

Phraseological complexity in EFL learners' spoken production across proficiency levels

Magali Paquot¹, Dana Gablasova², Vaclav Brezina²
and Hubert Naets¹

¹UC-Louvain / ²Lancaster University

This study explores phraseological complexity in English as a Foreign Language (EFL) learners' spoken production across proficiency levels in the *Trinity Lancaster Corpus*. Phraseological diversity and sophistication are operationalized as root type-token ratios and median mutual information scores of verb + object co-occurrences respectively. Results draw a complex picture of phraseological complexity in EFL learners' oral performance, with phraseological diversity increasing on the whole and phraseological sophistication decreasing significantly from B1 to B2. These findings can at least partly be explained by the fact that, unlike EFL learners at B1, EFL learners at B2 and above repeat fewer combinations and use more specific vocabulary, which sometimes leads to less idiomatic combinations that should nevertheless be considered as traces of qualitative development.

Keywords: phraseological complexity, EFL learner speech, proficiency, pointwise mutual information

1. Introduction

Studies on linguistic complexity, in both writing and speaking, have traditionally defined and measured complexity in terms of lexical and syntactic features (e.g. Bulté & Housen 2012). However, as Paquot (2019) recently argued, L2 complexity research needs to broaden its scope to account for the fact that language is highly formulaic in nature and words naturally combine to form conventional patterns of meaning and use (e.g. Sinclair 1991; Hanks 2013), a phenomenon at the lexico-grammatical interface that is not reflected and captured by traditional measures of syntactic and lexical complexity. Drawing on Ortega (2003: 492), who emphasised the role of diversity and sophistication in linguistic complexity, Paquot

(2019: 124) identified “the range of phraseological units that surface in language production and the degree of sophistication of such phraseological units” as major dimensions of the construct of phraseological complexity. Following this conceptualization, a learner text with a wider diversity of (target-like) phraseological units and a higher proportion of sophisticated units would be considered to be more complex than one where the same set of (more frequent) word combinations is often repeated.

Only a few studies, so far, have investigated foreign language learners’ phraseological competence in terms of the diversity and sophistication of the phraseological units used in L2 production. While phraseological units can be of a variety of types, including collocations, collocational frameworks and lexical bundles (Granger & Paquot 2008), the majority of studies have focused on statistical collocations or co-occurrences. In Paquot (2019), phraseological complexity was approached via a relational or syntactic model of co-occurrences to investigate English as a Foreign Language (EFL) learners’ use of adjective modifiers, adverb modifiers and verb + noun collocations at the B2, C1 and C2 proficiency levels of the *Common European Framework of References for Languages* (CEFR; Council of Europe 2001) in the *Varieties of English for Specific Purposes dAtabase* (VESPA) learner corpus (Paquot et al. 2013). The corpus was POS-tagged and parsed with *Stanford CoreNLP* and collocations were extracted on the basis of *amod* (adjective modifier), *advmod* (adverb modifier) and *dobj* (direct object) Stanford typed dependencies (De Marneffe & Manning 2003). Phraseological diversity, the first dimension of phraseological complexity, was operationalized as a root type–token ratio (RTTR) computed for each dependency relation. Phraseological sophistication was operationalized in two different ways. First, in the same way as sophisticated single words have sometimes been equated with university-level vocabulary (e.g. Laufer 1994; Laufer & Nation 1995), sophisticated word combinations were defined as academic collocations and operationalized as word pairs that appear in the *Academic Collocation List* (ACL; Ackermann & Chen 2013). Measures of phraseological sophistication were then computed as ratios of the number of sophisticated *amod*, *advmod* and *dobj* types/tokens to the total number of *amod*, *advmod* and *dobj* types/tokens respectively. Second, phraseological sophistication was also operationalized as the average point-wise mutual information (MI) score for *amod*, *advmod* and *dobj* dependencies since this association measure promotes word pairs that consist of relatively less frequent words which also tend to be more exclusively combined. Results showed that:

1. No measure of phraseological diversity displayed a linear increase from B2 to C2. The RTTR for *dobj* dependencies even displayed a slight decrease from B2 to C1 and C2, but the difference between groups was not statistically significant.

2. Measures of phraseological sophistication based on the ACL showed a linear increase from the lower to the higher proficiency levels but no differences were statistically significant.
3. MI-based measures of phraseological sophistication showed significant differences across proficiency levels:
 - a. Adjective + noun collocations showed a significant difference in mean MI scores between the B2 and C2 levels but differences were not large enough to distinguish between adjacent levels such as B2–C1 and C1–C2.
 - b. Adverbial modifiers (i.e. adverb + adverb, adverb + adjective, adverb + verb) singled out upper intermediate (i.e. B2) learner writing from the more advanced (i.e. C1 and C2) learner productions.
 - c. Verb + direct object structures were the best discriminators of the most advanced (C2) level: mean MI scores for *dobj* dependencies set C2 texts apart from B2 and C1 texts.
4. No statistically significant difference was found between learner groups when their texts were analyzed with traditional measures of syntactic or lexical complexity.

The discriminatory power of MI-based measures of phraseological complexity computed on the basis of relational collocations was further demonstrated in Paquot (2018) where a mixed-effect model was used to assess the effect of syntactic, lexical and phraseological complexity on human raters' overall judgment of writing quality in the VESPA corpus. After applying a stepwise model selection, the final model only included two MI-based phraseological measures, i.e. mean MI score for *dobj* dependencies and mean MI score for *amod* dependencies, as fixed effects (marginal $R^2 = 0.25$), thus revealing that the higher the average MI scores for *dobj* and *amod* dependencies in a student's paper, the higher it was rated on the CEFR scale.

Cross-sectional studies of phraseological complexity in L2 Dutch and L2 French recently confirmed the prominent role of measures of phraseological diversity and phraseological sophistication in the prediction of foreign language writing competence or success (Rubin, Housen & Paquot 2021; Vandeweerd, Housen & Paquot 2021), with phraseological diversity being a stronger predictor in these studies compared to Paquot (2018, 2019). Altogether, the studies above have demonstrated the usefulness of conceptualizing learners' phraseological competence in terms of complexity, with diversity and sophistication as its two main dimensions. They have also piloted and evaluated the usefulness of different approaches to operationalizing the measures of phraseological complexity (e.g. the MI score) that can be used in further studies on this phenomenon.

To the best of our knowledge, no (corpus-based) study so far, however, has examined phraseological complexity in spoken learner language.¹ In fact, studies that have investigated learners' use of collocations in speech are also quite rare. Biber and Gray (2013) examined frequent collocations with five semantically light verbs (*get, give, have, make* and *take*) in the speaking tasks of the TOEFL iBT® Test. They found that collocations were much more likely to be used in integrated tasks than in independent tasks, suggesting that “the extra planning time associated with the integrated tasks permits the recall and use of such collocational combinations” (Biber & Gray 2013: 31). Results also showed that low-proficiency (Level 1) candidates used very few collocations and that, in the integrated task, intermediate-proficiency (Levels 2 and 3) candidates relied on collocations to a greater extent than the most proficient candidates (Level 4) did. Kyle and Crossley (2015) made use of a wide range of measures of lexical and phraseological sophistication to predict holistic speaking proficiency scores assigned to independent tasks from the TOEFL public use dataset. No indices of bigram frequency were kept in the final model, which included five measures related to trigram frequency, content word range, academic vocabulary, word familiarity and content word frequency. In an exploratory factor analysis that examined the multidimensional nature of lexical sophistication in oral proficiency interviews (OPI) by Japanese learners of English, Eguchi and Kyle (2020) identified two phraseology factors: a factor including measures that involve strongly associated but common bigrams (e.g. bigram range, bigram association strength computed with t-score) and a factor promoting measures of ‘exclusive collocations’ (bigram and trigram association strength (MI)). A multiple linear regression was then used to investigate the relationship between the 10 lexical sophistication factors and holistic OPI scores. The two phraseology factors explained 3.1% and 1.6% of the variance respectively in the final model, with factors representing (1) abstract but common content words and (2) concrete function words with frequent neighbours contributing the most. Finally, Brezina and Fox (2021) examined the effect of L2 proficiency on adjective + noun collocations in L2 English spoken production from the *Trinity Lancaster Corpus*. Using measures of collocational frequency and collocational association, the study found that proficiency explained some variation in the target collocations by L2 speakers (e.g. a difference between B1 and C1/C2 level), suggesting that higher levels of proficiency result in greater use of more relatively frequent and exclusive adjective + noun collocations

1. Xu (2018) proposed to measure spoken collocational competence along three dimensions, i.e. accuracy, complexity and fluency. Collocational complexity, however, is conceptualized in a very different way from this study, with its two main dimensions being transparency and restrictedness, two criteria that are at the core of the definition of collocations in the traditional approach to phraseology (Granger & Paquot 2008).

as identified by logDice scores. However, the authors also stressed that L2 speakers at higher levels of proficiency do not systematically produce more formulaic word combinations and that “the evidence points to a more complex picture of non-linear development of phraseological competence, where individual differences between speakers play a major role” (p. 171).

The main objective of this study is therefore to start filling a research gap and explore EFL learners' use of collocations in speech in terms of phraseological complexity. Speech and writing are two fundamentally different modes of communication in terms of the cognitive demands on online processing of language (e.g. Kormos 2014), with interactive communication placing particularly high demands on the ability to process and produce language with limited opportunity to monitor and edit the production. This is likely to affect the phraseological patterns in these two modalities, with speech involving a higher reliance on automatized chunks during interaction (e.g. Biber et al. 1999). In this study of phraseological complexity in L2 spoken production, the focus is placed on verb + object structures as, to the best of our knowledge, they have never been investigated in studies that have adopted a frequency-based approach to phraseology to explore collocational use in learner spoken corpora. Investigating the use of verb-object collocations in L2 spoken production also seems particularly pertinent as they are considered a major stumbling block for EFL learners (Nesselhauf 2005; Laufer & Waldman 2011). The study addresses the following research questions:

1. To what extent does phraseological diversity increase with proficiency in EFL learners' oral performance?
2. To what extent does phraseological sophistication increase with proficiency in EFL learners' oral performance?

Building on the current state of the art, we formulate two main hypotheses:

1. Phraseological diversity will increase with proficiency in EFL learners' oral performance, reaching a 'plateau' at upper-intermediate level (cf. Biber & Gray 2013).
2. Given that collocational strength (as measured by MI scores) for V+O has been reported to increase with proficiency in EFL learner writing but was also shown to only increase significantly at the C2 level (Paquot 2019), and given that writing has been found to give rise to more linguistic complexity (Ellis & Yuan 2005; Vasylets et al. 2017), it is expected that phraseological sophistication will also increase with proficiency in EFL learners' oral performance but at a slow(er) pace (than has been reported for writing).

2. Method

2.1 Learner data

The learner data come from the *Trinity Lancaster Corpus* (TLC; Gablasova et al. 2019), which consists of transcribed recordings from the *Graded Examinations in Spoken English* (GESE) developed and delivered by Trinity College London, a major international examination board. The corpus includes transcribed interactions between examiners (L1 speakers of English) and GESE candidates, L2 learners of English with B1 to C2 proficiency levels on the CEFR scale in a variety of countries such as China, India, Italy, Mexico, Russia and Spain. To ensure representativeness of L2 speakers’ language competence at each proficiency level, the majority of exams selected for the corpus were sampled from L2 speakers in the middle range of the grade/proficiency band as determined by the marks awarded for each speaking task (Gablasova et al. 2019: 139–140).

While our initial aim was to use the full TLC, a careful examination of the corpus composition showed that there was a need to restrict the number of confounding variables and work with a more homogeneous dataset to answer our research questions. For the purpose of the current study, the Italian and Spanish subsets of the corpus were selected for analysis, as they contained sufficient numbers of speakers at each CEFR level relevant to the study (i.e. B1, B2, C1 and C2). These subsets were also considered to provide a balanced sample of L2 speakers in terms of gender and age. Age of the speakers, and consequently, both cognitive and linguistic maturity of the speakers could play a role in the degree of complexity of the language used by L2 users (e.g. Nippold, Ward-Lonergan & Fanning 2005). Table 1 provides the number of candidates per level and L1. As can be seen, there are fewer C2 language samples, which is representative of the TLC corpus in general but also means that any result about that proficiency level will have to be interpreted with caution.

Table 1. Number of texts per proficiency level and L1 background

	B1	B2	C1	C2	
Italian_L1	97	100	51	9	257
Spanish_L1	197	169	97	23	486
Total	294	269	148	32	743

In terms of tasks, there are four different speaking tasks represented in the TLC with two dialogic and highly interactive tasks that are administered across all proficiency levels, i.e. the ‘discussion’ task and the ‘conversation’ task, allowing to keep the effect

of the task constant. At the higher proficiency levels (C1 and C2), the 'discussion' task starts with the examiner asking about ideas and information expressed in the 'presentation' task (this monologic task is based on a topic freely chosen and prepared by the L2 speakers in advance of the exam, allowing them to select a topic they are familiar with and interested in), and a conversation ensues from there. At the lower levels (B1 and B2), which do not include the 'presentation' task, the candidates introduce their topic briefly and the discussion develops from there. In the 'conversation' task, the candidate is invited to engage in a conversation about two topics of general interest. The topics for candidates at the B1, B2 and C1 levels are selected from a list of topics (e.g. "Society and living standards", "Personal values and ideal" and "National environmental concerns") known to the candidates before the exam; candidates at the C2 level can be given any topic considered suitable by the examiner (see Gablasova et al. 2019: 143–145, for more details about the tasks). While the 'discussion' task gives full flexibility to the L2 speakers in terms of the choice of the topic, the 'conversation' task is more limited in terms of the topic areas discussed by the L2 speakers. However, as the topic areas on the list were modified regularly over the five years of the TLC data collection and the topics are further personalised by the examiners, the 'conversation' task includes a high number of different topics distributed across different L2 proficiency levels. The topic of the spoken exchange was thus not included as a predictor (independent) variable in this study (see also Brezina & Fox 2021); however, see the Discussion section for a discussion of potential topic effects that should be taken into consideration when explaining the phraseological patterns found in this study.

2.2 Data preparation: Co-occurrence extraction and analysis

Unlike in previous studies where relational co-occurrences were extracted on the basis of parsed data, the extraction method used in this study relies on regular expressions to extract verb + nominal objects from the CLAWS C6 part-of-speech tagged corpus version available in the *Sketch Engine* (Kilgariff et al. 2014).² The main reason for not exploring the possibility of using parsed spoken learner language was that, at the time of data retrieval and analysis (2018), the TLC was only available to the research community via the Sketch Engine.

After a lengthy process of writing, evaluating (with precision and recall metrics) and improving CQL (Corpus Query Language) syntax searches, we extracted 15,643 concordance lines with the following query from the Italian and Spanish L1 TLC components:

2. <<http://www.sketchengine.eu>>


```
[tag="V[VHD].*"&word!="(think|know)"] [tag="(D|A|U|PPH1|PPHO1|
PPHO2|PPIO1|PPIO2|PPY).*" ] {0,3} ([tag="(R|U|MC|MD).*" ] {1,3} [tag=
"(J|U|CC).*" ] {1,2} | [tag="(J|U|CC).*" ] {0,3} ) [tag="(NN|U).*"&word!="lot" ] {0,3}
(meet [tag="NN.*"&word!="lot" ] [tag!="NN.*" ] 0 1) within <u/>
```

All hits were checked manually by a research assistant with the help of a coding scheme that was developed, piloted and revised during an internship at the ESRC Centre for Corpus Approaches to Social Science, Lancaster University (UK). The final set includes 13,985 verb + nominal object structures (89% precision). While recall (70.5%) may seem relatively low, we were not able to improve the query further as missed V + O structures include part-of-speech errors, verb and object in separate turns and inversions.

The study makes use of Mutual Information (MI) to explore EFL learners' use of statistical collocations as this association measure has been shown to tap into the sophistication of word combinations by promoting the relatively less frequent and semantically complex word pairs in learner production (Paquot 2019; Paquot et al. 2021). The reference corpus used to compute MI scores was the Web corpus ENCOW14 (sentence shuffle AX version; 9,578,828,861 tokens; Schäfer 2015).³ Given the lack of sufficiently large and comparable spoken reference corpora, we opted for the ENCOW14 as web corpora from the COW family include a variety of text types such as quasi-spontaneous language and blogs largely targeting a general audience across domains (Schäfer & Bildhauer 2012). Compared to other reference corpora such as the *British National Corpus*, they are also large enough for many English collocations to appear (Paquot et al. 2021). As the ENCOW14 corpus is distributed with Stanford-typed dependencies, it seemed more practical to extract V + O structures on the basis of *dobj* dependencies instead of regular expressions from this corpus. As a result, a list of *dobj* dependency-based pairs of words (as lemmas) with frequency information was used as input to the *Ngram Statistics Package* (NSP),⁴ and an MI score was calculated for each lemma pair that appears with a frequency of at least five occurrences in ENCOW14.

Each verb + direct object combination in the TLC transcripts was then lemmatized and looked up in the list of dependencies extracted from the ENCOW14 to determine its MI score in a reference corpus of contemporary general English. If a verb + direct object combination was not found or appeared fewer than five times in ENCOW14, it was removed from further analysis. This criterion was applied to identify the set of word pairs that best fit our definition of phraseological complexity

3. <<http://corporafromtheweb.org/encow14/>>

4. <<http://www.d.umn.edu/~tpederse/nsp.html>>

(which includes the notion of *target-like* phraseological units; see above) and help answer our research questions while minimizing interaction effects with the different construct of accuracy as well as aspects of L2 learners' creativity. Finally, the last step involved computing a median MI score for each learner transcript on the basis of all the verb + object combinations (tokens) extracted with the help of the regular expression described above. Unlike in previous studies of phraseological complexity where means were often selected, median scores were used in the current study to better account for the sometimes highly skewed distribution of MI scores in learner production. Median association measures were calculated with R (R Core Team 2019). No medians were computed for transcripts with fewer than three verb + nominal objects and such transcripts were removed from further analyses.

2.3 Statistical evaluation

To answer our research questions, we made use of mixed-effects modelling to assess the influence of proficiency on (1) phraseological diversity operationalized as RTTRs of *dobj* dependencies and (2) phraseological sophistication operationalized as median MI scores for *dobj* dependencies, while taking into account any random variation observed across the participants. Given the ordinal nature of the proficiency variable, we defined orthogonal contrasts for sequential-differences testing. We also included L1 and task as fixed effects (with 2-by-2 interactions) in our models because previous research has identified these variables as potential important explanatory factors in EFL learners' use of word combinations (e.g. Biber & Gray 2013; Alexopoulou et al. 2017). By doing so, we followed Gries's (2018) recommendation to approach even a monofactorial hypothesis (here, proficiency has an effect on phraseological complexity in EFL learner speech) with a multifactorial design "to determine either (a) whether it adds anything to what we already know about the phenomenon (by statistically controlling for what we already know) or (b) whether it replaces (parts of) what we already know about the phenomenon" (Gries 2018: 296). All statistical analyses were performed with R (R Core Team 2019) and the *lmerTest* (Kuznetsova et al. 2017), *MuMIn* (Barton, 2020), *effects* (Fox & Weisberg, 2019), *MASS* (Venables & Ripley, 2002) and *performance* (Lüdtke et al. 2020) packages. Model diagnostics revealed no problematic aspects (residuals appeared to be independent and not autocorrelated; no multicollinearity was found among main effects).

3. Results

The first model we computed tests to what extent phraseological diversity increases with proficiency in EFL learners’ oral performance. It explains 12% of the variance in phraseological diversity (R^2 marginal: 0.12; R^2 conditional: 0.27). First, as reported in Table 2, no significant main effect of, or interaction effect with, L1 was found in this model. By contrast, there is a significant main effect of PROFICIENCY for B1 > C1 and B1 > C2, a significant main effect of TASK, and an interaction between PROFICIENCY and TASK that is significant for PROFICIENCY B1 > C1 and B1 > C2.

Table 2. Summary results of the regression model for phraseological diversity (RTTR)

	Estimate	SE	T <i>Satterthwaite</i>	p
(Intercept)	2.728	0.059	46.556	< 0.000 ***
PROFICIENCY_B2-B1	0.086	0.080	1.066	0.287
PROFICIENCY_C1-B2	0.624	0.098	6.373	0.000 ***
PROFICIENCY_C2-C1	0.264	0.202	1.308	0.191
L1_Spanish	0.112	0.068	1.650	0.099
TASK_discussion	-0.169	0.063	-2.685	0.007 **
PROFICIENCY_B2-B1:L1_Spanish	0.115	0.089	1.282	0.200
PROFICIENCY_C1-B2:L1_Spanish	0.001	0.109	0.013	0.990
PROFICIENCY_C2-C1:L1_Spanish	-0.121	0.219	-0.555	0.579
PROFICIENCY_B2-B1:TASK_discussion	-0.054	0.072	-0.744	0.457
PROFICIENCY_C1-B2:TASK_discussion	-0.596	0.088	-6.806	0.000 ***
PROFICIENCY_C2-C1:TASK_discussion	-0.156	0.167	-0.936	0.350
L1_Spanish:TASK_discussion	0.059	0.066	0.900	0.368
Varying intercepts	PARTICIPANT 0.076 (0.28)	Residual 0.366 (0.60)		

On the whole, phraseological diversity operationalized as RTTR increases with proficiency. As shown in Figure 1, the interaction effect between PROFICIENCY and TASK is so that phraseological diversity increases more in the ‘conversation’ task, with a significant difference between B2 and C1.

The second model evaluates the extent to which phraseological sophistication also increases with proficiency in EFL learners’ oral performance. Unlike for phraseological diversity, in this second model, there is only a significant main effect of PROFICIENCY on phraseological sophistication but there is no main effect of TASK

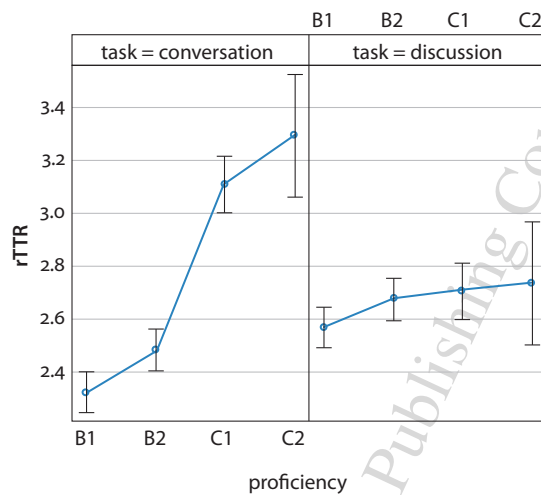


Figure 1. The interaction effect of PROFICIENCY and TASK on RTTR

or L1. As shown in Table 3, significant interaction effects were found between PROFICIENCY and L1, and between PROFICIENCY and TASK but only when PROFICIENCY goes from B1 to B2. The model, however, came with a very small amount of variance explanation (R^2 marginal: 0.062; R^2 conditional: 0.082).

Table 3. Summary results of the final regression model for phraseological sophistication (median MI)

	Estimate	SE	<i>p</i>		
(Intercept)	2.079	0.118	17.676	< 0.000	***
L1_Spanish	−0.261	0.135	−1.935	0.053	
PROFICIENCY_B2-B1	−1.087	0.162	−6.693	0.000	***
PROFICIENCY_C1-B2	0.041	0.193	0.212	0.832	
PROFICIENCY_C2-C1	0.285	0.396	0.721	0.471	
TASK_discussion	0.060	0.142	0.425	0.671	
L1_Spanish:PROFICIENCY_B2-B1	0.528	0.174	3.035	0.002	**
L1_Spanish:PROFICIENCY_C1-B2	0.031	0.208	0.149	0.882	
L1_Spanish:PROFICIENCY_C2-C1	−0.193	0.419	−0.461	0.645	
L1_Spanish:TASK_discussion	−0.036	0.148	−0.241	0.810	
PROFICIENCY_B2-B1:TASK_discussion	0.459	0.162	2.831	0.005	**
PROFICIENCY_C1-B2:TASK_discussion	−0.343	0.195	−1.762	0.078	
PROFICIENCY_C2-C1:TASK_discussion	−0.149	0.369	−0.404	0.686	
Varying intercepts	PARTICIPANT	Residual			
	0.037 (0.19)	1.76 (1.33)			

Figure 2 shows that phraseological sophistication does not increase linearly with higher proficiency. Quite the contrary, there is a significant decrease in MI scores from B1 to B2. This decrease is particularly marked in the conversation task for both Italian and Spanish EFL learners; it is also particularly marked for Italian speakers across tasks.

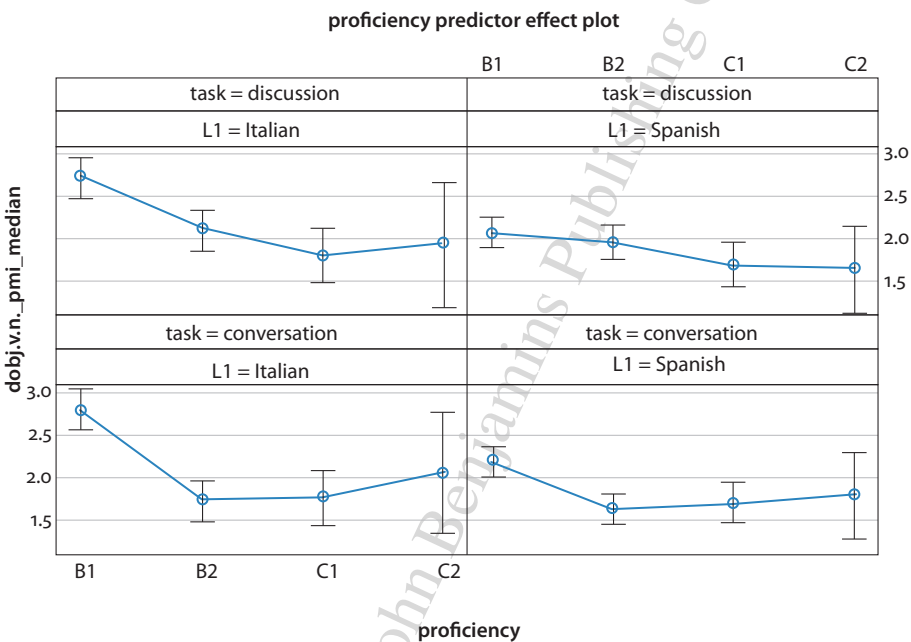


Figure 2. The effect of PROFICIENCY by L1 and TASK on median MI scores

4. Discussion

Two mixed-effect models were computed to address the research questions and test our hypotheses. The first model served to examine the extent to which RTTR scores increase from the B1 level to the C2 level on the CEFR scale in a subset of the TLC made up of conversation and discussion tasks by Italian and Spanish EFL learners. Results partly confirm our working hypothesis by showing that phraseological diversity increases with proficiency in EFL learners' oral performance. However, they also reveal that phraseological diversity only increases significantly between B2 and C1, thus distinguishing clearly between the intermediate and advanced learners. These findings provide a slightly different picture than reported in Biber and Gray (2013) for collocations with high-frequency verbs, and do not support the second part of our hypothesis that the development of phraseological diversity

would reach a plateau at the upper-intermediate level. Note, however, that these results are closer to those of Römer and Garner (2019), who reported an increase in Verb-Argument Construction (VAC) productivity with increased proficiency in the TLC, with learners at B2 and above producing a wider range of verbs in the target VACs than B1 learners.

Interestingly, this increase is more visible in the conversation task than the discussion task, which we very tentatively attribute to the nature of the task. On the one hand, the discussion task focuses on a topic that the L2 speaker selected and is thus presumably more familiar with. Preparing the topic and thinking about it, it is likely that the vocabulary in the semantic field of the selected topic has also been more activated. It could thus be assumed that this activation and a certain amount of planning could contribute to higher complexity across different dimensions of language (e.g. lexical, grammatical as well as phraseological) as suggested by SLA research on the effects of planning on L2 spoken production (e.g. Yuan & Ellis 2003). On the other hand, by focusing on two topics of general interest, the conversation task provides a greater opportunity for learners to use a more varied set of word combinations than the discussion task, which may involve greater repetition in terms of vocabulary related to a particular semantic field (for an example of this effect, see Brezina & Fox 2021: 172). The results suggest that the greater opportunity offered by the different topics benefits the advanced learners to a larger extent, allowing them to display a greater variety and a more diverse set of word combinations than the intermediate students who tend to rely on a more limited set of word combinations despite the greater range of topic areas.

The second model was used to investigate to what extent phraseological sophistication operationalized as median MI scores increases with proficiency in EFL learners' oral performance. Results do not support our hypothesis that phraseological sophistication would increase with proficiency, if only at a lower rate. In fact, partly opposite results were found: median MI scores decrease from B1 to B2 before stabilizing somewhat (no other differences are significant across tasks and L1s). This effect is more visible in the conversation task; it is also more marked in the tasks performed by Italian learners. The discussion that follows will remain tentative until further research has been completed because the model comes with very little variance explained. There are at least three potential explanations for that, some related to the nature of L2 production and others related to the methodological approach adopted in evaluating the word combinations. First, recent studies have shown that prompt or topic can have a significant effect on EFL learners' use of collocations. Gablasova et al. (2017), for example, demonstrated how a discussion on a topic (e.g. global warming) that involves more technical (and rarer) terms may result in higher average MI scores than another topic (e.g. social conditioning) (see also Paquot et al. 2021). However, it was not possible for us to use prompt or

topic as an independent variable as the two tasks investigated offer a large variety of topics that are not systematically recorded in the TLC metadata. Second, it was also not possible to account for the effect of priming in the study, both in the form of self-priming and priming from the interlocutor (McDonough & Kim 2009; Gries 2019). It is likely that, as the data includes highly interactive spoken conversation, both the L2 and L1 speakers primed each other in the use of specific expressions. L2 speakers of different proficiency may be able to use language produced by the other speaker in the exchange in a different way (e.g. to repeat it verbatim as part of their own utterance or to integrate it more flexibly into their own production). It is possible that L2 speakers of lower proficiency levels may rely more heavily on the language produced by the examiner (an L1 speaker of English) and be more likely to repeat it than the L2 speakers of higher proficiency. Such tendency could result in a situation where L2 speakers at lower proficiency level may use some word combinations that are more complex while the speakers at higher proficiency levels, who are likely to be more independent, may use language typical of their own production (e.g. including mistakes and less native-like combinations). However, the effect of priming in spoken interactive communication is very complex and would require a separate study dedicated to understanding the pattern. Third, it may also be the case that average MI scores are not as reliable measures of phraseological sophistication in spoken production as they have been found to be in written learner corpora such as VESPA (Paquot 2019) or LONGDALE (Paquot et al. 2021) because these scores are computed on the basis of fewer data points per learner transcript in the TLC (median = 9, IQR = 6).

With these limitations in mind, there are nevertheless a few important points that can be made. There are indeed several factors that seem to explain the unexpected decrease in median MI scores between B1 and B2. First, the higher median MI scores at B1 come with a lower percentage of negative MI scores in EFL learners' oral performance. What a negative MI score tells us is that the two words co-occur less than would be expected. A verb + object structure can receive a negative MI score for a variety of reasons, including ~~lack of occurrence in the selected reference corpus despite the combination being correct~~, creative use on the part of the learners, or erroneous use. As shown in Table 4, there is a significant increase of verb + object structures with negative MI scores from B1 to B2 to C1. It is only at C2 that EFL learners start using verb + object structures with negative MI scores less.

Importantly, verb + object structures with negative MI scores are of a different nature with increased proficiency. Most structures with negative MI values at B1 are of the type illustrated in (1a–1e), that is, more or less acceptable (lemma) combinations with a high frequency verb that is being used where more precise verbs could have been used instead in a more strongly associated collocation (e.g. *exhibit/ display* + *behaviour* instead of *have* + *behaviour*, *collect/charge* + *tax* instead

Table 4. Ratio of verb + object structures with a negative MI score out of total number of verb + object structures by proficiency

	Q1	Median	Q3	Mann-Whitney U tests
B1	0	0.14	0.25	
B2	0.08	0.18	0.33	[U = 125815, $p < .001$] ***
C1	0.13	0.2	0.33	[U = 61669, $p < .001$] ***
C2	0.09	0.2	0.28	[U = 9784, $p = 0.35$]

of *have + tax*, *commit + fault* instead of *make + fault*, *sit + examination* instead of *have + examination*). By contrast, at B2 onwards, a larger proportion of verb + object structures with negative MI scores consist of combinations with other verbs than high-frequency verbs such as *have*, *make* or *do*. As shown in Examples (1f–1n), learners are often penalized for using such combinations when MI scores are used to describe their collocational competence as these combinations often receive very high negative MI scores.

(1) Verb + object structures with negative MI scores

- a. *children who play more violent video games er are more likely to have an aggressive behaviour* (Italian, B1, MI = -1.93)
- b. *the people who smoke or the people who to who who who e= who who drink alcohol we we have some taxes for that kind of of </s><s> but </s><s> of </s><s> we are talking about sugar drinks* (Spanish, B1, MI = -0.87)
- c. *erm last week er in the classroom at school ah at school er erm I I had the the examination for the erm last class* (Italian, B1, MI = -1.61)
- d. *er no er </s><s> no </s><s> we do the experience </s><s> uhu </s><s> and the but er is very very hard* (Spanish, B1, MI = -0.88)
- e. *with too much force er you can get a you can make a fault* (Spanish, B1)
- f. *erm at beginning er many students refuse history of art but after few lesson er they begin to appreciate the art* (Italian, B2, MI = -3.88)
- g. *erm you 're a mother tongue the language and erm this is er your erm rea= your real job or do you practise other job ?* (Italian, B2, MI = -3.97)
- h. *er I think that er it 's not very good for my wallet </s><s> mm </s><s> I afford the future because I </s><s> it could be difficult to find </s><s> yes </s><s> work yes* (Italian, B2, MI = -2.44)
- i. *after twenty three years of hard work </s><s> mm </s><s> we obtain the possibility to govern this this town hall* (Spanish, B2, MI = -3.12)
- j. *I would er taught myself problems that were very close to to reality* (Spanish, B2, MI = -2.67)
- k. *because otherwise we can lose the difference or we can lose we can get lost between the two worlds* (Italian, C1, MI = -4.13)

- l. *devices the internet too of course the internet poses erm untruthful information that makes people rethink things they did n't know there existed*
(Spanish, C1, MI = -6.02)
- m. *it 's stra= ye= rich people educated people has have mm this probl= er resent this problem*
(Italian, C2, MI = -2.49)
- n. *but I think somebody in this case this this person could go for other way and </s><s> mm </s><s> and reach another alternative*
(Spanish, C2, MI = -3.53)

Second, together with our finding that there is less phraseological diversity at B1, the higher median MI scores at B1 can also be explained by the following trends:

- B1 learners use a very limited number of strongly associated collocations (Table 5), which they are confident in using and repeating more, a finding that has already been reported in the literature for learner writing (e.g. Granger 1998; Nesselhauf 2005).
- B1 learners try less often to combine more sophisticated words, which, as we have seen above, seems to penalize upper-intermediate and advanced learners.

Note that B1 learners' specific use of V + O combinations may have inflated median MI scores based on tokens as computed in this study. Future studies of phraseological complexity may consider using collocational types instead of tokens to better distinguish between the two dimensions of diversity and sophistication. This may be particularly desirable when investigating phraseological complexity in interactive (unplanned) speech, which is characterised by greater repetitiveness than written production (e.g. Biber et al. 1999; Leech 2000) especially in the case of (lower proficiency) L2 learners.

These combined results show that the significant decrease in median MI scores at B2 and the lack of significant quantitative progress between B2 and C2 may actually hide qualitative development, with upper-intermediate and advanced learners using a more diverse set of verb + object combinations and trying to use richer vocabulary. EFL learners' attempt at using more sophisticated or specific single words with increased proficiency in speech has already been documented by Kyle and Crossley (2015) and Eguchi and Kyle (2020). In these two studies, the authors also stressed the importance of lexical sophistication in the assessment of EFL learners' oral performance, which supports particularly well Jones et al.'s (2017) statement that, in oral exams, "to be successful at higher levels, a wider range of types (...) must be evidenced" (p. 42). In our dataset, increased lexical sophistication is best visible in the type of verbs used by EFL learners across proficiency levels. For example, 84% of the verb types used in verb + object structures with negative MI scores at B1 belong to the first 1,000 words of the language, while only 74%, 69% and 72% of the verb types belong to the first 1,000 words at B2, C1 and C2

Table 5. Verb + object structures with positive MI scores repeated more than three times at B1 (all verbs and nouns are lemmas)

Verb	Direct object	MI	Verb	Direct object	MI	Verb	Direct object	MI
dance	tango	12.15	visit	museum	7.18	play	game	6
sing	song	8.91	drink	water	7.18	do	homework	5.91
rent	apartment	8.87	win	trophy	7.17	drive	car	5.89
drink	alcohol	8.85	wear	clothes	7.17	buy	ticket	5.87
drink	tequila	8.62	pay	rent	7.11	change	mind	5.87
wear	uniform	8.38	win	league	7.08	learn	language	5.79
wear	helmet	8.37	play	piano	7.01	need	help	5.68
score	goal	8.28	play	football	7	eat	junk	5.66
speak	language	8.09	ask	question	6.91	watch	film	5.62
wear	t-shirt	7.88	attend	school	6.78	read	book	5.62
wear	trousers	7.79	play	guitar	6.71	spend	time	5.62
watch	television	7.65	go	fishing	6.71	break	rule	5.59
eat	meal	7.63	eat	vegetable	6.67	eat	fish	5.56
watch	TV	7.59	watch	movie	6.67	listen	music	5.56
pass	exam	7.55	eat	pizza	6.62	save	money	5.52
open	door	7.44	attend	university	6.59	watch	football	5.52
play	tennis	7.43	eat	food	6.51	play	instrument	5.4
play	soccer	7.35	eat	salad	6.41	play	sport	5.29
solve	problem	7.34	send	message	6.39	write	book	5.16
play	volleyball	7.22	pay	attention	6.37	clean	house	5.05
play	basketball	7.21	eat	fruit	6.34	earn	money	5.04
study	biology	7.19	study	medicine	6.14	learn	vocabulary	5

respectively.⁵ However, as shown above (examples 1f-1n), increasing risk taking in terms of lexical sophistication seems to have a negative impact on the idiomaticity and/or accuracy of word combinations, with higher level learners' attempts leading to more erroneous or less idiomatic combinations (Thewissen 2013: 87).

5. Conclusion

This study is one of very few studies that examined learners' use of verb + object structures in oral performance. It is also the first study that analysed learners' phraseological competence in terms of phraseological complexity, with its two main dimensions of phraseological diversity and sophistication operationalized as RTTR and median MI scores respectively. While its findings can only remain tentative

5. We used the VP-Classic tool from Compleat Lexical Tutor v. 8.3 (<https://www.lex tutor.ca/vp/>) to explore lexical sophistication.

given the little variance explained, they draw a complex picture of phraseological complexity in EFL learners' oral performance across proficiency levels, with phraseological diversity increasing on the whole and phraseological sophistication decreasing significantly from B1 to B2. A number of reasons have been proposed for this unexpected behaviour, most particularly that EFL learners at B2 and above repeat fewer combinations and use more specific or sophisticated verbs in verb + object structures, which sometimes leads to more erroneous or non-idiomatic combinations that should nevertheless be considered as an indication of qualitative development and progress. By contrast, B1 learners just seem to 'play it safe', using a limited set of combinations they are confident in using and repeating them more. Before we are in a position to draw pedagogical implications, these findings are in need of replication. Another particularly important avenue for future research that we have started pursuing now is to study the effect of task and mode on phraseological complexity in written and oral performance by the same learners (see Vandeweerd, Housen & Paquot submitted).

There are also a number of methodological lessons to be learnt from this study. First, while MI scores have already proved [reliable](#) measures of phraseological sophistication in the context of academic writing (Paquot 2019; Paquot et al. 2021), caution should be exercised when analyzing these measures, especially when they are reported in the form of measures of central tendency (means, medians). First, it is important to examine how well a mean/median MI score summarizes the data points (in our case, verb + object structures) produced by each learner: it may be the case that depending on the corpus used, mode, learners' proficiency, or topic, there may be too much variability in MI scores and that a measure of central tendency may just not be able to provide a good representation of learners' phraseological competence and/or development. Second, as already suggested by Granger and Bestgen (2017), more attention should be devoted to combinations with negative scores and their impact on measures of central tendency for MI values. This is all the more important since EFL learners' production at B2 and above are typically characterized by the combination of more specific and less frequent words, which are rated with very high negative MIs, thus penalizing upper-intermediate and advanced learners if a median value is used despite the fact that these learners may be displaying qualitative development in phraseological sophistication. It thus seems like a good idea to compute separate measures for word combinations with $MI < 0$ or ≥ 0 (Granger & Bestgen 2017). Third, analyzing phraseological sophistication together with phraseological diversity seems warranted: As demonstrated in this study, lower proficiency learners may display high phraseological sophistication just by sticking to a limited number of highly associated combinations. Future research should investigate whether mean/median MI scores need to be computed on the basis of collocation tokens or types to better distinguish the dimension of

phraseological sophistication from phraseological diversity. Last, the distributional profile of verbs found in this study also points to the need for more studies that will examine the combined development of lexical complexity and phraseological complexity in learners' oral performance.

Acknowledgements

We would like to thank Zoé Broisson for her invaluable help with data extraction and coding.

This study was first presented at the 14th American Association for Corpus Linguistics (AACL) conference, 20–22 September 2018, Atlanta, GA, USA.

Funding

The first author gratefully acknowledges the financial support of the Belgian Fund for Scientific Research (F.R.S.-FNRS).

References

- Ackermann, Kristen & Chen, Yu-Hua. 2013. Developing the Academic Collocation List (ACL): A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes* 12(4): 235–247. <https://doi.org/10.1016/j.jeap.2013.08.002>
- Alexopoulou, Theodora, Michel, Marije, Murakami, Akira & Meurers, Detmar. 2017. Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning* 67(S1): 180–208. <https://doi.org/10.1111/lang.12232>
- Barton, Kamil. 2020. MuMIn: Multi-Model Inference. R package version 1.43.17. <<https://CRAN.R-project.org/package=MuMIn>> (20 December 2021).
- Biber, Douglas, & Gray, Bethany. 2013. Discourse characteristics of writing and speaking task types on the TOEFL iBT® test : A lexico-grammatical analysis. *ETS Research Report Series* 2013(1): i–128. <https://doi.org/10.1002/j.2333-8504.2013.tb02311.x>
- Biber Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman. Also published as Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, & Finegan, Edward. 2021. *Grammar of Spoken and Written English*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.232>
- Brezina, Vaclav & Fox, Lorrae. 2021. Adjective + noun collocations in L2 and L1 speech: Evidence from the Trinity Lancaster Corpus and the spoken BNC2014. In *Perspectives on the Second Language Phrasicon: The View from Learner Corpora*, Sylviane Granger (ed.), 152–177. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781788924863-008>
- Bulté, Bram & Housen, Alex. 2012. Defining and operationalising L2 complexity. In *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* [Language Learning & Language Teaching 32], Alex Housen, Folkert Kuiken & Ineke Vedder (eds), 23–46. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.32.02bul>

- De Marneffe, Marie-Catherine & Manning, Christopher D. 2013. Stanford Typed Dependencies Manual. <http://nlp.stanford.edu/software/dependencies_manual.pdf> (30 April 2021).
- Eguchi, Masaki & Kyle, Kristopher. 2020. Continuing to explore the multidimensional nature of lexical sophistication: The case of oral proficiency interviews. *The Modern Language Journal* 104(2): 381–400. <https://doi.org/10.1111/modl.12637>
- Ellis, Rod & Yuan, Fangyuan. 2005. The effects of careful within-task planning on oral and written task performance. In *Planning and Task Performance in a Second Language*, Rod Ellis (ed.), 167–192. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.11.11ell>
- Fox, John & Weisberg, Sanford. 2019. *An R Companion to Applied Regression* (3rd edn). Thousand Oaks CA: Sage. <<https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html>> (20 December 2021).
- Gablasova, Dana, Brezina, Vaclav & McEnery, Tony. 2017. Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning* 67(S1): 155–179. <https://doi.org/10.1111/lang.12225>
- Gablasova, Dana, Brezina, Vaclav & McEnery, Tony. 2019. The Trinity Lancaster Corpus: Development, description and application. *International Journal of Learner Corpus Research* 5(2): 126–158. <https://doi.org/10.1075/ijlcr.19001.gab>
- Granger, Sylviane. 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. In *Phraseology: Theory, Analysis and Applications*, Anthony Cowie (ed.), 145–160. Oxford: OUP.
- Granger, Sylviane & Bestgen, Yves. 2017. Using collgrams to assess L2 phraseological development: A replication study. In *Language, Learners and Levels: Progression and Variation*, Pieter de Haan, Rina de Vries & Sanne van Vuuren (eds), 385–408. Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger, Sylviane & Paquot, Magali. 2008. Disentangling the phraseological web. In *Phraseology: An Interdisciplinary Perspective*, Sylviane Granger & Fanny Meunier (eds), 27–49. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.139.07gra>
- Gries, Stefan T. 2018. On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally. *Journal of Second Language Studies* 1(2): 277–309. <https://doi.org/10.1075/jsls.00005.gri>
- Gries, Stefan T. 2019. Priming of syntactic alternations by learners of English: An analysis of sentence-completion and collostructional results. In *Using Corpus Methods to Triangulate Linguistic Analysis*, Jesse Egbert & Paul Baker (eds), 219–238. New York NY: Routledge. <https://doi.org/10.4324/9781315112466-9>
- Hanks, Patrick. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge MA: The MIT Press. <https://doi.org/10.7551/mitpress/9780262018579.001.0001>
- Jones, Christian, Byrne, Shelley & Halenko, Nicola. 2017. *Successful Spoken English: Findings from Learner Corpora*. New York NY: Routledge. <https://doi.org/10.4324/9781315101712>
- Kilgarrriff, Adam, Baisa, Vít, Bušta, Jan, Jakubiček, Miloš, Kovář, Vojtěch, Michelfeit, Jan, Rychlý, Pavel & Suchomel, Vít. 2014. The Sketch Engine: Ten years on. *Lexicography* 1: 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kormos, Judit. 2014. *Speech Production and Second Language Acquisition*. New York NY: Routledge. <https://doi.org/10.4324/9780203763964>
- Kuznetsova, Alexandra, Brockhoff, Per B., Christensen, Rune H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13): 1–26. <https://doi.org/10.18637/jss.v082.i13>

- Kyle, Kristopher & Crossley, Scott. 2015. Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly* 49(4): 757–786.
<https://doi.org/10.1002/tesq.194>
- Laufer, Batia. 1994. The lexical profile of second language writing: Does it change over time? *RELC Journal* 25: 21–33. <https://doi.org/10.1177/003368829402500202>
- Laufer, Batia & Nation, Paul. 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16(3): 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Laufer, Batia & Waldman, Tina. 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning* 61(2): 647–672.
<https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Leech, Geoffrey. 2000. Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning* 50(4): 675–724. <https://doi.org/10.1111/0023-8333.00143>
- Lüdecke, Daniel, Ben-Shachar, Mattan S., Patil, Indrajeet, Waggoner, Philip & Makowski, Dominique. 2021. Performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software* 6(60): 3139.
<https://doi.org/10.21105/joss.03139>
- McDonough, Kim & Kim, Youjin. 2009. Syntactic priming, type frequency, and EFL learners' production of *wh*-questions. *The Modern Language Journal* 93(3): 386–398.
<https://doi.org/10.1111/j.1540-4781.2009.00897.x>
- Nesselhauf, Nadja. 2005. *Collocations in a Learner Corpus* [Studies in Corpus Linguistics 14]. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.14>
- Nippold, Marilyn A., Ward-Lonergan, Jeannene M. & Fanning, Jessica L. 2015. Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language, Speech, and Hearing Services in Schools* 36: 125–138.
[https://doi.org/10.1044/0161-1461\(2005/012\)](https://doi.org/10.1044/0161-1461(2005/012))
- Ortega, Lourdes. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24(4): 492–518.
<https://doi.org/10.1093/applin/24.4.492>
- Paquot, Magali. 2018. Phraseological competence: A useful toolbox to delimitate CEFR levels in higher education? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly* 15(1): 29–43. <https://doi.org/10.1080/15434303.2017.1405421>
- Paquot, Magali. 2019. The phraseological dimension in interlanguage complexity research. *Second Language Research* 35(1): 121–145. <https://doi.org/10.1177/0267658317694221>
- Paquot, Magali, Hasselgård, Hilde & Oksefjell Ebeling, Signe. 2013. Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. In *Twenty Years of Learner Corpus Research: Looking Back, Moving Ahead* [Corpora and Language in Use – Proceedings 1], Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (eds), 377–387. Louvain-la-neuve: Presses universitaires de Louvain.
- Paquot, Magali, Naets, Hubert & Gries, Stefan T. 2021. Using syntactic co-occurrences to trace phraseological complexity development in learner writing: verb + object structures in LONGDALE. In *Learner Corpus Research Meets Second Language Acquisition*, Bert Le Bruyn & Magali Paquot (eds), 122–147. Cambridge: CUP.
- Römer, Ute & Garner, James R. 2019. The development of verb constructions in spoken learner English: Tracing effects of usage and proficiency. *International Journal of Learner Corpus Research* 5(2): 207–230. <https://doi.org/10.1075/ijlcr.17015.rom>

- Rubin, Rachel, Housen, Alex & Paquot, Magali. 2021. Phraseological complexity as an index of L2 Dutch writing proficiency: A partial replication study. In *Perspectives on the Second Language Phrasicon: The View from Learner Corpora*, Sylviane Granger (ed.), 101–125. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781788924863-006>
- Schäfer, Roland. 2015. Processing and querying large web corpora with the COW14 architecture. In *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3), Lancaster, 20 July 2015*, Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lungen & Andreas Witt (eds), 28–34. Mannheim: Institut für Deutsche Sprache. <https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/3836/file/Schaefer_Processing_and_querying_large_web_corpora_2015.pdf> (14 October 2021).
- Schäfer, Roland & Bildhauer, Felix. 2012. Building large corpora from the web using a new efficient tool chain. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds), 486–493. European Language Resources Association (ELRA) <http://www.lrec-conf.org/proceedings/lrec2012/pdf/834_Paper.pdf> (14 October 2021).
- Sinclair, John M. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.
- Thewissen, Jennifer. 2013. Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *The Modern Language Journal* 97(S1): 77–101. <https://doi.org/10.1111/j.1540-4781.2012.01422.x>
- Vandeweerd, Nathan, Housen, Alex & Paquot, Magali. 2021. Applying phraseological complexity measures to L2 French: A partial replication study. *International Journal of Learner Corpus Research* 7(2): 197–228. <https://doi.org/10.1075/ijlcr.20015.van>
- Vandeweerd, Nathan, Housen, Alex & Paquot, Magali. Submitted. Comparing the longitudinal development of phraseological complexity across oral and written tasks.
- Vasylets, Olena, Gilabert, Roger & Manchón, Rosa M. 2017. The effects of mode and task complexity on second language production. *Language Learning* 67(2): 394–430. <https://doi.org/10.1111/lang.12228>
- Venables, William N. & Ripley, Brian D. 2002. *Modern Applied Statistics with S* (4th edn). New York NY: Springer. <https://doi.org/10.1007/978-0-387-21706-2>
- Xu, Jing. 2018. Measuring “spoken collocational competence” in communicative speaking assessment. *Language Assessment Quarterly* 15(3): 255–272. <https://doi.org/10.1080/15434303.2018.1482900>
- Yuan, Fangyuan & Ellis, Rod. 2003. The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics* 24(1): 1–27. <https://doi.org/10.1093/applin/24.1.1>

Author queries

Please provide a complete reference for the citation ‘(Lüdecke et al. 2020)’ in this article.

Please provide a citation for the reference id “CIT0290 (Lüdecke, Daniel, Ben-Shachar, Mattan S., Patil, Indrajeet, Waggoner, Philip & Makowski, Dominique. 2021)” since citation is missing in the article.