)). Interestingly, compared to the analysis presented in this article, the relationship between MCI and proficiency was slightly weaker and the results contrasted even more with those of previous studies.

7. The comparability of proficiency levels in different studies is a sensitive issue. In our corpus, the texts graded as intermediate to advanced (groups 2-4) can be broadly regarded as levels CEFR B2-C1. As the basic level is a non-certified level in the CELU system, the texts belonging to group 1 cannot be referenced with certainty, but, presumably, they belong to a level below B2 (e.g. A2/B1). Consequently, the level of our participants is probably higher than that of Brezina and Pallotti's participants in their study on Italian (CEFR A2-B2) and comparable to that of Bulté and Roothooft's participants (CEFR A2/B1-C1).

8. We carried out a tentative analysis to explore the influence of the participants' L1 on the results. As there were many L1 among the participants, we made two groups: speakers of a Germanic language (N = 66) and speakers of a Romance language (N = 20). We conducted a factorial ANOVA, with MCI as dependent variable, and level and L1 group as fixed factors. The test revealed no significant main effect of any of the variables, nor significant interaction between them. However, these results do not necessarily rule out the influence of the L1 background on the linguistic profile of the texts.

References

Alcoba, S. (1999). La flexión verbal. In V. Demonte & I. Bosque (Eds.), *Gramática descriptiva de la lengua española (vol. 3). Entre la oración y el discurso. Morfología* (pp. 4915–4992). Espasa Calpe.

American Council on the Teaching of Foreign Languages. (2012). *ACTFL proficiency guidelines*. http://www.actfl.org/

Bermúdez-Otero, R. (2013). The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features. *Probus*, *25*(1), 3–103. https://doi.org/10.1515/probus-2013-0009

Brezina, V., & Pallotti, G. (2015). *Morphological complexity tool*. http://corpora.lancs.ac.uk/vocab/analyse_morph.php

Brezina, V., & Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second Language Research*, *35*(1), 99–119. https://doi.org/10.1177/0267658316643125

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins Publishing Company.

Bulté, B., & Roothooft, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System*, *91*, 102246. https://doi.org/10.1016/j.system.2020.102246

Collentine, J. (1995). The development of complex syntax and the selection of mood by foreign language learners of Spanish. *Hispania*, *78*(1), 122–135.

Collentine, J. (2010). The Acquisition and Teaching of the Spanish Subjunctive: An Update on Current Findings. *Hispania*, *93*(1), 39–51.

Consorcio Español Lengua Segunda o Extranjera (CELSE). (n.d.). *CELU: Certificado de Español Lengua y Uso*. Retrieved February 10, 2020, from https://www.celu.edu.ar

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. https://www.coe.int/en/web/language-policy/cefr

Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Daller, H., & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 150–164). Cambridge University Press.

Davies, M., & Hayward Davies, K. (2018). *A Frequency Dictionary of Spanish* (2nd ed.). Taylor & Francis.

De Clercq, B., & Housen, A. (2019). The development of morphological complexity: A cross-linguistic study of L2 French and English. *Second Language Research*, *35*(1), 71–97. https://doi.org/10.1177/0267658316674506

Kusters, W. (2008). Complexity in linguistic theory, language learning and language change. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 23–41). John Benjamins, Amsterdam/Philadelphia

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, *34*, 12–24. https://doi.org/10.1016/j.jslw.2016.10.003

Lan, G., Liu, Q., & Staples, S. (2019). Grammatical complexity: 'What Does It Mean' and 'So What' for L2 writing classrooms? *Journal of Second Language Writing, 46*, 100673. https://doi.org/10.1016/j.jslw.2019.100673

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, *29*, 16–27. https://doi.org/10.1016/j.jslw.2015.06.003

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3ª). Mahwah, NJ: Lawrence Erlbaum Associates.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230511804

Miestamo, M. (2008). Grammatical complexity in a cross-linguistic perspective. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 23–41). John Benjamins Amsterdam/Philadelphia

Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*. John Benjamins Publishing.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Olbertz, H. (1998). *Verbal periphrases in a functional grammar of Spanish*. De Gruyter Mouton. https://doi.org/10.1515/9783110820881

Pallotti, G. (2009). CAF: Defining, Refining and Differentiating Constructs. *Applied Linguistics*, *30*(4), 590–601. https://doi.org/10.1093/applin/amp045

Pallotti, G. (2015a). Una nuova misura della complessità linguistica: l'Indice di Complessità Morfologica (ICM). *Rassegna Italiana Di Linguistica Applicata (RILA)*, *2*, 195–218.

Pallotti, G. (2015b). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117–134. https://doi.org/10.1177/0267658314536435

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10–27. https://doi.org/10.1016/j.jslw.2014.09.003

Presson, N., Sagarra, N., MacWhinney, B., & KowaIsky, J. (2013). Compositional production in Spanish second language conjugation. *Bilingualism: Language and Cognition*, *16*(4), 808–828. https://doi.org/10.1017/S136672891200065X

Real Academia Española. (2008). *Corpus de Referencia del Español Actual (CREA) - versión 3.2*. http://www.rae.es

Roca, I. (2010). Theme vowel allomorphy in Spanish verb inflection: An autosegmental optimality account. *Lingua*, *120*(2), 408–434. https://doi.org/10.1016/j.lingua.2009.05.005

Slabakova, R. (2014). The bottleneck of second language acquisition. *Foreign Language Teaching and Research*, *46*(4), 543–559.

The jamovi project. (2020). *jamovi (version 1.2) [Computer Software]*. https://www.jamovi.org

ITL-20009 (ITL - International Journal of Applied Linguistics)
"Morphological Complexity and Rated Writing Proficiency: The Case of Verbal Inflectional Diversity in L2 Spanish"

Editor's suggestions based on the reviews	**Changes made in acceptance of suggestions**
It is absolutely crucial to address reviewer 2's comments about the computation of MCI, as this was insufficiently addressed in the revised manuscript	Revised, see footnote 6 on page 33
The literature review would benefit from a more in-depth discussion of morphological complexity	Included, see pages 2-3
The discussion should be more strongly linked to your literature review too. Please, make sure you refer back to the theories/studies discussed in the Literature reveiw and do not introduce any new theoretical concepts in the Discussion	We linked more explicitly our RQ and discussion to our literature review. We omitted in the Discussion a number of theoretical concepts that were not previously introduced.
I would also like to recommend you have another look at the language and organization of the manuscript. For instance, the discussion could be more concise	Revised
Reviewer#1 suggestions	**Changes made in acceptance of suggestions**
The language could still be worked on a little more (e.g. in the abstract, where logical connectors could be used to establish more coherence between sentences). The introduction is improved too, although I feel the text could also be condensed more and more directly to the point.	Abstract and Introduction revised revised.
Important additions have been made to the literature review, yet there are still a number of topics that aren't addressed. For example, I would appreciate a short theoretical discussion of morphological complexity (not necessarily focusing on the measure, but on the construct). Complexity is discussed in general and morphological complexity is mostly discussed in relation to the MCI, yet the construct of morphological complexity is not really elaborated on currently from a theoretical perspective.	Added on pages 2-3
P.7 specify which syntactic measures MCI correlated with in Bulté & Roothooft's study.	Added on page 7
I appreciate the inclusion of a more detailed section on Spanish TAM (morphology) acquisition as well, but there are two further points of improvement. (1) The discussion at times could be more concise, for example, as it treats the indicative/subjunctive or the preterite/imperfect contrasts. These are certainly relevant, but perhaps they could be synthesized	We are now more concise in the discussion on Spanish morphology acquisition. We are also more explicit in relating that research with our hypothesis and with the morphological complexity approach. Pages 9-12.

more to serve the function of the article, which brings me to the second point. (2) The discussion now occurs mostly independently from the notion of complexity and the literature introduced here is not used to inform the hypotheses. Both in the literature review and in the methodology, an explicit attempt should be made to reinterpret previous findings in light of the complexity methodology that is proposed, as this will lead to more substantiated hypotheses that bridge the gap between morphological complexity research and more traditional TAM research.	
To be absolutely certain that the procedure in CLAN was performed correctly, could you clarify if the code for calculating D excluded all morphological information? (This is not done automatically, if I'm not mistaken.) In addition, was word class information maintained when calculating D? This could lead to more accurate results in cases of homographs belonging to different word classes.	It excluded all morphological information and included word class information. Details on the VOCD command we used have been added on footnote 3 on page 33.
With regard to the treatment of auxiliaries as one operation, this implies that "había trabajado" is treated as a single unit (as made clear in Table 2), right? This is a departure from the other studies you cite, no? Could you elaborate on this decision and its consequences?	It is right, compound forms were treated as a single operation. However, we are not sure that this is a departure from the cited studies. Pallotti (2015, A simple view of linguistic complexity. *Second Language Research*, 31,1) writes on page 121: "Periphrastic morphemes, like *be V-ing* or *have V-ed* in English, will be counted as one single operation". There is also an example of a compound form treated as a single operation on page 202 in Pallotti (2015) (Una nuova misura della complessità linguistica: l'Indice di Complessità Morfologica (ICM). Rassegna Italiana Di Linguistica Applicata (RILA), 2). Unfortunately, we could not find specific information about this issue in Brezina & Pallotti (2019), De Clercq & Housen (2019) nor Bulté & Roothooft (2020). That is why we followed Pallotti (2015). In our corpus there are 130 compound forms, on average 1.14 forms per text. Their frequency, on average, is 0.03. We think that a different analysis would not have had a significant impact on our results.
Could you very briefly characterize the corpus used to establish the frequency dictionary of Spanish in the methodology (e.g. frequencies in oral vs. written data)?	Added information on page 33, footnote 4.

The procedure and motivation for labelling some suffixes as "less frequent" is not entirely clear and potentially circular. At the very least, it is not derived from the hypotheses. Formal aspects (regular vs. irregular) seem to be conflated with aspects that are perhaps influenced by difficulty, but are at the very least "relative" (frequent vs. infrequent). Note that the separate analysis for irregular verbs seems more relevant from the "absolute" perspective of the MCI. It seems to me that the question of which sources contribute to morphological diversity/diversification should be made on theoretical grounds – for example, based on the TAM literature that is introduced.	Our hypothesis on RQ2 was that Spanish will have a similar development to that observed in French by De Clercq and Housen (2019) i.e. inflectional diversity will increase primarily as a result of the contribution of a progressively more extensive range of regular forms. With this analysis we tried to find out whether higher-level texts contained a greater diversity of inflectional suffixes, beyond the present and infinitive suffixes that flood the entire corpus. It seems that the data suggest that this is the case: as the level becomes higher, the diversity of inflectional suffixes increases (corresponding to different verb tenses), and irregular and suppletive forms decrease. However, we acknowledge that labelling some suffixes as "less frequent" is problematic from the "absolute" perspective of the MCI. We therefore decided to remove that analysis and stick to the separate analysis of regular and irregular forms, which is more relevant from an "absolute" perspective. We report the results obtained using the same procedure as De Clercq and Housen, 2019 i.e. computing the MCI after collapsing all the irregular forms in a single category. The regular forms that contribute to increase the morphological diversity are difficult to examine from a purely structural perspective of complexity and without relating them to morphosyntactic paradigms established in the target language grammar. We have dared to include an exploratory analysis of past and subjunctive forms that we carried out using the native speakers' data See pages 22-23 and footnote 5 on page 33
Table 5 is discussed fairly superficially, with the text only mentioning that over 80% of the variance is explained by the two independent variables, but the impact of the variables themselves is not addressed. Note too that, since proficiency level wasn't included in the model, this analysis doesn't say anything about how the proportion of less frequent or irregular suffixes is	Removed as it was including the category of "less frequent suffixes".

related to the proficiency levels. In consequence, this analysis can't really answer RQ2. The statistical test is to be changed to include proficiency levels in order for this test to become relevant – I do believe principle of the regression analysis can be maintained though.	
Figure 1 is perhaps more relevant to answer RQ1 (disregarding the choice to distinguish "less frequent suffixes" for a moment), but even better would be to represent all sources of diversification in this table (as a bar graph showing the proportion of each category), no?	Removed as it was including the category of "less frequent suffixes". We include now a bar plot representing past tenses and subjunctive mood that could be more acceptable from a theoretical point of view.
Could the text also include the relevant statistical information, e.g. when describing how strongly each variable predicts MCI? In particular, could the standardized coefficients be discussed more explicitly? Since the different complexity measures use different scales, standardized coefficients are crucial to interpret your findings correctly.	Removed as it was including the category of "less frequent suffixes".
As a general point, I wonder to what extent the fact that ceiling levels may conceal differences in the source of complexification implies that short production task may be less suited to accurately represent the morphological diversity of these learners' interlanguage. Perhaps different types of data are required in order to get a better idea of the morphological inventory of learners, where data from different genres is pooled together for the analysis of each learner. If this is the case, the fact that you included two different texts per learner is not necessarily a downside (although the genre is similar in both texts). Perhaps the segment size setting of the MCI also has an impact on the representation of morphological complexity.	Perhaps our text is not sufficiently clear about this issue. We did not include two different texts per learner. We had 113 texts from 113 different learners. We clarify better this point in the text on page 13.
Finally, I would also appreciate more examples in your discussion, especially since these quantitative measures (representing the number of different exponents) seem to conceal qualitative differences (representing the type of exponents) in your data (as is evidenced by Figure 1). The same is true for your observation that an increase in C/TU would be accompanied by an increase in the use of the subjunctive. What does your data look like here? Do subjunctives appear more frequently here? Are subjunctives sometimes formally identical to indicative forms in Spanish (cf. French subjunctive "il faut que je pense" = indicative "je pense")? It seems to me like even more can be done with the data in the discussion.	We are afraid that meaningful examples are too long to be included in the text as they require almost a complete essay to be reproduced. We provide one text as an example that we think is representative of the diversity that can be found in the lowest-level group. However, we add now some qualitative observations that can help to understand the quantitative results (page 23) Regarding the subjunctive and the C/TU, they seem to be independent in our corpus, despite the possible theoretical relationship we refer to in the article. We include this point in the Discussion (page 29)

	Subjunctives are distinguishable from indicatives in oral and written Spanish. They were also distinguishable in the codes used to compute the MCI. Imperfect-subjunctive always have a *-ra* or *-se* marker, which is not shared by any other tempus. The present-subjunctive could be not distinguishable from the present-indicative if we had calculated the stem vowel as part of the inflectional exponent. In that case a form as *canta* (3th person singular present indicative: *cantar*) would be codified as T_n-a and a form as *coma* (3th person singular present subjuntive: *comer*) would also be codified as T_n-a. They would have not been distinguishable in the MCI calculation. But in Spanish there are theoretical basis to include the stem vowel in the lexical base and not in the exponence and thus we could preserve the distinction in the MCI calculation codifying those forms as T_n-0 (*canta*) and T_n-a (*coma*). If a form as *come* and a form as *com-a* coincide in the same text the different lexical bases add to diversity using different codes: T1-0 (*come*) and T2-a (*coma*).
P.2: "complexity-accuracy-fluency approach" => "Complexity Accuracy and Fluency (CAF) approach". There are also a number of article mistakes on these pages: "It has been defined in at least two ways: the relative complexity or difficulty and the structural or absolute complexity" (and elsewhere) - P.10: "much of research" => much/most research - P. 11, "with very few exceptions, subjunctive is restricted" => the subjunctive is restricted - P. 18: "a stem alteration that produce" => produces - P. 27: "as it is well attested… L2 morphology" – this sentence isn't very clear to me. - P. 28: "which might be a not very perceptible" => which might not be a highly perceptible - P.30: "the later increases" => the latter	Revised
Reviewer#2 suggestions	**Changes made in acceptance of suggestions**
Nevertheless, there is still one point that strikes me: the authors have operationalized the MCI differently than what Brezina and Pallotti did. In	We completely agree: the particular identification and extraction of the

particular, at the core of the MCI lies the way a researcher identifies and extracts inflectional exponences. Brezina and Pallotti, for Italian and English and De Clercq and Housen for French, have identified the exponences the following way: root (pure lexical meaning, outside the exponence) and exponence composed by [Stem Vowel + TAM morpheme + PersonNumber morpheme]. This implies that, in the verb "cantabamos" we have: cant [root] + abamos [exponence]. The authors decided to calculate the MCI differently, as they explain at page 17: they decided to calculate the SVs as part of the root, in a single unit that they call "verb lexical bases". The exponency computation and extraction is therefore as follows: root, outside the exponence [lexical root + SV] and exponence composed by [TAM morpheme + PersonNumber morpheme]. The verb "cantabamos" is analyzed as follows: canta [root] + bamos [exponence]. The authors acknowledge that this operationalization of MCI reduces the inflectional diversity of the texts (p. 17: "a notable consequence of this decision is that global inflectional diversity is reduced"). The authors acknowledge, for instance, that if their MCI version is followed, the inflectional properties of the Spanish infinitive are lost, : for instance, hablar, comer, vivir would be recognized, by Brezina and Pallotti's MCI version as bearing 3 inflectional types: hab -er; com-er; viv-ir. According to the authors, there would be only one inflectional type in these 3 verbs, namely -r: habe-r;come-r; vivi-r. I would add here that also other inflected forms would be lost, according to the authors operationalization of MCI: the only verb "cantar" has two important and frequent forms tha would not be aclculated as bearing an exponent by the tool: "canta", which is both third form (el cant-a, regular for third person singular present) and imperative (canta tu!, regular for this mood). How would these two forms be calculated, by the authors' version of MCI? As bearing a zero morphological functor? And in the case of the second person at the present, how would this be calculated? The I and II conjugations would only add an -s to the verb lexical base, but the III conj. regular verb would change the SV? Es: dormir [dormi --> lexical base] + r [exponence], but [duerme -->lexical base] + s [exponence]?	inflectional exponents are crucial for the calculation of the MCI. (1) In a form as *canta,* the inflectional exponent for the MCI calculation is -0 (see also a similar example in page 19 in our article). It must be noticed that the alternative analysis (to calculate the SV inside the inflectional exponent) would treat the third person singular present indicative of the class -*ar* verbs (I conjugation) as *canta* (T_n-a) in the same way as the third person singular present subjunctive of the class -e*r* verbs (II conjugation) as *coma* (T_n-a). In our analysis, the exponent for the first form *canta* is -0 and for the second one *coma* is -*a*, where *com-* is an allomorphic lexical base (*come*-0, but *com*-a). We believe that, in this sense, our analysis suits better than the other one the morphological processes underlying the forms of the Spanish verbs. It is true that with our analisys a form as *canta* (present) is treated in the same way as *canta!* (imperative): T_n-0. Imperatives are present in our corpus but they are very scarce. (2) As for the second person at the present of indicative like *vives*, from *vivir*, our analysis considers it an allomorphic lexical base (as duerme- from *duermes*, dormi- from *dormir* and duerm- from *duermas*). The form *vivir* would be calculated in the MCI as T_1-r and *vives* as T_2-s; *comes* would be T_1-s, *comas* T_2-as, *comieron* as T_3-ron. This is the way we used to reflect the diversity due to transformations in the SV. In our analysis they are allomorphic themes (roots+SV, our lexical bases). Previously to the analysis we describe in this article, we calculated the MCI in the learners' texts following the Brezina and Pallotti and De Clercq and Housen procedure, i.e. analyzing the SV as part of the inflectional exponent. The results were similar to the ones we present in the article. The MCI mean (SD) scores were

As RQ1 is answered bu comparing the MCI analysis of this paper to the ones of Brezina & Pallotti and De Clercq & Housen, and as a strong divergence is found, it would be interesting to see if a different computation of the MCI (i.e. an exact replication of Brezina, Pallotti, De Clercq and Housen) would give more convergent results between the 3 papers or if the divergence still remains. In other words: how much of the papers' divergence is(can be explained by the different operationalization of the MCI made by the authors? As a reviewer, I feel that the authors have widely analyzed their results, acknowledging and highligting the "weak" points. The different operationalization that I pointed above remains, nevertheless, undiscussed. It is hard for me to calculate how much difference it can bring in the results, but it would be intersting to see if exactly the same operationalization of the MCI had been used, how the results would have changed. The other papers' authors (i.e. Pallotti, Hosen etc.) might, in future, stress the difference bewteen their MCI and authors' one, and that's probably all fine. Just decide if it is important for you and for the jourbnal to ask the authors to add what I have so far suggested (if my argumentation has been clear and convincing enough, of course).	as follows: group 1 (basic): 7.203 (0.676), group 2 (Intermediate): 7.361 (0.562), group 3 (Upper-Intermediate): 7.310 (0.536) and group 4 (Advanced): 7.471 (0.492). There was not any significant correlation between the level and the MCI scores (Spearman's *ro* 0.125, $p = 0.188$). The results were surprising as they diverged such a lot from previous studies. Thus, we hypothesized that the diversity due to the SV, determined by the root, was overshadowing the diversity due to strictly inflectional constituents as TAM an PN morphemes. We also found theoretical basis to carry out an analysis where the SVs were analyzed inside the lexical bases. The results of the second analysis, the one we present in the article, still diverge from those in previous studies, but we decided to report the second analysis because the relationship between MCI and proficiency is a bit stronger and, in this sense, closer to the results of previous studies. We include a note about this point. See footnote 6 on pages 33-34.