Book of abstracts

The Graduate Student Conference in Learner Corpus Research 2023

Bridging Borders

Bozen/Bolzano (Italy)

25, 26, 27 October 2023

An online event hosted by Eurac Research and held under the aegis of the Learner Corpus Association

Edited by the Organising Committee

Olga Lopopolo, Arianna Bienati, Paolo Brasolin, Elena Ferrato, Jennifer-Carmen Frey, Aivars Glaznieks, Marta Guarda, Egon Stemle, Fabio Zanda

Acknowledgment and thanks are given to the Scientific and Evaluation Committee

Andrea Abel, Free University of Bolzano/Bozen

Katherine Ackerley, Università di Padova

Marcus Callies, Universität Bremen

Erik Castello, Università di Padova

Sylvie De Cock, Université catholique de Louvain

Sandra Deshors, Michigan State University

María Belén Díez-Bedmar, Universidad de Jaén

Hildegunn Dirdal, Universitetet i Oslo

Signe Oksefjell Ebeling, Universitetet i Oslo

Luciana Forti, Università per Stranieri di Perugia

Jennifer Carmen Frey, Eurac Research

Robert Fuchs, Universität Hamburg

Dana Gablasova, Lancaster University

Gaëtanelle Gilquin, Université catholique de Louvain - FNRS

Aivars Glaznieks, Eurac Research

Sandra Götz-Lehmann, Philipps Universität Marburg

Sylviane Granger, Université catholique de Louvain

Marta Guarda, Eurac Research

Hilde Hasselgård, Universitetet i Oslo

Shin'Ichiro Ishikawa, Kobe University

Iztok Kosem, Univerza v Ljubljani

Tove Larsson, Northern Arizona University

Cristóbal Lozano, Universidad de Granada

Anke Lüdeling, Humboldt Universität Berlin

Akira Murakami, University of Birmingham

Susan Nacey, Inland Norway University of Applied Sciences

Pascual Pérez-Paredes, Universidad de Murcia

Ute Römer, Georgia State University

Anna Shadrova, Humboldt Universität Berlin

Stefania Spina, Università per Stranieri di Perugia

Agnieszka Leńko-Szymańska, Uniwersytet Warszawski

Jennifer Thewissen, Universiteit Antwerpen

Stefanie Wulff, Bremen Universität

Book of Abstracts of The Graduate Student Conference in Learner Corpus Research 2023

Editors: Olga Lopopolo, Arianna Bienati, Paolo Brasolin, Elena Ferrato, Jennifer-Carmen Frey, Aivars Glaznieks, Marta Guarda, Egon Stemle, Fabio Zanda

Eurac Research, Bolzano 2023

Conference website: https://lcrgradconf2023.eurac.edu/

Table of contents

Valency Patterns of TAKE in EFL and ESL

Yating Tao (University of Louvain, Belgium)

English as a foreign language (EFL), also known as Learner Englishes (e.g. English used in China), and English as a second language (ESL), also known as New Englishes (e.g. English used in Singapore), have long been studied in two individual fields, namely Contact Linguistics and Second Language Acquisition, thus leading to a "paradigm gap" (Sridhar & Sridhar, 1986). Despite their differences (e.g. language input and functions), EFL and ESL varieties seem to share the same psycholinguistic processes of second language acquisition (Buschfeld, 2013; Percillier, 2016). Some studies also report a fuzzy borderline between ESL and EFL varieties (e.g. Cyprus in Buschfeld, 2013; Tswana English in Gilquin & Granger, 2011). Further, ESL varieties are not necessarily the result of colonization (Buschfeld and Kautzsch, 2014 on Namibia; Edwards on the Netherlands, 2014).

Against this backdrop, this study aims to bridge the gap between these two types of English varieties at the lexis-grammar interface, which is argued to be particularly prone to innovation (Schneider, 2007). More specifically, I focus on the valency patterns of TAKE in two corpora: student writing samples from the Hong Kong and Singapore components of the International Corpus of English (ICE) and Chinese student essays from the International Corpus of Learner English (ICLE), with British university student essays from the Louvain Corpus of Native English Essays (LOCNESS) as a reference. The valency patterns of TAKE are analyzed at two levels. At the higher level, all instances of TAKE are classified in the manner of valency pattern, that is, by distinguishing between valency patterns with it, monovalency, divalency, trivalency and quadrivalency and their subcategories. At the lower level, I examine phraseological uses including phrasal verbs, light verb constructions and idioms. To detect the potential links in valency patterns and phraseological uses among these English varieties, I employ chi-square test and hierarchical clustering analysis.

References

Buschfeld, S. (2013). English in Cyprus or Cyprus English: An Empirical Investigation of Variety Status. Amsterdam: John Benjamins.

Buschfeld, S., & Kautzsch, A. (2014). English in Namibia. English World-Wide. A Journal of Varieties of English, 35(2), 121–160. https://doi.org/10.1075/eww.35.2.01bus

Edwards, A. (2014). English in the Netherlands: Functions, forms and attitudes (Unpublished Doctoral Thesis). University of Cambridge, United Kingdom.

Gilquin, G., & Granger, S. (2011). From EFL to ESL: Evidence from the International Corpus of Learner English. In J. Mukherjee & M. Hundt (Eds.), Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap (pp. 55-78). Amsterdam: John Benjamins.

Percillier, Michael (2016). World Englishes and Second Language Acquisition: Insights from Southeast Asian Englishes. Amsterdam: John Benjamins.

Schneider, E. W. (2007). Postcolonial English: Varieties around the World. Cambridge: Cambridge University Press.

Sridhar, K. K., & Sridhar, S. N. (1986). Bridging the paradigm gap: Second language acquisition theory and indigenized varieties of English. World Englishes, 5(1), 3–14. https://doi.org/10.1111/j.1467-971X.1986.tb00636.x

The Use of Prepositional Phrase Bundles in Turkish Student Essays: A Core Expression Analysis

Ömer Faruk Kaya (Trakya University, Turkey)

In recent years, corpus-linguistic methods have been extensively used to study the use of lexical bundles (LB) by L2 writers in terms of structure and discourse functions. Previous research has shown that non-native speakers use fewer LBs than their native counterparts, but little is known about the partial uses of LBs. It was argued that automated analyses may fail to detect attempted uses of LBs, which draws an inaccurate picture of the use of LBs in learner corpora. Therefore, Shin et al. (2018) developed a core expression analysis to investigate the errors found in learner corpora. Core expression analysis is a manual check of the phrases formed around a core word (e.g., one of is the core expression in the bundles is one of the and one of the most). To date, no study has investigated the use of preposition-based LBs in the Turkish context using a core expression approach. To address this gap, the present research investigates four-word preposition-based LBs in a learner corpus (191,253 words) of human-rated L2 English opinion essays (n= 691) produced by Turkish university students by using a modified version of core expression analysis. As an extension to Shin et al., this study uses the regex function of the Antconc tool (an NLP tool to retrieve n-grams) to account for LBs with two prepositions. In a corpus-based analysis, the present study aims to (a) compare LB productions of native (LOCNESS) and non-native writers using a log-likelihood test, (b) classify student errors and find the most common errors based on hit frequency (c) examine the effects of errors on writing performance through linear mixed-effect models. It can be assumed from the findings in the literature and the possible effect of L1 Turkish that English-speaking Turkish university students might frequently misuse preposition-based LBs.

Does L1 Morphology Actually Influence ‘Passive’ Unaccusatives?: A Corpus Study of Unaccusative Errors by Japanese Learners

Yu Tazaki (Ohio University, United States of America)

Second language (L2) learners of English regardless of their first language (L1) backgrounds frequently produce and accept ungrammatical passive sentences with one type of intransitive verb, namely, unaccusative verbs such as Miki was disappeared. A number of previous studies have attempted to reveal the culprit of the unique phenomenon without reaching a consensus. This study critically reviews one of the major accounts for the ungrammatical passivization, L1 morphological transfer, which claims that L1 morphology with unaccusative verbs leads learners to generate and accept ‘passive’ unaccusatives. According to the analysis, since Japanese has morphological differences between unaccusatives and their transitive counterparts (e.g., English gather/gather, Japanese atum-e-ru/atum-ar-u), Japanese learners of English seek an equivalent and add passive morphology be + en to unaccusatives. If the analysis is on the right track, there is no wonder that Japanese learners add progressive morphology -ing to unaccusatives although unaccusatives in the progressive structure are grammatical. The present study hypothesizes that Japanese speakers produce unaccusatives in the progressive as well as in the passive. This research tests the hypothesis by analyzing a large corpus data that consists of approximately 70,000 words obtained from Japanese learners of English in junior and senior high schools. The findings in the study showed that unaccusatives appeared in the passive structures, yet few in the progressive structures. Therefore, the results demonstrated that L1 morphology has weak or no effects on ‘passive’ unaccusatives. From the analysis of the corpus, this study concluded that learners are likely to produce ‘passive’ unaccusatives due to the resemblance between unaccusatives and passives in terms of sentence structures.

Pronoun Drop as Evidence of Proficiency: Using Corpus Linguistics to Study Subject Pronoun Expression in Learners of Spanish

Miguel Hernandez Alonso (University of Utah, United States of America)

Subject pronoun expression (SPE) in Spanish has been widely researched, specifically in contrast to SPE in non-pro-drop languages. English-speaking learners of Spanish tend to prefer overt subject pronouns over null subject pronouns due to cross-linguistic transfer. Examining SPE in L2 Spanish helps us understand learners' acquisition of conjugations and their awareness of the syntactic structures. Despite extensive research on SPE across Spanish dialects and bilingual communities, there seems to be a gap regarding non-heritage learners of Spanish, specifically in the K-12 context. Therefore, this study aims to investigate the use of overt and null subject pronouns by L2 Spanish learners in dual language immersion schools.

We contrast samples from the Corpus of Utah Dual Language Immersion (CUDLI, 2019): CUDLI is a multilingual corpus of second language writing that contains responses to the presentational writing portion of the Assessment of Performance toward Proficiency in Languages (AAPPL) test. The Spanish subcorpus that we use contains more than 80,000 files by over 21,000 students, and it allows us to compare the production of students with similar instruction across different levels. We will analyze texts from grades 4th to 9th.

This study investigates the morphological features of verbs associated with students' pronoun drop, to determine if there are relevant patterns of influence. Grammatical number and person, among others, have been pointed out as habitual constraints in SPE, as well as lexical frequency, at least in Spanish L1 users (Erker & Guy, 2012). Thus, a statistical analysis would complement our findings about the dual-language-immersion students' acquisition of this feature of the Spanish language.

The findings of this research could contribute to improve language teaching and learning. We want to provide evidence for teachers to improve students' understanding of verbs and pronouns in pro-drop languages. These data could also add to the creation of pedagogic materials.

References

Erker, D., & Guy, G. R. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. Language, 88(3), 526–557. http://www.jstor.org/stable/23251863

Rubio, F. & Schnur, E. (2019). The Corpus of Utah Dual Language Immersion (CUDLI) [Learner Corpus]. University of Utah. https://l2trec.utah.edu/learner-corpora/cudli/index.php

Exploring the use of cohesive devices in writing among secondary-level learners of Chinese: a corpus study

Meng Zhou (University of Utah, United States of America)

The purpose of the study is to investigate how secondary-level (6th, 8th, and 9th grades) learners of Chinese from Utah Dual Language Immersion (DLI) programs use cohesive devices in their writing production. In Chinese, cohesive devices (e.g., conjunctive adverbs and conjunctions) can be used to connect words, phrases, clauses, and sentences. This study will focus on cohesive devices that connect clauses and sentences, which make learners' written text fluent and cohesive. Previous research found that the appropriate use of cohesive devices can improve second language learners' writing performance in terms of coherence and comprehensibility (Ke, 2005; Xiao, 2010). In addition, according to ACTFL proficiency guidelines for writing (American Council on the Teaching of Foreign Languages, 2012), learners should be able to produce connected sentences at the intermediate level; connected discourse of paragraph length and structure is required for learners at an advanced level. Therefore, learning to use cohesive devices to connect sentences is essential for learners to improve their Chinese proficiency.

The researcher will analyze the Chinese sub-corpus of Written Corpus of Utah Dual Language Immersion (Corpus of Utah Dual Language Immersion [CUDLI], n.d.), a corpus of written responses to the presentational writing portion of the ACTFL Assessment of Performance toward Proficiency in Languages (AAPPL). The Chinese sub-corpus contains data from 12,467 students from Chinese DLI programs and 47,988 text files. The analysis will focus on the frequency, variety, and accuracy of cohesive devices learners use in each grade. Then the study will investigate if learners progress using cohesive devices and how they improve. The results will be presented via descriptive statistics.

Regarding pedagogical implications, the study will provide suggestions to Chinese DLI teachers on how to help learners improve their writing proficiency. According to the result, the teacher can improve their teaching methods in classes, such as teaching cohesive devices explicitly, giving more corrective feedback on cohesive devices, and adjusting the writing prompts.

References

American Council on the Teaching of Foreign Languages. (2012). ACTFL proficiency guidelines 2012: Chinese - Simplified Characters Writing. https://www.actfl.org/educator-resources/actfl-proficiency-guidelines/chinese-simplified-characters/chinese-simplified-characters-writing

Corpus of Utah Dual Language Immersion (CUDLI). (n.d.). [Data set]. Second Language Teaching and Research Center, University of Utah.

Ke, C. (2005). Patterns of acquisition of Chinese linguistic features by CFL learners. Journal of the Chinese Language Teachers Association, 40(1), 1-23.

Xiao, Y. (2010). Discourse features and development in Chinese L2 writing. In M. E. Everson & H. H. Shen (Eds.), Research Among Learners of Chinese as a Foreign Language (pp. 133-151). Hawaii, HI: University of Hawaii.

Adverb Placement in Learner Writing: The Effect of Linguistic Features, Cross-Linguistic Transfer and Register

Vildan Ozkan (Northern Arizona University, United States of America)

Due to their syntactic mobility, adverbs lend themselves well to first-language (L1)-related syntactic transfer studies (Hasselgård, 2015); however, there is a limited body of research looking into what other factors may influence clause placement. With the exception of Larsson et al (2020), which investigated the impact of linguistic, extralinguistic and L1 transfer-related factors on adverb placement, most previous studies have focused on overuse-underuse analysis and misuse of adverbs. Moreover, very few studies go beyond the realm of well-studied Germanic and Romance languages. The present study is a partial replication study of Larsson et al.'s (2020) and revisits the question of whether adverb placement can help us detect L1 transfer; it also looks at the possible effect of linguistic (e.g., presence/absence of an auxiliary), extralinguistic variables (e.g., native-speaker status, L1 background) and register. However, the aim is not only to see if those findings can be replicated on German and native-speaking students, but also to extend our knowledge by also looking at Turkish, a language that is very different from English typologically.

The data were obtained from two different corpora; the International Corpus of Learner English (ICLE) and the Louvain Corpus of Native English Essays (LOCNESS). The study looks at the following 15 epistemic adverbs (Granath, 2005): maybe, probably, possibly, really, simply, actually, apparently, certainly, clearly, definitely, evidently, obviously, perhaps, surely, and of course.

The results show that German learners behaved more native-like than their Turkish counterparts who overused the clause initial (e.g., Probably she is here) and underused the clause-medial position. Similar to Larsson et al.'s (2020) findings, the main predictors of adverb placement are linguistic rather than extralinguistic; although some traces of L1 transfer were found with Turkish data. Moderate effect of register on variation in learner writing was also reported.

An investigation of pause location in spoken English corpora

Dilara Dikilitas (Northern Arizona University, United States of America)

We all pause, but, um, do we all pause … in a similar manner? The first language (L1) of speakers is proposed as one factor that influences their pausing behavior in a target language, especially with regard to the frequency and location of the pauses (Goldman-Eisler, 1968). It is also assumed that the preference for a certain type of pause (silent or filled pauses) is related to its location in an utterance (Dumont, 2018). Given the complex relationship between these variables, more research is needed to understand language learners' behavior better. This corpus-based study investigates the pause behavior of speakers of three L1 backgrounds: English, French, and Spanish. The data is taken from two English-spoken corpora: The Louvain International Database of Spoken English Interlanguage (LINDSEI) and The Louvain Corpus of Native English Conversation (LOCNEC). Descriptive statistics and binomial logistic regression were used for analysis. The study answers two research questions: (1) In what ways do L1 speakers of English, French, and Spanish differ in terms of the location and type of pause produced in English interviews? (2) Does L1 background and/or presence vs. absence of a clause boundary have an effect on what type of pause (silent vs. filled) is produced? The results show that the L1 Spanish group differed considerably from the other groups with regard to both frequency and location of pauses. Additionally, both L1 and the presence of a clause boundary were found to be predictors of pause type.

Self-mentions in persuasive writing in L2 French: a study on the use of JE, NOUS and ON by L1 Turkish learners

Sulun Aykurt-Buchwalter (LIDILEM, France)

Self-mentions are metadiscourse markers that are mostly expressed through the use of personal pronouns such as I and we. Their use generally results from a “conscious choice by writers to adopt a particular stance and a contextually situated authorial identity” (Hyland, 2005 : 53). Self-mentions in L1 Turkish students' writing in L1 and L2 English has been studied (Akbas, 2013), but there is no existing research on L1 Turkish writers in L2 French.

In French, in addition to the personal pronouns "je" (I) and "nous" (we), the indefinite pronoun "on" can sometimes be used as an equivalent to nous and therefore may express self-mention. How do Turkish learners of French employ self-mentions in argumentative writing, in comparison with native French writers ?

Our learner corpus is composed of 30 essays written by B1-level learners and 30 essays written by B2-level learners. All learners are first and second year university students. The reference corpus is composed of 30 essays written by native French students. Two prompts were used : the first is to write a formal letter to the mayor to persuade them not to cancel a concert ; the other is to write an email to other students to raise funds for a kitchen garden, on behalf of a student association. In addition to the occurrences of je and nous, we analysed each occurrence of on and only counted the ones that can be interpreted as self-mentions. In order to take into account the length of the essays, we calculated the relative frequency of these occurrences.

Results indicate that native French writers prefer the singular pronoun "je" in the formal letter task, whereas they tend to use "nous" in the email task. They rarely use "on" in either task. L1 Turkish learners use self-mentions more frequently than native writers. Learners overuse the "on" pronoun. This overuse diminishes at B2-level. The overuse of self-mentions in comparison with native writers may be the result of inappropriate teaching or hypercorrection. The overuse of "on", especially in formal letters, should be interpreted as misuse.

References

Akbas, E. (2013). Are They Discussing in the Same Way? Interactional Metadiscourse in Turkish Writers’ Texts. In Łyda, A., Warchał, K. (eds) Occupying Niches: Interculturality, Cross-culturality and Aculturality in

Academic Research. Second Language Learning and Teaching. (pp. 119-133). Springer. https://doi.org/10.1007/978-3-319-02526-1_8

Hyland, K. (2005). Metadiscourse : Exploring Interaction in Writing. Bloomsbury Publishing Plc.

Advantages of corpus-based longitudinal analyses on experimental tasks: a case study on the acquisition of the passive in L1 Italian

Elena Ferrato (Eurac Research, Università di Verona, Libera Università di Bolzano)

The present study provides a complementary approach to experimental studies for understanding young children's ability to master passive structures in Italian as their L1.

Early works focused on L1 English suggested a delay in passive acquisition even until the age of 6 or 7 (Baldie, 1976), leading to a first maturational account (Borer and Wexler, 1987), while recent cross-linguistic research has shown that children can derive passive clauses around the age of three (Pinker et al., 1987; Demuth, 1989; Fox and Grodzinsky, 1998) this also applies to Italian as a native language (Manetti, 2012; Volpato et al., 2013, 2014). However, several elicited production and comprehension tasks (Pinker et al., 1987; Crain et al. 1987(2009); O'Brien et al. 2006) have been criticized for facilitating children's mastery of passive derivation before the age of four (i.e., making the internal argument the topic of discourse): without any facilitation, children's ability to master passive derivation is believed to "mature" until that age (Snyder and Hyams, 2015).

To address this debate, the present study conducted an analysis based on longitudinal corpora, the CHILDES Italian subsection (MacWhinney, 2000), focused on the spontaneous production of passives by 18 L1 Italian children aged 1;04 – 3;04, younger than the participants of the experimental tasks on the same topic. The CHILDES longitudinal data are less susceptible to possible manipulation of the production context since they are acquired by recording children in a natural setting.

The results showed that participants were able to spontaneously produce all types of passive structures in Italian around 2 years of age. Moreover, the context of production has been analyzed, and contextual clues that may have facilitated children were found in half of the structures considered: in such cases, as in (1) and (2)[1], the internal argument bears a [+WH] or [+Topic] feature, thus it may elicit the production of a passive structure instead of an active one, as claimed by Snyder and Hyams (2015).

(1) Mother, talking about a toy (Cipriani et al., 1989):

non riesci a toglier=le la testa?

neg can-2sg to remove-inf=3sg.f.dat the.sg.f head.sg.f

‘can’t you remove her head?’

(2) Marco, 2;05:

no, perché è tenut-o drento [//] dentro.

neg, because be-3sg hold.ptcp.pst.sg.m [wrong spelling] [//] inside.

‘No, because it is kept inside’.

On the other hand, the example in (3)[2] pertains to the other half of passive structures produced without any facilitation: while further research on a larger amount of data is needed for more general claims, these occurrences may suggest that children’s ability to master passive structures is acquired earlier than previously thought.

(3) Diana, 2;06 (Tonelli et al., 1998):

e quetti [: quest-i] vanno sciugiti [:asciugat-i] bene.

and these.pl.m go-prs dry-ptcp.pst.pl.m well.

‘And these need to be dried well.’

In conclusion, these findings contribute to the growing body of cross-linguistic research that suggests children’s early acquisition of passives and highlights the advantages of corpus-based longitudinal analyses over experimental tasks.

Notes

1. Despite the lack of agreement in (2) between the gender of the internal argument and the past participle (“the head” in Italian is feminine, but the child uses the masculine form of the past participle), the passive structure is clearly referred to the discourse topic.

2. The passive auxiliary “andare” implies the presence of an external force that starts the action and has a deontic modal nuance. It is not clear what is the contextual referent of “these”, but it is certainly not the discourse topic.

References

Baldie, B. J. (1976). The acquisition of the passive voice. *Journal of Child Language*, *3*(3), 331–348.

Borer, H., & Wexler, K. (1987). The maturation of syntax. In T. Roeper & E. Williams (A c. Di), *Parameter Setting* (pp. 123–172). Reidel.

Cipriani, P., Pfanner, P., Chilosi, A., Cittadoni, L., Ciuti, A., Maccari, A., Pantano, N., Pfanner, L., Poli, P., Sarno, S., & others. (1989). Protocolli diagnostici e terapeutici nello sviluppo e nella patologia del linguaggio. *Pisa: Italian Ministry of Health, Stella Maris Foundation*.

Crain, S., Thornton, R., & Murasugi, K. (2009). Capturing the evasive passive. *Language Acquisition*, *16*(2), 123–133. Paper originally presented at the 12th Annual Boston University Conference on Language Development (1987).

Demuth, K. (1989). Maturation and the acquisition of Sesotho passive. *Language*, *65*, 56–80.

Fox, D., & Grodzinsky, Y. (1998). Children's passive: A view from the by-phrase. *Linguistic Inquiry*, *29*, 311–332.

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

Manetti, C. (2012). *The acquisition of passives in Italian: Evidence from comprehension, production and syntactic priming studies* [PhD dissertation,]. University of Siena.

O'Brien, K., Grolla, E., & Lillo-Martin, D. (2006). Long passives are understood by young children. In D. Bamman, T. Magnitskaia, & C. Saller (A c. Di), *Proceedings from the 30th Boston University Conference on Language Development* (pp. 441–451). Cascadilla Press.

Pinker, S., LeBeaux, D. S., & Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition*, *26*, 195–267.

Snyder, W., & Hyams, N. (2015). Mimimality effects in children's passives. In E. Domenico, C. Hamann, & S. Matteini (A c. Di), *Structures, Strategies and Beyond: Essays in Honour of Adriana Belletti* (Vol. 223, pp. 343–368). John Benjamins.

Tonelli, L., Marco, A., Vollmann, R., & Dressler, W. U. (1998). Le prime fasi dell'acquisizione della morfologia. *Parallela*, *6*, 281–301.

Volpato, F., Tagliaferro, L., Verin, L., & Cardinaletti, A. (2013). The Comprehension of (Eventive) Verbal Passives by Italian Preschool Age Children. *Advances in Language Acquisition (Selected Proceedings of GALA, 2011)*, 243–250.

Volpato, F., Verin, L., & Cardinaletti, A. (2014). The Acquisition of Passives in Italian: Auxiliaries and Answering Strategies in an Experiment of Elicited Production. *New Directions in the Acquisition of Romance Languages*, 371–394.

L1 Influence on Chinese English Learners' Use of Spatial and Metaphorical Senses of the English Preposition "IN". A Learner Corpus Study

Lingling Xu (University of Birmingham, United Kingdom)

This study explored how the accuracy of Chinese English learners' use of the preposition "IN" is influenced by their first language (L1). "IN" was selected mainly because it is highly polysemous thus posing a huge difficulty for English learners. Besides, previous studies have predominantly focused on spatial senses (e.g., Alonso et al., 2016) while metaphorical senses have not received equal attention. The focus was on both the spatial and metaphorical senses of "IN", and the potential impact of the overlap between these senses and their Chinese equivalents .

The research question the study specifically addressed is:

Does the semantic overlap between the individual senses of "IN" and its Chinese equivalent "Li" influence the accuracy of Chinese English learners' use of "IN"?

I drew data from EF-Cambridge Open Language Database (EFCAMDAT; Geertzen et al., 2013). It includes approximately 1,200,000 writings in L2 English. Most of the writings have been error-annotated, which makes it straightforward to identify errors. In the present study, I only targeted the error-annotated writings by Chinese English learners.

Since the data was highly imbalanced, I down-sampled accurate uses. Specifically, I first randomly identified 200 correct uses and 200 errors. Each correct use and error was coded in terms of "Accuracy" (Correct/Incorrect), "Sense Conveyed" based on i Ferrando's (2000) taxonomy of the senses of "IN", "Sense-Type" (spatial/metaphorical), "Corresponding-Sense-Chinese" (whether the Chinese equivalent expression shares the sense conveyed) and "Corresponding-Expression-Chinese" (whether the Chinese equivalent expression can be used to express the same meaning as the original English expression).

To address the question, I modeled "accuracy" as a function of various predictors. Our models showed that L1 influence may be more prominent on the sense level than on the expression level. However, this is largely driven by the difference between spatial and metaphorical senses. When only metaphorical senses were targeted, no such association was identified.

References

Alonso, R. A., Cadierno, T., & Jarvis, S. (2016). Crosslinguistic Influence in the Acquisition of Spatial Prepositions in English as a Foreign. Crosslinguistic influence in second language acquisition, 95.

Ferrando, I. N. (2000). A cognitive-semantic analysis of the English lexical unit" in". Cuadernos de investigación filológica, 26, 189-220.

Geertzen, J., Alexopoulou, T., & Korhonen, A. (2013, October). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). In Proceedings of the 31st Second Language Research Forum. Somerville, MA: Cascadilla Proceedings Project (pp. 240-254).

Using Lexical Complexity Measures to Predict Grades in Student Writing

Christian Holmberg Sjöling (Luleå University of Technology, Sweden)

Writing is one of the key components of language proficiency. To a large extent it is dependent on vocabulary knowledge, as proven by the fact that lexically complex texts are often considered as being of high quality by assessors. Every year in Sweden, upper secondary school students are required to take the national tests of English, which are created by the Swedish National Agency for Education (SNAE) and intended to establish to what extent the students' proficiency is in line with the course expectations. This paper aims to study the role that lexical complexity measures play in the grading of such texts written by Swedish upper-secondary school students. The data consist of one learner corpus of graded example texts (n=142) provided by SNAE to teachers in the assessment instructions to illustrate how the tests are to be assessed and an additional learner corpus of student texts graded by teachers (n=175) during the actual exams. The assessment instructions indicate that there should be a cline from the lowest to highest grade in terms of lexical and phraseological complexity. Therefore, a wide range of lexical (e.g., word frequency, dispersion and diversity) and phraseological measures (e.g., n-gram register and association strength) were applied to discern if a sequential distribution between different grades exists. The analysis was carried out using TAALED (Kyle et al., in press) and TAALES (Kyle and Crossley, 2015). Preliminary results show that very few complexity features predict grade in any meaningful way, suggesting that lexical complexity is largely overlooked in the assessment of national tests in Sweden. The SNAE and practicing teachers also appear to value different aspects of lexical complexity, which may have consequences for their classroom practice and the development of students' writing proficiency. This discrepancy and its possible implications are further discussed in relation to ensuring a fair and reliable assessment practice.

References

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. Tesol Quarterly, 49(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S. A., & Jarvis, S. (in press). Assessing the validity of lexical diversity using direct judgements. Language Assessment Quarterly.

Effects of incorporating learner corpus data into EAP writing materials development: L1 Chinese students' awareness of the use of epistemic lexical verbs in L2 English

Xie Fei (University of Birmingham, United Kingdom)

Since the mid-1980s, there has been an increased interest in using corpus data to develop EAP teaching/learning materials. However, most researchers and materials developers mainly relied on native English-speaker corpora, and learner corpora still remained marginal. Nesselhauf (2004) argue that, for language learning, it is not only important to know what typical native language is, but what typical difficulties learners have. To achieve this purpose, the present study explored the value of learner corpora in EAP writing materials development in helping learners to increase their knowledge of epistemic lexical verbs (ELVs).

Firstly, a L1 English corpus and L1 Chinese learner corpus of English consisting of MA academic essays were compiled and compared the use of ELVs. After, two types of learning materials were developed: the first included data from L1 English corpus, the second added both L1 Chinese learner corpus data and prior corpus findings. These two materials were transferred to Qualtrics and allocated to the control group (CG) and experimental group (EG) of MA L1 Chinese students respectively. Their performance was assessed at the pre-, post-, and delayed post-tests; their processing time on each test and their IELTS overall score were used as dependent variables to test their relative effectiveness on learning performance. The quantitative findings were triangulated with data collected via questionnaire and semi-structured interview to examine learners' attitudes regarding incorporating learner corpus data in materials development.

A mixed-effects regression model showed that both EG and CG demonstrated a higher accuracy at post- and delayed post-test compared to pre-test. Also, both groups' learning performance increased at a rate of 0.10 IELTS overall score. However, no significant difference was identified between two groups at any three testing points. The qualitative data suggest over 85% students in both groups preferred to see learner corpus data and believed it would better to address their own problematic usage.

References

Nesselhauf. N. (2004). Learner corpora and their potential for language teaching. In J.M. Sinclair (Eds.), How to use corpora in language teaching (pp.125-156). John Benjamins Publishing Company.

Spoken learner language via videomediated communication: a multimodal corpus pragmatics approach

Gerard O' Hanlon (Mary Immaculate College, Limerick, Ireland, Ireland)

Video-mediated communication – commonplace in professional and social settings – also permits multimodal affordances in pedagogical contexts. All communication is multimodal (Jewitt et al., 2016) be it face-to-face or distant, spoken or written, synchronous or asynchronous (Adami, 2017). This involves modes such as gaze, gesture and nods combining to orchestrate (online) communicative events. Therefore, to fully understand language acquisition, it is necessary to investigate multimodality (Urbanski and Stam, 2022).

Most corpora are monomodal (i.e. transcribed/written data) (Knight & Adolphs, 2020). Multimodal corpora (i.e. the employment of additional layers of annotated metadata, e.g. gestures) are gaining traction thanks to technological advances (Chanier & Lamy, 2017). Thus, a video-mediated repository of spoken learner interaction will add greatly to learner corpus research.

Annotation is common to both multimodality and corpus pragmatics. It is held that pragmatics should be further incorporated into language learning pedagogy (Bardovi-Harlig, 2017). This PhD project therefore aims to highlight pragmatic features in spoken learner interactions using a multimodal corpus approach.

I address the following questions:

-How do English language learners perform request sequences in pairs during spoken video-mediated interactions?

-What role does multimodality play in the competence and performance of these L2 pragmatic interactions?

This research gathers a video-mediated dataset via Zoom and analyses it using AntConc and ELAN. Four adult English language learners at B2 level participated in the initial study. Open roleplays were used to elicit spoken request sequences. The data were transcribed and manually annotated for spoken pragmatic and embodied features of requests (i.e. gesture, head, face, gaze) in ELAN.

Preliminary results show that English language learners employ multimodal potentialities in online dyadic video-mediated interactions to augment requests, including paralanguage (e.g. pauses, loudness, rapid speech), gesture (literal, deictic, stress and non-literal movements), gaze patterns related to politeness and imposition, and facial expressions signalling dispreferred responses. Findings from this research have potential applications in the fields of EAP and teacher training, for example.

References

Adami, E. (2016) 'Multimodality' in Garcia, O., Flores, N. and Spotti, M., eds., Oxford Handbook of Language and Society, Oxford: Oxford University Press, 451–472

Bardovi-Harlig (2017) 'Acquisition of L2 Pragmatics' in Loewen, S. and Sato, M., eds., The Routledge Handbook of Instructed Second Language Acquisition, London and New York: Routledge, 224-245

Chanier, T. and Lamy, M. (2017) Researching Technology-mediated Multimodal Interaction, in Chapelle, C.A. and Sauro, S. (eds), The Handbook of Technology and Second Language Teaching and Learning, Oxford: Wiley Blackwell 428-443

Jewitt, C., Bezemer, J. and O'Halloran, K. (2016) Introducing Multimodality, London: Routledge

Knight, D. and Adolphs, S. (2020) 'Multimodal Corpora' in Paquot, M. and Gries, S.Th., eds., A Practical Handbook of Corpus Linguistics. Cham: Springer International Publishing, 353-371

Urbanski, K. and Stam, G. (2022) 'Overview of multimodality and gesture in second language acquisition', in Stam, G. and Urbanski, K. eds., Gesture and Multimodality in Second Language Acquisition, A Research Guide. London: Routledge. 1-25

Investigating Spelling Errors in High-Functioning Dyslexic and Non-Dyslexic Learners of English as a Foreign Language

Radar Laurie (UCLouvain, Belgium)

In this presentation, we compare the proportion and the nature of spelling errors made by learners of English as a foreign language (L2) with dyslexia to those made by a control group. Our approach relies on data representing both the written product and the writing process.

Dyslexia is commonly associated with spelling difficulties, which persist into adulthood (Callens et al., 2012), even among high-functioning dyslexics (Gallagher et al., 1996), i.e. adults who were diagnosed with dyslexia in childhood but have compensated sufficiently to undertake higher education. Most studies focusing on the impact of dyslexia in writing concern children and L1, and the few studies focusing on adults show contradictory findings (Tops et al., 2014). It therefore seems relevant to investigate spelling errors and their nature in high-functioning dyslexics writing in L2 as dyslexia-related difficulties may be more pronounced in an L2 (Helland & Kaasa, 2005).

In line with previous research on dyslexia (Tops et al., 2014), we distinguished three categories of spelling errors: phonological spelling errors, occurring when a word is inaccurately spelled based on its pronunciation, grammatical spelling errors, occurring when spelling violates a language-specific spelling rule, and orthographical spelling errors, occurring when an incorrect grapheme is used to represent a phoneme. Our objective is to investigate which type(s) of spelling errors is/are most prevalent and to what extent in several high-functioning dyslexic learners and non-dyslexic learners based on short narrative texts written in L1 and L2.

The data collected are part of PROCEED (Gilquin, 2022), a learner corpus of English that includes written texts, but also keylogging (Inputlog; Leijten & Van Waes, 2013) and screencasting data (OBS Studio; Jim & OBS Studio Contributors, 2021), making the writing process visible. Stimulated recall sessions are also performed to get a better understanding of learners' cognitive activities during the writing task.

References

Callens, M., Tops, W., & Brysbaert, M. (2012) Cognitive profile of students who enter higher education with an indication of dyslexia. PloS one, 7(6), e38081.

Gallagher, A. M., Laxon, V., Armstrong, E., & Frith, U. (1996) Phonological difficulties in high-functioning dyslexics. Reading and Writing, 8, 499-509.

Gilquin, G. (2022) The Process Corpus of English in Education: Going beyond the written text. Research in Corpus Linguistics, 10(1), 31-44.

Helland, T., & Kaasa, R. (2005) Dyslexia in English as a second language. Dyslexia, 11(1), 41-60.

Jim & OBS Studio Contributors (2021) OBS Studio. https://obsproject.com.

Leijten, M., & Van Waes, L. (2013) Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. Written Communication, 30(3), 358-392.

Tops, W., Callens, M., Bijn, E., & Brysbaert, M. (2014) Spelling in adolescents with dyslexia: Errors and modes of assessment. Journal of Learning Disabilities, 47(4), 295-306.

Effects of Shared First Language and Speakers' First Language on Comprehensibility of Paired Interaction

SungEun Choi (Northern Arizona University, United States of America)

Few spoken learner corpora are easily accessible to the public and young researchers. Given this reality, graduate students whose research interests include spoken interaction face difficulty developing study stimuli that include conversational data. Utilizing the paired interaction samples from the Wildcat Corpus of native- and foreign-accented English (XYZ), the current study investigated how the differing degrees of shared first language between listeners and speaker pairs affect L1-English listeners' comprehensibility. Specifically, the research questions the study aimed to explore are as follows: (1) To what extent does the shared L1 status affect English-L1 listeners' comprehensibility scores? (2) To what extent does the speakers' L1 background affect English-L1 listeners' comprehensibility scores? Speakers consisted of three groups: (a) L1-fully shared pairs (English-English pairs); (b) L1-partially shared pairs (English-Chinese pairs, English-Korean pairs); and (c) L1-not shared pairs (Chinese-Chinese, Korean-Korean, and Chinese-Korean pairs). Speakers participated in a spot-the-difference task called the Diafix task, which was designed to elicit a spontaneous dialogic interaction between two speakers. Thirty English-L1 listeners listened to the excerpts of paired interaction and provided the scalar ratings of their comprehensibility. A mixed-effects model indicated that the listeners rated the L1-fully shared group and the L1-partially shared group as highly comprehensible. However, their comprehensibility ratings were significantly lower when they listened to L1-not shared pairs ($p = .028$). Results have implications for language teaching and learning and spoken communications in global contexts. Furthermore, the study will give ideas for graduate students on applications of the currently available spoken corpora for designing a study on second language acquisition.

Reference

van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The wildcat corpus of native-and foreign-accented english: Communicative efficiency across conversational dyads with varying language alignment profiles. Language and Speech, 53(4), 510-540. https://doi.org/10.1177/0023830910372495

Learner Corpus Approach for Automatic Distractor Generation

Nikita Login (National Research University Higher School of Economics, Russian Federation)

According to (Granger, 2008), language testing is one of the areas of practical application of learner corpora. Automatic question generation (AQG) is one of the rapidly developing Natural Language Processing applications in education - a review of (Kurdi et al., 2020) indicates an increase in number of publications on AQG during recent years and mentions automatic distractor generation (ADG) as one of its most challenging subtasks. However, it is common to rely on pre-defined heuristics in ADG for language testing, while the usability of learner corpus data in this task remains a poorly researched area.

This paper describes a machine learning-based ADG approach (called DisSelector) for corpus-sourced EFL lexical questions derived from REALEC (Vinogradova & Lyashevskaya, 2022) corpus. For each of 3,000 examples of corpus lexical errors a set of candidate distractors was retrieved from other examples with the same correction word, and each candidate was manually labelled as a plausible or implausible distractor. A number of classification models (including classical machine learning algorithms and gradient boosting implementations) were trained on the data. Word and sentence vectors from BERT (Devlin et al., 2019) and Word2Vec (Mikolov et al., 2013) models together with corpus word frequencies were used as input features for the classifiers. The highest F1-score (0.72) in classification experiment was attained by a XGBoost (Chen & Guestrin, 2016) model. DisSelector showed an improvement over the unsupervised baseline during both automatic and manual evaluation.

The scores obtained by DisSelector are expected to increase with the enlargement of labelled training dataset and introduction of new methods of dataset filtering, feature extraction and model training. The plans also include comparing distractors generated by described models with distractors obtained by zero-shot learning technique from large language models, such as ChatGPT and GPT-3 (Brown et al., 2020).

References

Tianqi Chen, & Carlos Guestrin (2016). XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM.

Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American

Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics.

Granger, S. (2008). Learner Corpora.

Kurdi, S. (2020). A Systematic Review of Automatic Question Generation for Educational Purposes. International Journal of Artificial Intelligence in Education, 30(1), 121-204.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Proceedings of Workshop at ICLR, 2013.

Vinogradova, O., Lyashevskaya, O. (2022). Review of Practices of Collecting and Annotating Texts in the Learner Corpus REALEC. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds) Text, Speech, and Dialogue. TSD 2022. Lecture Notes in Computer Science(), vol 13502. Springer, Cham. https://doi.org/10.1007/978-3-031-16270-1_7

BILCC: A new resource for exploring L2 Chinese pragmatic competence

Alessia Iurato (Ca' Foscari University of Venice, Bremen University)

This paper examines the validity of a new resource for the study of the pragmatic competence in L2 Chinese: The Bimodal Italian Learner Corpus of L2 Chinese (BILCC; Iurato, in press A).

Given the lack of data from Italian learners in existing L2 Chinese corpora (Iurato 2022a; 2022b), the increasing number of L2 Chinese learners in Italy (Conti & Romagnoli, 2021), and the scarcity of pragmatically tagged learner corpora in LCR (Callies, 2015), BILCC was designed to fill these gaps by providing data from Italian learners for the study of the pragmatic knowledge of Chinese shì...de clefts.

Eight theoretically motivated tasks (Sinclair, 2005; Tracy-Ventura & Myles, 2015; Bell & Payant, 2021; Lozano, 2021) were designed to collect written and oral data: two discourse completion tasks and two picture-based narratives (written tasks); one closed role-play, two open role-plays, and one picture description (oral tasks). The tasks were completed by 103 Italian learners, grouped into beginner, intermediate, and advanced levels according to their HSK Chinese language proficiency test scores. 35 L1 Chinese speakers also completed the tasks for comparison purposes.

Considering that Chinese clefts have corrective contrastive and non-contrastive functions (Iurato, in press B), BILCC distinguishes two different pragmatic functions for annotating the corpus data at the pragmatic level: intensification and corrective contrast. Intensification refers to non-contrastive clefts where the focus has the original function of highlighting a piece of information (Garassino, 2016; Korzen, 2014). Contrast refers to contrastive clefts with a corrective contrastive focus (Hartmann & Veenstra, 2013).

The study will address the following research questions:

RQ1: Does the variability of the tasks used to compile BILCC allow us to observe differences in the linguistic competence of Italian learners at different proficiency levels in the use of Chinese clefts?

RQ2: Are there differences in the use of the construction between learners and native speakers?

Preliminary results show that shì...de clefts are used more by intermediate and advanced learners. Learners use more contrastive than non-contrastive clefts in both oral and written production, whereas native speakers tend to use clefts non-contrastively. Learners' poor L2 pragmatic competence is arguably caused by the challenges posed by the syntax-pragmatic interface, since the acquisition of information structures such as shì...de clefts requires learners' interlanguage to implement a cross-domain interaction between the syntactic module and the pragmatic system, as argued by the Interface Hypothesis (Sorace & Serratrice, 2009).

References

Bell, P. & Payant, C. (2021). Designing learner corpora. Collection, transcription, and annotation. In N. Tracy-Ventura and M. Paquot (eds), The Routledge Handbook of Second Language Acquisition and Corpora (pp. 53-67). London-New York: Routledge.

Callies, M. (2015). Learner corpus methodology. In S. Granger, G. Gilquin, F. Meunier (eds), The Cambridge Handbook of Learner Corpus Research (pp. 35-55). Cambridge: Cambridge Univeristy Press.

Garassino, D. (2016). Using cleft sentences in Italian and English. A multifactorial analysis. In A.-M. De Cesare & D. Garassino (eds), Current issues in Italian, Romance and Germanic non-canonical word orders. Syntax-information structure-discourse organization (pp. 181-204). Bern, Peter Lang.

Hartmann, K. & Veenstra, T. (2013). Introduction. In K. Hartmann & T. Veenstra (eds), Cleft structures (pp. 1-32). Amsterdam: John Benjamins.

Iurato, A. (2022a). Learner corpus research meets Chinese as a second language acquisition: Achievements and challenges. Annali di Ca' Foscari. Serie Orientale, 58(1): 709-742.

Iurato, A. (2022b). Analyzing Chinese learner corpus research and Chinese learner corpora: Main features, critical issues and future pathways. Kervan. 26(1): 531-562.

Iurato, A. (in press A). Designing and compiling the written sub-corpus of the Bimodal Italian Learner Corpus of Chinese (BILCC): Methodological issues. In S. Zuccheri (ed.), Studies on Chinese Language and Linguistics in Italy (pp. 197-228). Bologna: Bologna University Press.

Iurato, A. (in press B). Exploring the pragmalinguistic knowledge of the 是 shì...的 de cleft construction in L1 Italian learners' L2 Chinese: Triangulation of corpus and experimental data. In I. Kecskes & H. Zhang (eds), Chinese as a Second Language from Different Angles. Leiden: Brill.

Korzen, I. (2014). Cleft sentences. Italian-Danish in contrast. In A.-M. De Cesare (ed.), Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic: Contrastive, Corpus-Based Studies (pp. 217-275). Berlin, De Gruyter.

Lozano, C. (2021). CEDEL 2: Design, compilation, and web interface of an online corpus of L2 Spanish acquisition research. Second Language Research, 1-19. DOI: 10.1177/02676583211050522.

Romagnoli, C. & Conti, S. (eds) (2021) La Lingua Cinese in Italia. Studi su Didattica e Acquisizione. Roma: Roma Tre Press.

Sinclair, J. (2005). Corpus and text – basic principles. In M. Wynne (ed.), Developing linguistic corpora: A guide to good practice (pp. 1-16). Oxford: Oxbow Book Company.

Sorace, A. & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. International Journal of Bilingualism, 13(2): 195-210.

Tracy-Ventura, N. & Myles, F. (2015). The importance of task variability in the design of learner corpora for SLA research. International Journal of Learner Corpus Research, 1(1): 58-95. doi 10.1075/ijlcr.1.1.03tra issn 2215–1478.

Prepositional phrases as postmodification patterns of the noun phrase in learner English writing across cohorts

Pamela Saavedra-Jeldres (The University of Warwick, United Kingdom)

This is a learner corpus study in progress, and it analyses the grammatical complexity of the noun phrase, based on Biber, et al's hypothesised developmental index (2011). This presentation focuses on features of postmodification, namely prepositional phrases starting with 'of' and 'other prepositions' as described by the framework, and how these features are associated with the learners' cohorts (year 3, 4, 5 undergraduate programme) in academic writing as a foreign language. Dataset were taken from the CELTEC, a 246,808-word learner corpus compiled with the writings of future teachers of English in Chile for the purpose of this study. Preliminary results indicate that prepositional phrases starting with other prepositions than 'of' is the single most common modification pattern in this corpus across the three cohorts with a frequency of occurrence of 56.3 normalised to a thousand words. Besides, a slow increase in the frequency of occurrence can be observed in the three cohorts: year 3 has a frequency of 54, year 4 of 56, and year 5 a frequency of 60.5 prepositional phrases per one thousand words. Additionally, prepositional phrases starting with 'of' has an overall frequency of occurrence of 19.3 which makes it the fourth most frequent modification pattern in CELTEC. Similarly, prepositional phrases starting with 'of' show a moderate but steady increment in the frequency of occurrence across cohorts: year 3 with 16.6; year 4 with 19.1; and finally, year 5 a frequency of 23 prepositional phrases per one thousand words. Preliminary, postmodification in the noun phrase by means of prepositional phrases seems to increase both in frequency and length across cohorts.

Development of word formation regularities by learners of German as a foreign/second language

Christine Marie Coca (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany)

Recognizing word formation regularities and using them to comprehend and create complex words is not a particular challenge for native speakers. This word formation competence is otherwise crucial for learners of German as a second language. Its importance has long been justified in the specialist literature (cf. Storch/Storch-Luche 1979, Thurmair 1997, Storch 2001, Simeckova 2004, Mac 2009, Gärtner 2012) primarily by the economic possibility of vocabulary expansion. However, acquiring the competence of word formation is impeded by the multitude and complexity of word formation rules, as well as the numerous irregularities and polyfunctionalities of the German word formation. Furthermore, the lack of sufficient qualitative and quantitative data from learner language research hinders the establishment of a systematic word formation didactic.

Previous studies on word formation in learner languages represent exclusively cross-sectional studies. Berth 2009, Zeldes 2013 and Coca 2021 showed, based on the corpora of the Falko family, a significant underuse of the investigated word formation patterns, although advanced learners of German as a second language can usually recognize and use them productively.

The question of the development of such word formation regularities remains unanswered due to the data basis of the corpora studied thus far. However, it is of great importance especially regarding the resulting implications for the didactics of German as a second language. Within the framework of a pseudo-longitudinal study, this development will be outlined on the basis of the most productive native adjective-forming suffixes of German.

As a corpus basis, texts from the "Written Text Production" section of the DSH examination (Deutschen Sprachprüfung für den Hochschulzugang) were used. Three subcorpora were created from a complete semester data set of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), based on the three DSH levels. The primary objective of the project is to enable an investigation of the development of the word formation competence of the learners.

The analysis includes quantitative and qualitative aspects. Contrastive Interlanguage Analysis (CIA) is used to compare the productive use of the investigated suffixes at individual

developmental levels. Computer-Aided Error Analysis (CEA) is used to examine error types in the developmental levels, capturing the qualitative development of word formation competence in the adjective domain. The compilation and annotation of a learner corpus presents a particular challenge.

References

Berth, Michael (2009): Treffungen, Sinkung und Benützung. Korpuslinguistische Untersuchung des Erwerbs von derivationsmorphologischen Wortbildungsregularitäten bei fortgeschrittenen Lernern des Deutschen als Fremdsprache. Magisterarbeit HU Berlin. URL: http://e-doc.hu-berlin.de/master/berth-michael-2009-0917/PDF/berth.pdf.

Coca, Christine (2021): Die Allomorphe -heit, -keit und -igkeit in der Lernersprache. Eine korpuslinguistische Untersuchung bei fortgeschrittenen Lernern des Deutschen als Fremdsprache. Masterarbeit FAU.

Gärtner, Angelika (2012): Wortbildung: Problemfelder im DaF-Unterricht. In: Info DaF 4/39, S. 499-513.

Mac, Agnieszka (2009): Einige Überlegungen zur Wortbildungslehre im fremdsprachlichen Deutschunterricht. In: Glottodidactica 35, S. 101-117.

Simeckova, Alena (2004): Zur jüngeren germanistischen Wortbildungsforschung und zur Nutzung der Ergebnisse für Deutsch als Fremdsprache. In: Deutsch als Fremdsprache 41/3, S.140-151.

Storch, Günther (2001): Deutsch als Fremdsprache – eine Didaktik. München: Wilhelm Fink Verlag.

Storch, Günther/Storch-Luche, Monika (1979): Wortbildungsübungen im Fremdsprachen-unterricht. In: Zielsprache Deutsch 10/4, S. 11-23.

Thurmair, Maria (1997): Nicht ohne meine Grammatik! Vorschläge für eine Pädagogische Grammatik im Unterricht Deutsch als Fremdsprache. In: Jahrbuch Deutch als Fremdsprache 23, S. 25-45.

Zeldes, Amir (2013): Komposition als Konstruktionsnetzwerk im fortgeschrittenen L2-Deutsch. In: Zeitschrift für germanistische Linguistik 41, S. 240-276.

The influence of L1 Dutch on cohesion in L2 German writing: Results from a contrastive corpus- based analysis of L1 and L2 students´ writing in German

Helena Wedig (University of Antwerp, Netherlands)

Second language (L2) writers struggle with cohesion (Crossley & McNamara, 2012). One of the reasons is that they tend to rely on native language (L1) strategies to create cohesive texts which may differ from the strategies in the L2 (Breindl, 2016). In stark contrast to the burgeoning research into cohesion in L2 English (e.g., Appel & Szeib, 2018; Crossley & McNamara, 2011) research into L2 German has been scarce, with only a handful of studies contrasting L1 texts with L2 texts by students with diverse L1s (e.g., Walter, 2007; Breindl, 2016). Whereas these studies investigate general characteristics of L2 German, we lack studies contrasting L2 texts produced by homogenous L1 groups with those produced by heterogeneous L1 groups. My study aims to fill this gap by performing a contrastive analysis of connectives in L2 German, comparing L2 writers with L1 Dutch and L2 writers with diverse L1s. The analysis is based on the Belgisches Deutschkorpus (Beldeko) (Strobl, 2020) and the German Summary Corpus (GerSumCo). Beldeko consists of 301 L2 German texts written by students with L1 Dutch. GerSumCo is still growing and to date contains 102 texts. Both corpora contain summaries of the same source texts that were produced under comparable conditions. Automated annotations for connectives were added to the corpora, which were manually corrected by three trained annotators based on our guidelines. First preliminary results show that writers with L1 Dutch use more temporal and expansion connectives than contingency and comparison connectives. In my presentation, I show the results of the comparison of this connective use in Beldeko with the connective use in GerSumCo. Based on previous research, I expect to find a different density and distribution of connectives depending on the L1s of the writers (as seen in Breindl, 2016).

The StuWiss Corpus as a Base for Studying Formulaic Language in Student Academic Writing in L1 and L2 German – Considerations and Challenges in Corpus Design and Annotation

Andrea Lösel (Leipzig University, Germany)

The StuWiss Corpus as a Base for Studying Formulaic Language in Student Academic Writing in L1 and L2 German – Considerations and Challenges in Corpus Design and Annotation The PhD project presented explores the formulaic lexical resources L1 and L2 students draw on in their academic writing in German. For this, the StuWiss corpus (Studentisches Wissenschaftliches Schreiben – Student Academic Writing) is being built. It currently consists of 208 master theses (4.8 million tokens), submitted to five German as a Foreign Language (GFL) departments at German universities. The collection of a comprehensive metadata set, with information on the master's thesis and the educational and linguistic background of each submittee, allows for the composition of sub corpora, specifically the division of the main corpus into two similarly large L1 and L2 sub corpora. The first goal of the PhD project is to determine the inventory of multi-word units that students possess overall, but also in L1 and L2 comparison. This quantitative (statistical) access via linear and syntactic n-grams is complemented by qualitative (syntactic and functional) analyses of the found multi-word units. These analyses can provide insights into the formal morphosyntactic nature as well as the functional domains and usage specifics of the multi-word units in student academic writing – specifically in L1 and L2 comparison. The paper presented at the LCR Graduate Student Conference will place emphasis on the considerations and challenges of designing, compiling and preparing the StuWiss corpus, from the data collection to the data cleaning to the data annotation, including syntactic dependency annotation, to the evaluation of data quality leading up to the first test runs computing n-grams.

References

Andresen, Melanie; Zinsmeister, Heike (2017): The Benefit of Syntactic vs. Linear N-Grams for Linguistic Description. In: Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017). Linköping: Linköping University Electronic Press, S. 4–14.

Brommer, Sarah (2018): Sprachliche Muster. Eine induktive korpuslinguistische Analyse wissenschaftlicher Texte. Berlin; Boston: de Gruyter.

Chen, Yu-Hua; Baker, Paul (2010): Lexical Bundles in L1 and L2 Academic Writing. In: Language Learning & Technology 14/2, S. 30–49. Jaworska, Sylvia; Krummes, Cedric; Ensslin, Astrid (2015): Formulaic Sequences in Native and NonNative Argumentative Writing in German. In: International Journal of Corpus Linguistics 20/4, S. 500– 525.

A Validation Study of Automated Analysis of Noun Phrase Complexity

Ichika Yamaguchi (Tokyo University of Foreign Studies, Japan)

Automated tools for writing developmental measurements are increasingly used for analyzing various types of learners' production in recent years, even though these tools were originally developed and validated for texts produced by advanced learners. The specific characteristics found in the text produced by less advanced learners, such as grammatical errors and improper punctuation usage, could potentially affect the accuracy and reliability of these automated tools. This study aims to examine the reliability of indices of noun phrase complexity computed in TAASSC, Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (Kyle, 2016) for L2 English produced by learners ranging from A1 to B2 level on the CEFR scale. To this end, manual and automated analyses of 100 texts from the EF-Cambridge Open Language Database (EFCAMDAT) (Geertzen et al., 2013) are compared. The target indices include measurements of adjectival modifiers per nominal, prepositions per nominal, verbal modifiers per nominal and relative clause modifiers per nominal. Based on the quantitative comparison, this study demonstrates how the accuracy of automated analysis of noun phrase complexity is affected by the proficiency levels of the learners. Furthermore, through both the quantitative and qualitative analyses, this study identifies which specific features found in learners' text are causing a decrease in accuracy. Based on the results, it is further discussed how to improve the accuracy of automatic computation of indices through the pre-processing steps prior to automated analyses, in which raw data are cleaned and transformed into a way that is suitable for further analysis. The findings of this study will inform future research that seeks to identify criterial features for distinguishing learners of different proficiency levels based on the characteristics of noun modification.

Complexity Development of Intermediate German Learners of English: A Longitudinal Corpus Analysis

Philine Kim Tschirner (Philipps-University Marburg, Germany)

Complexity is a topic of wide range, overall separated into various levels, namely lexis (Linnarud, 1986), morphology (Brezina & Pallotti, 2019), syntax (Lee, 2004) and phraseology (Paquot, 2019). It is generally assumed that syntactical complexity of a language correlates positive with overall language competence and development. (Bulté 2008 & Paquot, Naets, Gries 2021) For example, length of T-Units, clauses per T-Unit or dependent clauses per T-Unit. (Lee 2004: 108) Previous studies already treated the topic, whereas their corpora were much smaller. (Kyle, Crossley, Verspoor, 2021). This given example also worked longitudinal and examined those students throughout years.

The present project therefore poses the following research question: How does the quantitative and qualitative linguistic complexity in written English of intermediate learners of a German high school develop between 9th and 12th grade?

To answer these research questions, the "Marburg Corpus of Intermediate Learner English" (MILE; Kreyer, 2015) gives the opportunity to analyze data of 90 students from 9th to 12th grade. Within more than 500,000 words in total, the MILE corpus for the first time gives researchers the opportunity for conducting a truly longitudinal analysis of a large number of intermediate learners of English over four years. Also, metadata such as gender and age was made available. This corpus will be analyzed using the "Tool for Automatic Analysis of Syntactic Sophistication and Complexity" (TAASSC; Kyle, 2016)

The method that will be used replicates previous studies connected to syntactic complexity (Crossley & McNamera, 2012) and is called the "bottom-up" principle, which means that data will be analyzed in the first step and will be put into bigger context in the second. The project analyses the complexity development considering Lu's 14 parameters of measuring complexity (Lu 2010: 479).

Results are expected to show the impact of given metadata on complexity development and will be discussed in the light of their language-pedagogical. Finally, the project results in an individual and multifactorial learner curve.

References

Brezina, Vaclaw; Pallotti, Gabriele (2019). Morphological Complexity in Written L2 Texts. Second Language Research. 35 (1): 99-119

Bulté, Bram; Housen, Alex; Pierrad, Michel; Van Daele, Siska (2008). Investigating Lexical Profieciency Development over Time – The Case of Dutch-Speaking Learners of French in Brussels. Journal of French Language Studies 18 (3): 277-298.

Crossley, Scott; McNamera, Danielle (2012). Predicting Second Language Writing Proficiency: The Roles of Cohesion and Linguistic Sophistication. Jounral of Research in Reading 35 (2): 115-135.

Kreyer, Rolf (2015). The Marburg Corpus of Intermediate Learner English (MILE). In Learner Corpora in Language Testing and Assessment, Marcus Callies & Sandra Götz, eds. Amsterdam: John Benjamins. 13-34.

Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication (Doctoral Dissertation)

Lee, Jiyoujng (2004). Syntactic Complexity, Clausal Complexity, and Phrasal Complexity in L2 Writing: The Effects of Task Complexity and Task Closure. The Journal of Asia TEFL. South Korea: 108-124

Linnarud, Moira (1986). Lexis in Composition: A Performance Analysis of Swedish Learners Written English. Malmö: C.W.K. Gleerup.

Paguot, Magali (2019). The Phraseological Dimension in Interlanguage Complecity Research. Second Language Research 35 (1): 121-145.

Paquot, Magali; Naets, Hubert; Gries, Stefan (2021). Using Syntactic Co-occurrences to Trace Phraseological Complexity Development in Learner Writing: Verb + Objekt Structures in LONGDALE. In: LeBruyn, Bert Simonne Walter; Paquot, Magali (hrsg): Learner Corpus Research Meets Second Language Acquisition. Cambridge: Cambridge University

Comparative analysis of collocations of the Spanish verb echar in the learner corpora CEDEL2 and CAES: errors and L1 influence

Maria Annese (Università degli Studi G. d'Annunzio Chieti-Pescara, Italy)

Phraseology is known to be a problematic domain for learners of foreign languages (Mendizábal de la Cruz and Sastre Ruano, 2017); nevertheless, it constitutes an important area of language as it links it to the target culture (Castillo Carballo, 2003). The problems that might arise in the process of acquiring phraseological units concern their recognition and subsequent use in a pragmatically appropriate way: at the latter phase, in particular, errors of interlinguistic nature might arise (Orol González and Alonso Ramos, 2013). Against this background, this poster aims to investigate the use of phraseological units based on the Spanish verb echar (Martín Salcedo, 2015) in an analysis comparing its occurrences in two Learner Corpora: CEDEL2 and CAES. In particular, it aims to determine whether and to what extent L1 affects the acquisition of this verb and the level of the CEFR at which verbal phraseological units in which the verb echar appears are acquired. This is followed by a brief analysis of the phraseological errors related to this verb (Pérez Serrano, 2014). Regarding the frequency of errors, a substantial decrease in intermediate levels of language proficiency was hypothesized. As for interlinguistic influence, examples related to two L1s, namely English and Italian, have been analyzed. In this regard, a lower frequency of errors was expected in the Italian-Spanish interlanguage due to the proximity between the two languages (Bailini, 2016). The possible applications of this study concern the development of targeted and specific teaching activities for each level of language proficiency (Timofeeva Timofeev, 2013; Mendizábal de la Cruz and Sastre Ruano, 2017; Peramos Soler et al., 2010) to foster the acquisition of the phraseological units containing echar.

References

Orol González A., Alonso Ramos M. (2013) A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish, in Procedia - Social and Behavioral Sciences, Volume 95, Pages 563-570, https://doi.org/10.1016/j.sbspro.2013.10.683.

Martín Salcedo, J. (2015), Échale ganas que te echamos una mano. Fraseología con el verbo echar, in V CONGRESSO NORDESTINO DE PROFESSORES DE ESPANHOL, 2014, Teresina, Anais do V Congreso Nordestino de Profesores de Espanhol, Brasilia: Ministerio de Educación, Cultura y Deporte.

Castillo Carballo, M. A. (2002), Conocimiento cultural en la adquisición de la L2: la fraseología, in El Español, Lengua del Mestizaje y la Interculturalidad, Actas del XIII Congreso Internacional de la ASELE, Murcia.

Bailini, S. (2016), La interlengua de lenguas afines: el español de los italianos, el italiano de los españoles, Milano: Led Edizioni.

Mendizábal de la Cruz, N., Sastre Ruano, M. Á. (2017), Problemas de las unidades fraseológicas verbales y su aplicación a la enseñanza del español como lengua extranjera, in Palabras Vocabulario Léxico: La lexicología aplicada a la didáctica y a la diacronía, Florencio del Barrio de la Rosa, Pages 49-62.

Pérez Serrano, M. (2014), Análisis de errores colocacionales en un corpus de aprendientes de ELE, in marcoELE. Revista de Didáctica Español Lengua Extranjera, 19 (2014), Redalyc, https://www.redalyc.org/articulo.oa?id=92152427011

Timofeeva Timofeev, L. (2013), La fraseología en la clase de lengua extranjera: ¿misión imposible?, in Onomázein, 28, december, Pages 320-336, Santiago, Chile: Pontificia Universidad Católica de Chile.

Peramos Soler, N., Leontaridi, E. & Ruiz Morales, M. (2009), Las unidades fraseológicas del español: su enseñanza y adquisición en la clase de ELE, in J.F. Barrio Barrio (coord.), Actas de las Jornadas de Formación del Profesorado en la Enseñanza de L2/ELE y la Literatura Española Contemporánea, Sofía: Ministerio de Educación de España y Universidad de Sofía “San Clemente de Ojrid”, 185-204. http://www.educacion.es/exterior/bg/es/publicaciones/actas_sofia_3_junio_web.pdf

Writing Revisions of German Learners of English: A Longitudinal Corpus Based Approach

Helena Hanneder (Philipps-Universität Marburg, Germany)

As part of a PhD project, we will analyse 9000 revisions of students from the truly longitudinal Marburg Corpus of Intermediate Learner English (MILE). It includes handwritten exams from 94 German secondary school students from grades nine to twelve. The data calls for larger quantitatively driven investigations of the sections the students crossed out and how they revised them, as this can tell us about the interlanguage development of the learners. We will focus on two research questions from a longitudinal perspective: In which areas did students revise their writing (content, form, mechanics)? Were their formal revisions successful? We will use a general logistic regression to find the effects of type of revision and revision length on revision success. For that, we will reveal the underlying layers of the final product to be able to see what the final text is hiding:

(1) Mr. Schwarzenegger has a positiv opinion of the USA and *from* California. (0049-2-11-00006)

In 1, we would not have known that the student was struggling with the preposition. Only a process-based approach can reveal the grammatical constructions the students were struggling with or partly aware of, as they improve or worsen their writing through revisions. This poster will briefly discuss how the data was annotated with a taxonomy based on previous ones (Lindgren 2005, Faigley & Witte 1980), while focusing on results from the data annotation of 2000 revisions, accomplished up to this point in the PhD programme. The most important findings were that students were more concerned with content than previously expected and they could reivse predominantly successfully. Statistically significant findings showed that spelling revisions were more successfully corrected than revisions concerning abbreviations and multiple word units. The results show promising preliminary findings that could help in the description of the interlanguage of intermediate learners.

References

Bonk, C. J., & Reynolds, T. H. (1992). Early adolescent composing within a generative-evaluative computerized prompting framework. Computers in Human Behavior, 8(1), 39–62.

Faigley, L., & Witte, S. (1981). Analyzing Revision. College Composition and Communication, 32(4), 400–414.

Kreyer, R. (2019). “Dear [Man men and women] madam, dear xxx sir”— What we can learn from revisions in authentic learner texts. Corpus Linguistics, Context and Culture. De Gruyter. 63-386.

Bridging borders through study abroad: Introducing the Corpus of Written Spanish- L2 and Heritage- Study Abroad (COWS-L2H-SA)

Sophia Minnillo (University of California, Davis, United States of America)

Few learner corpora have captured language learning longitudinally during study abroad (SA), with the notable exception of the LANGSNAP corpus (Mitchell et al., 2017). SA has been found to provide opportunities for enhanced language learning, although the affordances of SA may vary based on program features and students' individual differences (Pérez-Vidal & Sanz, 2023). Given the heterogeneity of learning outcomes between different programs and students, it is imperative that SA-focused learner corpora capture a greater variety of program types and student profiles.

To fill this gap, this poster will present the Corpus of Written Spanish- L2 and Heritage- Study Abroad (COWS-L2H-SA), an addition to COWS-L2H (Yamada et al, 2020) that includes students from the same U.S. university who study abroad for 2.5 months. The students participate in one of two Spanish language-focused programs in Argentina or Spain. The participant sample includes both L2 and heritage language learners and participants from beginner, intermediate, and advanced proficiency levels. The corpus showcases writing that students complete at the beginning, middle, end, and two months after their SA program. In the task, students make requests via emails written to interlocutors differing in social distance– a professor and a student.

In the poster, I will present preliminary findings about students' longitudinal writing development during and after SA in terms of (1) global communicative effectiveness (based on expert ratings) and (2) lexical complexity, which encompasses lexical diversity (mean textual lexical diversity; MTLD), lexical density (ratio of content to total words), and lexical sophistication (mean frequency of content words in the EsPal reference corpus; Duchon et al., 2013). By contextualizing students' writing development through information about their learner profiles and engagement with features of their SA program, this study increases understanding of the diverse manners in which language learning can occur (un)successfully during SA.

References

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. Behavior Research Methods, 45(4), 1246-1258.

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). Anglophone students abroad: Identity, social relationships and language learning. Routledge.

Pérez-Vidal, C., & Sanz, C. (Eds.) (2023). Methods in Study Abroad Research: Past, present, and future. John Benjamins.

Yamada, A., Davidson, S., Fernández-Mira, P., Carando, A., Sagae, K., & Sánchez-Gutiérrez, C. (2020). COWS-L2H: A corpus of Spanish learner writing. Research in Corpus Linguistics, 8(1), 17-32.

Italian word lists' representativeness of student writing: a corpus-based study

Alessandra Pierantoni (Alma Mater Studiorum - Università di Bologna, Italy)

To what extent can word lists be used as a tool to summarize language for teaching purposes? Word lists have been a research topic for many decades now, but despite their usefulness as a learning tool for students (see e.g. Durrant, 2016), they also have some limitations. One of them is the fact that studying the words using only lists where words are isolated does not give credit to the pedagogical importance of "constellations of words" (Palmer, cited in McArthur, 1998), i.e., words that should be taught together.

Another issue concerns the historically limited availability of student corpora focusing on academic writing (Nesi & Gardner, 2012), giving few possibilities to fully understand student vocabulary choices. This is one possible reason behind the fact that most lists, including academic word lists, are constructed based on the texts that students should read, not those they write. Another factor is that these lists were initially aimed at L2 students; with (advanced) L1 students they could be seen not only as a target to achieve, but also as an instrument to be verified and updated in light of possible language change.

In this paper, I will specifically explore how relevant Italian word lists are for texts written by native speakers, in particular if the words present in the lists are representative of the way young people write. To do this, I will use ITACA (https://itaca.eurac.edu/), a corpus of texts written by L1 Italian students in the province of Bolzano. These texts will be profiled by using the Academic Italian Word List (Spina, 2010) and the Vocabolario di base ("Core Vocabulary", De Mauro, 2015). I expect the results to show differences between the words used by the students and the word lists, leading to the conclusion that academic language is changing.

References

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? English for Specific Purposes, 43: 49-61.

McArthur, T. (1998). Living words: Language, lexicography and the knowledge revolution. Exeter: University of Exeter Press.

Nesi, H., Gardner, S. (2012). Genres across the disciplines: Student writing in higher education. Cambridge: Cambridge University Press.

A phraseological view to AI-powered writing assistant ChatGPT: A corpus-based study

Shuyuan Tu (Georgia State University, United States of America)

Recent advancements in artificial intelligence (AI) have opened doors for developing novel writing tools that enhance and create innovative ways to assist writers and learners during and after their writing process. Specifically, it can provide immediate and customized feedback and stimulate learners' interests and motivation (Huang et al., 2022; Kuhail et al., 2023). Studies in modern corpus linguistics have demonstrated that language is highly patterned and primarily comprises fixed or semi-fixed units (e.g., Biber, 2009; Hoey, 2005; Sinclair, 1991; Roemer, 2009) that serve critical discourse functions (Tan & Roemer, 2022). Therefore, this study intends to investigate ChatGPT's reliability by analyzing the ChatGPT- edited essays and comparing it with expert editors from a phraseological view.

The L2 learner essays in this study were extracted from the Edited Essays module from the International Corpus Network of Learners of English (ICNALE) (Ishikawa, 2018). Learner essays (n=140) and their fully edited versions (n=140) were selected for this study. Similarly, to test the reliability of ChatGPT in editing learner essays and determining if the process is consistent with human editors, ChatGPT was prompted using the same rubric and required to generate the fully edited versions (n=140) of the 140 original learner essays at once using ChatGPT API. The 3, 4, and 5 phrase-frames were identified based on the established frequency and range thresholds using AntGram (Anthony, 2019) and were manually filtered for meaningfulness. The variability and predictability in the filtered phrase-frames were analyzed, and functional analysis was further conducted. Tentative results indicate that ChatGPT contributes to more variable ($p<0.001$) and less predictable ($p<0.001$) 3 phrase-frames but does not show significant differences in frequency. This study could provide insight into the reliability of ChatGPT in editing essays in L2 writing by investigating its capacities. It could also raise awareness of the potential of AI-powered writing assistants in facilitating L2 writing from a phraseological perspective.