



The Graduate Student Conference in Learner Corpus Research 2021



Book of abstracts

11-12 October 2021

Conference website: <https://eng.inn.no/lcrgradconf>

Email: lcrgradconf@inn.no

Meet us on Twitter @lcrgradconf

#lcrgradconf



Welcome to Norway!

Yes, we know that you are not 'really' here in Hamar, Norway with us. But we have been thinking about you and this event for more than a year now, and we feel that you are here with us in spirit.



This conference is a child of the pandemic. As many of you know, the [Learner Corpus Association](https://www.learnercorpusassociation.org) (LCA) has held amazing biannual events since 2011. We last met physically for LCR2019, held at the University of Warsaw. The sixth LCR conference had then been scheduled for the fall of 2021, to be held at the University of Padua. The decision by the local organizing committee (LOC) of that event to postpone to 2022 was completely understandable given the uncertain times we have living in.



But we in Norway were beginning to experience withdrawal symptoms.

In brief, we ♥ LCR. A two-year gap between conferences is more than long enough. Three years is just too much. We felt need to return to the lively discussions and exchange of ideas typical for LCR conferences. When we approached the LCA with our suggestion of holding a virtual conference in the 'lost' year of 2021, they eagerly supported the idea, suggesting a new variant for the association. And so was born the first-ever Graduate Student Conference in Learner Corpus Research (LCRgradconf), proudly hosted by the Faculty of Education at Inland Norway University of Applied Sciences (INN): the first of what we hope will become a long tradition.

LCRgradconf is a virtual conference. This is not a measly second-best option. On the contrary, we see a [large number of advantages](#) to this format: lower cost (that is: FREE!) & greater accessibility mean that more people can attend from all over the world, with a substantially smaller carbon footprint. Our digital platform of course allows for spoken interaction, but also for written action through the 'chat' function-- a new dimension to conferences that enhances exchanges. And virtual conferences offer added convenience and comfort -- you can present your paper in slippers if you want!

We have pulled together a great program for you, filled with papers from Early Career Researchers (ECRs) currently working on their MAs or PhDs, in all different stages of the educational process. In addition, we have a number of ECR video posters available [HERE](#), an exciting format made possible by virtue of our virtual environment. And we will meet some of these poster presenters during a live event with the experts of their choice. Further, our live



conference events are neatly sandwiched by our two keynote speakers - Luciana Forti and Sylviane Granger - who between them will cover the span of learner corpus research: from ECR to Professor Emerita. And we have arranged all sorts of events for both ECRs and senior researchers, designed to give you plenty of opportunity to get to know each other, expand your networks, and have fun.

A virtual conference does not organize itself, and organization incurs expenses (see [here](#) for why). We are able to offer free registration to all conference delegates thanks solely to the generous sponsorship of the LCA and of INN's Faculty of Education. We also would like to thank the LCA board, our scientific committee, and all the LCA members who agreed to take an active role in our conference – as expert for our poster presenters, as 'chat' leader during our social events, and as active participants during the session. And also as keynote speaker! Thank you all.



The success of this conference is now up to YOU. Please read our four [simple tips](#) about how to be a good virtual conference delegate and follow them! Multitasking and conference participation do not mix well. Instead, put the **Gone Fishing** sign on your office door and spend two days with your fellow LCR researchers!

Happy conferencing!

The LCRgradconf local organizing committee

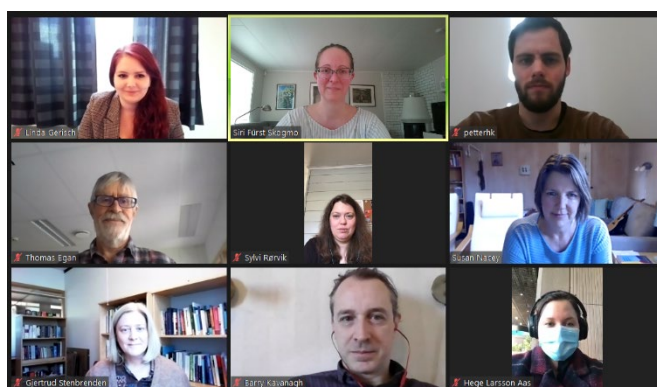
Siri Frst Skogmo

Susan Nacey

Linda Gerisch

Gjertrud Stenbrenden

Hege Larsson Aas



on behalf of the [English Language in Use](#) research group

Faculty of Education

Inland Norway University of Applied Sciences



From the President of the Learner Corpus Association

Ever since its emergence in the early 1990s, the study of L2 learning on the basis of learner corpora, a research strand that has come to be called Learner Corpus Research (LCR), has grown and matured significantly, attracting the interest of an increasing number of researchers in second language acquisition, foreign language teaching, language assessment and natural language processing. One telling sign of this development is the large number of students who have opted to write their MA or PhD theses in the field. The time therefore seemed ripe to organize an event which would give them the opportunity to present their research and receive feedback from their peers and more senior researchers in a relaxed, non-threatening atmosphere. And so the idea of holding an LCR Graduate Conference was born, and the first such conference was organized this year by the English Language in Use research group of Inland Norway University of Applied Sciences. The conference is taking place under the aegis of the Learner Corpus Association, an international association which aims to promote learner corpus research, to stimulate exchanges among scholars working in the field and to support the dissemination of findings to the broader scientific community. One of the key bonuses for young scholars is that membership, which is free for students, gives access to a comprehensive learner corpus bibliography that is updated on a continuous basis and currently features over 2,000 references.

The conference programme bears witness to the wide diversity of approaches and topics that are currently being investigated internationally. It features cross-sectional and longitudinal studies of a range of morphological, grammatical, lexical, phraseological and discourse features in L2 spoken and written learner corpora. The rapid expansion of LCR is evidenced by the number of L2 languages covered: while L2 English still predominates, other L2 language varieties are also being investigated (L2 Chinese, L2 Finnish, L2 German, L2 Japanese, L2 Russian and L2 Spanish). A positive hive of activity that bodes well for the future of learner corpus research.

In the name of the Learner Corpus Association I would like to express my warmest thanks to Susan Nacey and the whole conference organizing team for inaugurating what I am confident will be a long-running series of LCR Graduate Conferences.

Sylviane Granger

President of the Learner Corpus Association



Scientific committee

Ensuring the high quality of the LCR Graduate Student Conference

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

Gaëtanelle Gilquin, Université catholique de Louvain - FNRS

Sylviane Granger, Université catholique de Louvain

Sandra Götz, Justus Liebig Universität Giessen

Hilde Hasselgård, Universitetet i Oslo

Shin'Ichiro Ishikawa, Kobe University

Magali Paquot, Université catholique de Louvain

Pascual Pérez-Paredes, Universidad de Murcia

Ute Römer, Georgia State University

Jennifer Thewissen, Universiteit Antwerpen

Stefania Spina, Università per Stranieri di Perugia



4 simple tips on how to be a good virtual conference delegate

Before the conference:

1. Block off your calendar(s) for the entire duration of the conference. **Because you will be at a conference.**

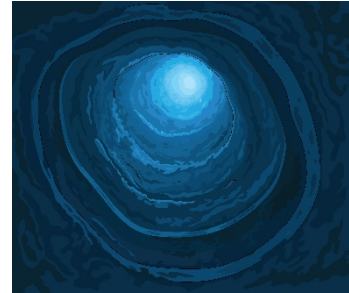


During the conference:

2. Sit somewhere where no one can find you. **Exceptions may be made for cats.**



3. Set up an email message saying you are unavailable & then turn off your email program and keep it off. **Just do not go down that rabbit hole, however tempting.**



4. Attend all the events, keynotes, papers and networking events. **Remember that you are networking with the Sylviane Grangers of tomorrow.**





Once upon a time ... A tale of learner corpus research

Sylviane Granger

Professor Emerita of English Language and Linguistics at the University of Louvain, Belgium

Sylviane is a leading researcher in the field of learner corpus studies and has published widely on learner corpus design and annotation, the analysis of phraseology in learner language with a particular focus on academic phraseology and the integration of learner corpus insights into reference and instructional materials.



Abstract:

The aim of my presentation is to take you on a journey from the very beginnings of learner corpus research (LCR) in the early 1990s through to the present day. As this journey coincides for the most part with my own scientific career, I will also take this opportunity to share some personal memories of my own journey in LCR. In the first part of the presentation I will describe how it all started, laying particular emphasis on the initial, twofold objective – to contribute to Second Language Acquisition (SLA) theory and to inform foreign language teaching (FLT) – and the two main methodologies, Contrastive Interlanguage Analysis (CIA) and Computer-aided Error Analysis (CEA). I will then turn to the current state of the field, highlighting in particular some developments that in my view are particularly positive, without glossing over those that I personally consider less welcome. I will cover the following areas: corpus collection, methodology, quantitative analysis, linguistic focus, interpretation and teaching applications. In the final part I will return to the initial objectives of LCR and argue for an equally active continuation of both SLA-oriented and FLT-oriented studies. While the former have recently been strongly promoted, the latter have lagged behind and do not appear to be given the consideration they deserve.

References: **See page 55**





Language learning and corpora in the journey through a PhD

Luciana Forti

Researcher and Lecturer in Linguistics at the University for Foreigners of Perugia, Italy

Luciana obtained her PhD in 2019 with a thesis on Data-driven learning (DDL) effects and the uses of corpora for the development of phraseological competence in Italian L2. She teaches theories and research methods in second language acquisition and conducts research on DDL for Italian and within a research project on Italian L2 phraseological complexity measures.



Abstract:

When I started my PhD, I was determined to investigate the potential of using corpora in Italian L2 pedagogy. More specifically, I wanted to conduct an empirical study to look into the effects of Data-driven learning (DDL) on the development of phraseological competence through time, work closely with learners enrolled in language courses at my university, and learn statistics. I was interested in considering variables that had received little attention in previous studies, such as semantic transparency and L1 influence, and I was particularly excited about working on a language other than English. However, I had very little clue as to how I would go about all this. So, I delved into the many methodological issues involved in researching a topic at the intersection between numerous fields such as the one I was committing to, and slowly sifted through the various stages of putting my thesis together. I was not alone in this process.

In this talk, I will show how the challenges I encountered throughout my PhD were the very things that created a number of valuable connections with many people: my supervisor, the peers and experts I met at conferences, the peers and experts I met during my research stays at UCLouvain and Lancaster University, the peers I met in various associations and groups that bring PhD candidates together, and the experts who, eventually, evaluated and discussed my work leading up to the final PhD defense. These are the people that I still feel surrounded, supported and inspired by today, as an early career researcher and lecturer. They defy the idea that a PhD is a lonely endeavour, although at times it may feel very lonely indeed. With this talk, I hope to demonstrate how rewarding doing a PhD combining language learning and corpora can be, also thanks to the people we meet along the way, whether life beyond the PhD is in academia or not.



Social/networking events

All the fun of a face-to-face conference (almost)

For anytime:

Digital Treasure Hunt (with real treasure)

Explore the conference web pages to find five letters. When you have all the letters, arrange them to find the key word.

The key word is related to the conference in some way.

Type in your key word here before the end of the conference to win the prize:

<https://nettskjema.no/a/lcrgradconf-treasure>.

A winner will be randomly selected among the participants who found the "treasure".



Clues:

- Final letter in the month the conference's YouTube channel was created.
- First letter in the name of the university hosting the conference, in its original language. This is also the first letter in the names of two of the members of our scientific committee from the University of Oslo.
- Ninth letter in the name of the research group organizing the conference.
- The fourth most frequent word in the conference's paper titles (Monday + Tuesday program).
- Seventh letter in the name of the wiki that the conference's code of conduct is based on.

Good luck!

Monday, 11 October:



'Continuing the conversation' chat with Sylviane Granger (16:00-16:30)

Just in case there's more to say after our first keynote! Join us to mingle and chat about anything LCR-related!

Icebreaker event (17:00-18:00)

Get to know your fellow conference delegates – from early career researchers to senior researchers in the field. You do not want to miss out on this opportunity to talk to like-minded people from around the world and discover connections you never knew existed.



Work-life balance

with Tove Larsson, University of Northern Arizona (18:45-19:30)

In order to thrive in academia, we have to lead a healthy and sustainable lifestyle, but how can we even begin to think about trying to attain work/life balance when our to-do lists seem endless? Managing this kind of balance requires a lot of determination and constant reminders of why it is important to do so. In this session, I'll share some thoughts on some hard-won lessons that I have learned over the years, and invite you to think about what may help you maintain a healthy balance between work and non-work activities.

Please prepare by watching Tove's 14-minute video presentation on the topic (from AAAL), linked from our YouTube channel: [HERE](#)

You can also read her blog post: [HERE](#)

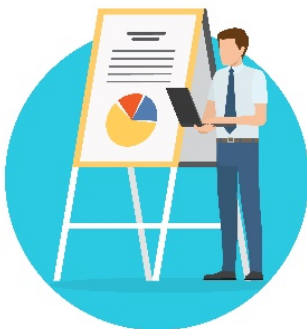


Quiz night (20:30-21:30)

Just for fun – but there will be prizes for the winning team! We promise a mix of pop culture, history, geography and current events, with questions related to language and linguistics sprinkled in here and there. You will work together in smaller teams in a breakout room, and type the answers you come up with into a digital form. Bring your brains (and close your browsers) – and spend an hour chatting about topics you never knew you would ever discuss at an academic conference!

Tuesday, 12 October

Poster event (10:00-10:45)



A live event with some of our poster presenters, where they discuss aspects of their work with the expert of their choice. Check out their posters on our [YouTube channel](#), then join us for this exciting event - a learning experience for all of us, where we can all also contribute actively!

Our conversation pairs:

- [Jennifer Jordan](#) & Jennifer Thewissen
- [Anood Al Shibi](#) & Sylviane Granger
- [Virginia Rapún Mombiola](#) & María Belén Díez Bedmar
- [Jessica Chamoun](#) & Nicolas Ballier

Social media for academic purposes with Elen Le Foll, Universität Osnabrück (11:45-12:30)

A networking haven for (early career) researchers? Productive procrastination? Or simply a waste of time? Elen Le Foll shares her experiences of tweeting as an ECR (@ElenLeFoll) before opening up the floor for a lively discussion on the benefits and risks of using social media for academic purposes.



Meet the experts (13:30-14:30)



This is your chance to talk to the experts in the field! We'll open up for breakout-room chats with senior researchers, you can talk or just be a fly on the wall in what promises to be interesting discussions about life, the universe and everything – or learner corpus research. Bring a hot/cold drink (depending on your time zone and climate) and join us! Our confirmed experts:

- María Belén Díez Bedmar
- Pascual Pérez-Paredes
- Jennifer Thewissen

'Continuing the conversation' chat featuring Luciana Forti (15:30-16:00)

Last chance to mingle with everyone before the conference draws to a close! One last chat, inspired by our second keynote talk.














Video posters

Posters plus brief video presentations

Please navigate to the LCRgradconf YouTube channel to view our posters and leave feedback:



And join us for our [Poster Event](#) on Tuesday at 10:00!











- **Anood Al Shibli** (University of Sheffield): **Page 3016**
Investigating of 19 grammatical categories in the Omani Learners of English Corpus 
- **Jessica Chamoun & Nicolas Bailier** (Université de Paris): **Page 18**
Automatic Classification of Arab learners of English based on complexity metrics 
- **Martin Johansson** (Uppsala University): **Page 19**
Adverbial Conjunctions in Senior High School Students' Written English: A Corpus Analysis of Swedish and British Students' English Conjunction Usage 
- **Jennifer Jordan** (Kwansei Gakuin University / Cardiff University): **Page 20**
Examining Lexical Phrases in a Longitudinal Learner Corpus 
- **Jen-Yu Li** (Université Rennes 2): **Page 21**
Dependency parsing for the retrieval of erroneous collocations in a learner corpus 
- **Virginia Rapún Mombiela** (University of Tartu): **Page 23**
The process of Spanish article acquisition by undergraduate students of Spanish Language and Literature 
- **Ryoko Osada** (Temple University, Japan Campus): **Page 25**
Native vs. Non-native
Uses of Transition Phrases: Evidence from the International Corpus of Japanese as a Second Language 
- **Simona Sobolevská** (Charles University): **Page 27**
Fluency markers in the speech of advanced learners of English before and after a study stay in an English-speaking country 
- **Ekaterina Uetova** (unaffiliated): **Page 28**
The Acquisition of Russian Agreement and Case Government by French-speaking Students 



Full papers

- **Seda Acikara** (Northern Arizona University): **Page 30**
The impact of communicative purpose on the discourse functions of personal pronouns in dialogic assessment tasks
- **Golpar Bahar** (University of Siegen, Germany): **Page 31**
Investigation of the Use of English Relative Clauses in Learner Corpora
- **Kanittarat Boottawong** (Université Saint-Louis, Belgium): **Page 32**
Thai Learner English: Capturing the interaction between grammatical accuracy and syntactic complexity
- **Dorra Chaabane** (Paris Nanterre University): **Page 34**
Bidirectional Language Transfer: The influence of L2 English on L1 French
- **Nicole C. De Los Reyes** (Georgia State University): **Page 35**
The use of *ni* by beginner-level Japanese as a Foreign Language learners: An exploratory study of phraseological tendencies and semantic preferences
- **Nida Dusturia** (University of Bremen, Germany): **Page 36**
A Learner Corpus Study of Indonesian EFL Learners' Connector Usage: Comparing Written and Spoken Argumentative Texts
- **Linda Gerisch** (Inland Norway University of Applied Sciences): **Page 37**
A Corpus-Based Network Approach to Second Language Acquisition and Teaching
- **Helena Hanneder** (Philipps-Universität Marburg): **Page 39**
Working Towards a Gold Standard in Writing Revision Analysis
- **Yevheniia Hasai** (Hamburg University): **Page 40**
Multilingual lexical transfer in a longitudinal learner corpus
- **Alessia Iurato** (Ca' Foscari University of Venice – Bremen University): **Page 41**
The acquisition of the Chinese 是 $shì$...的 de construction by L1 Italian learners: A preliminary analysis based on the compilation of a learner corpus and experimental data



- **Niina Kekki** (Centre for Language and Communication Studies, University of Turku): **Page 43**
The use of synonymous adjectives of Finnish as a second language learners: Applying the MuPDAR(F) approach 
- **Lin Li** (Massey University, New Zealand): **Page 44**
Challenges in Building Comparable L1 and L2 Corpora 
- **Nikita Login & Anna Viklova** (National Research University Higher School of Economics)
Predicting the difficulty level of learner-corpus-based grammar questions 
- **Olga Lopopolo** (Eurac Research – University for Foreigners Perugia): **Page 47**
Cross-linguistic influence in the acquisition of progressive aspect: a corpus-based study on multilingual learners of English as Third Language 
- **Ana Lúcia S. Pitanguy Marques** (Universidade Federal de Minas Gerais – UFMG – Belo Horizonte (Minas Gerais) - Brazil): **Page 48**
Specialized corpora for primary school young learners 
- **Sophia Minnillo** (University of California, Davis): **Page 49**
Preterit-Imperfect Development in Three Written Spanish Learner Corpora 
- **Yoon Namkung** (Georgia State University): **Page 50**
A Phraseological Analysis of Korean EFL Learners' Use of *This/These* in Academic Writing 
- **Anna Scherbakova** (NRU HSE, Moscow): **Page 52**
A Corpus-Based Case Analysis on Syntactic Complexity in Russian ESL Learners' Writing 
- **Julia Schlauch** (Justus-Liebig-Universität, Gießen, Germany): **Page 53**
A longitudinal study of teenage learners of German as a second language in intensive classes in Germany 
- **Christian Holmberg Sjöling** (Mid Sweden University): **Page 54**
A longitudinal study of Swedish upper secondary school students' vocabulary development 

Investigating of 19 grammatical categories in the Omani Learners of English Corpus

Anood Al Shibli

University of Sheffield

asal-shibli1@sheffield.ac.uk

The Common European Framework of Reference (CEFR) provides a clear description of the level of competence of language users at different levels of proficiency (North, 2014). Its framework aims to have objective criteria to determine learners' level of proficiency and sustain the recognition of different institutional qualifications despite their different teaching and learning contexts (Green, 2010). The CEFR, however, does not assign specific English language indicators to CEFR proficiency levels (Saville, 2010). Therefore, in 2005, the Cambridge ESOL group of the Cambridge Assessment collaborated with Cambridge University Press and other stakeholders to advocate the English Profile Programme (EPP). This programme aims to provide "reference level descriptions and to add grammatical and lexical details of English to CEFR's functional characterisation of the different levels by using the resources of Cambridge Learner Corpus (CLC)" (John A. Hawkins and Luna Filipović, 2012, p. 5). One part of this project was the English Grammar Profile (EGP), which describes grammatical competence at each CEFR proficiency level (O'Keefe and Mark, 2017). It investigated 19 grammatical categories (e.g., past, present, passive, future, adjectives etc.) to write can-do statements. Can-do statements indicate that learners master different grammatical structures at a selected level of proficiency¹. Refer to O'Keefe & Mark (2017) for more information on the methodology used to have the EGP.

Therefore, the current study adapts O'Keefe & Mark (2017) EGP methodology to look at the grammatical competence in written production in the Omani Learner of English Corpus (the OLEC). According to the institution, English learners in the OLEC are at level four whose level of proficiency is described as B1 CEFR level of proficiency. The purpose of this study is to describe what learners in the OLEC can-do when they use the 19 grammatical categories (e.g. learners can use 'but' to join a limited range of common adjectives after 'be'). The current study is different as it only focuses on one L1 different from the EGP, where it examined more than 150 L1s. Hence, it is unique because it can be contrasted with the EGP can-do lists to show learners' level of proficiency in the OELC.

Methodologically, with Sketch Engine features, concordance lines were generated and then filtered manually to see if the particular grammatical structure is used among as many learners as possible and if it is used correctly. Then, the decision of the selected grammatical structure is made if it met both criteria or not.

The results showed that only 13 grammatical categories met the criteria to write can-do statements. Specifically, out of 1,222 can-do statements in the EGP, the OLEC has only 68. These 68 can-do statements are generally listed under A 1 and A 2 CEFR levels of proficiency, contradicting the assigned level of proficiency of the institution (they assigned B 1) (Ministry of Manpower, 2018). Another interesting result is that seven out of the 68 can-do statements are categorised under B 1 in the EGP. These seven can-do statements were mastered because of the type of writing (argumentative essays). These findings can inform language teaching English in the institution in terms of reviewing the curriculum to improve the level of proficiency of these learners. The second finding, for instance, indicated that learners' written language improves because the writing curriculum in the institution is only based on writing argumentative essays (Ministry Of Manpower, 2018). Accordingly, it is advised to have different types of writing to have a chance to learn other grammatical structures.

References:

- Green, A. (2010) 'Requirements for Reference Level Descriptions for English', *English Profile Journal*, 1(1), pp. 1–19.
- John A. Hawkins and Luna Filipović (2012) *Criteria features in L2 English: Specifying the Reference Levels of the Common European Framework*. Cambridge: Cambridge University Press.
- Ministry of Manpower (2018) *Continuous Assessment Guidelines – Level 4*.
- North, B. (2014) 'Putting the Common European Framework of Reference to good use', *Language Teaching*, 47(2), pp. 228–249. doi: 10.1017/s0261444811000206.
- O’Keeffe, A. and Mark, G. (2017) 'The English Grammar Profile of learner competence', *International Journal of Corpus Linguistics*, 22(4), pp. 457–489.
- Saville, N. (2010) 'The English Profile Programme: Background, current issues and future prospects', *Language Teaching*, 43(02), p. 238.

Automatic Classification of Arab learners of English based on complexity metrics

Jessica Chamoun & Nicolas Ballier

Université de Paris CLILLAC-ARP F-75013

tayeh.jessica@gmail.com, nicolas.ballier@u-paris.fr

This study investigates the possibility of predicting the assigned CEFR levels of English texts written by Arab learners of English who live in the UAE. We are working with 390 essays from the ZAEBUC corpus¹. The task for students consisted in writing essays on how the UAE can promote a culture of tolerance in society. A rating of each text into one of the six CEFR bands (A1 to C2) was provided for each text by three experts. There were three different raters but the experiment was based on the majority band:

A1	A2	B1	B2	C1
7	96	195	82	10

Previous research have shown the relevance of textual metrics for level learner predictions (Lu 2014). Drawing on similar previous research based on other L1 learners of English (Ballier et al., 2019), we investigate automatic classification based on metrics such as Average misspelling_every_50words, misspelling_count, misspelling_percentage, and others. This allows us to determine the reliability of these metrics or the need to create other metrics to classify learners' essays into CEFR levels.

Using the pipelines described in (Sousa et al. 2020), we used several types of textual metrics applying to different linguistic dimensions, such as LCA (Lu, 2014) and TAALES (Kyle, 2018) for lexical complexity. L2SCA (Lu, 2010) and TAASC (Kyle et al., 2018) are used for syntactic complexity, TAACO for cohesion, and the PyEnchant for misspelt words. The Python textstat7 library was also used to compute readability metrics.

Having computed 768 metrics in total, we tried to predict the CEFR levels of learners on the basis of these metrics. Using extreme gradient boosting (XGB, Chen et al. 2015) with a 70/30 split ratio for training and testing on our data, we managed 72.17 % accuracy on our test set.

References

- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., & Cho, H. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4).
- Khushik, G. A., & Huhta, A. (2020). Investigating Syntactic Complexity in EFL Learners' Writing across Common European Framework of Reference Levels A1, A2, and B1. *Applied Linguistics*, 41(4), 506-532.
- Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior research methods*, 50(3), 1030-1046.
- Lu, X. (2014). *Computational methods for corpus annotation and analysis*. Springer.
- Sousa, A., Ballier, N., Gaillat, T., Stearns, B., Zarrouk, M., Simpkin, A., & Bouyé, M. (2020). From Linguistic Research Projects to Language Technology Platforms: A Case Study in Learner Data. In *Proceedings of the 1st International Workshop on Language Technology Platforms* (pp. 112-120)

Adverbial Conjuncts in Senior High School Students' Written English: A Corpus Analysis of Swedish and British Students' English Conjunct Usage

Martin Johansson

Uppsala University, Department of English
martin.johansson.4952@student.uu.se

This MA thesis investigates Swedish senior high school students' use of adverbial conjuncts in written English. Three corpora were examined: two learner corpora of Swedish students' written argumentation (ULEC and CISHAW) and a control corpus of British students' written argumentation (LOCNESS). AntConc was used to examine the use of conjuncts in the corpora, based on the following research question:

- How do the ULEC and the CISHAW writers use conjuncts in relation to the LOCNESS writers, with respect to semantic categories, placement, and syntactic realizations?

The results indicate that the learners' and the native speakers' distributions of conjuncts follow similar patterns: listing, contrastive, and resultive conjuncts are notably more common than summative and transitional conjuncts; initial position is more common than medial position, which is more common than end position; and single-adverb conjuncts are remarkably more common than other syntactic realizations. These findings generally support earlier research (e.g. Junmei 2015; Van Vuuren & Berns 2018; Altenberg & Tapper 1998). Some further observations are made. It is discussed whether differences in (in)formality and in frequencies of resultive and contrastive conjuncts may suggest that novice debaters tend to restrict their written argumentation to quick, informal conclusions, whereas slightly more experienced debaters tend to use formal, contrastive discussions. The findings on (in)formality mostly echo previous research (e.g. Ha 2014), but the hypothesis about quick conclusions versus contrastive discussions is not much discussed in previous research.

All in all, the study provides some support for the idea that English language teaching should address variability in conjunct usage, primarily in terms of semantic categories, and possibly also in terms of syntactic realizations and adverbial placement.

References

Primary Sources

ULEC, the Uppsala Learner English Corpus

CISHAW, the Corpus of Intermediate Senior High School Students' Argumentative Writing [unofficial, self-built]

LOCNESS, the Louvain Corpus of Native English Essays

Secondary Sources

Altenberg, B. and Tapper, M. 1998. The Use of Adverbial Connectors in Advanced Swedish Learners Written English. In: S. Granger, ed. 1998. *Learner English on Computer*. Harlow: Longman. Ch. 6.

Ha, M-J. 2014. A Corpus-Based Study on Korean EFL Learners' Use of English Logical Connectors. *International Journal of Contents*, 10(4), pp. 48-52. doi: <http://dx.doi.org/10.5392/IJoC.2014.10.4.048>

Junmei, J. 2015. A Study on the Developmental Characteristics of Adverbial Conjunctions by Chinese Non-English Majors. *Journal of English Language Teaching*, 5(3), pp. 19-27. Available at: <https://files.eric.ed.gov/fulltext/EJ1097393.pdf> [accessed May 3, 2021]

Van Vuuren, S. and Berns, J. 2018. Same difference? L1 Influence in the Use of Initial Adverbials in English Novice Writing. *International Review of Applied Linguistics in Language Teaching*, 56(4), pp. 427-461. doi: 10.1515/iral-2016-0077

Examining Lexical Phrases in a Longitudinal Learner Corpus

Jennifer Jordan

Kwansei Gakuin University / Cardiff University (PGR)

jordan.jen@gmail.com

Research in formulaic language use has increased in both in L1 and learner corpora. L1 studies indicate that formulaic language is ubiquitous in both spoken (e.g. Conrad & Biber, 2005) and written academic language (Biber & Barbieri, 2007) meaning it should also be salient as an L2 learning target. Lexical phrases, are a category of formulaic language described by Nattinger and DeCarrico as a pedagogically salient “collocations, such as how do you do? and for example that have been assigned pragmatic functions” (p. 36). They have much in common with other definitions of FL and in the context of academic writing they can be used to fulfil pragmatic and discourse functions. This presentation discusses lexical phrases in terms of two under researched areas 1) their development in a longitudinal (2 year) learner corpus 2) their use as signposting phrases that offer genre-specific scaffolding. A corpus of academic writing following Japanese L1, English L2 writing over 4 semesters was developed for this study. The corpus contains 5 submissions, the first essay and the final essay of each semester. N-grams were extracted from each corpus and then filtered manually to derive a list of lexical phrases used by the learners. The phrases were then separated into two types, discourse marking phrases used to connect ideas in the text and signposting phrases used to introduce certain steps or moves. The results of will be discussed in terms of the two types of lexical phrases identified and with relation to how they are used as learners progress through a language program. The results indicate a steady increase in the density of overall lexical phrases, but the kinds of phrases used were subject to different patterns of fluctuation.

References

- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286. doi:<http://dx.doi.org/10.1016/j.esp.2006.08.003>
- Conrad, S. M., & Biber, D. (2005). The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. *Applied Linguistics Faculty Publications and Presentations*. Retrieved from <https://pdfs.semanticscholar.org/69ea/05319bada16b3b82764eb7e69209b0794c02.pdf>
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Dependency parsing for the retrieval of erroneous collocations in a learner corpus

Jen-Yu Li

Linguistique Ingénierie et Didactique des Langues (LIDILE), Université Rennes 2
jenyu.li@etudiant.univ-rennes2.fr

Mastery of phrasemes makes a language learner become more proficient and fluent (Meunier and Granger, 2008). It is essential both for writing (Garner *et al.*, 2020; Granger and Larsson, 2021) and for oral expression (Uchihara *et al.*, 2021). Collocations, as a subset of phrasemes, are currently viewed as a necessary component of second language (L2) lexical competence (Nesselhauf, 2003). Corpus-based phraseological analysis can help understand the development of L2 acquisition (Liu and Lu, 2020), and such research requires automatic collocation extraction from learner corpora.

Recently, dependency grammar has gained more and more support in Natural Language Processing (NLP) and was reported to be able to improve the quality of collocation extraction (Uhrig and Proisl, 2012). For example, Verb-Noun (VN) pairs like *give (sth) (to) dog*, *remove (sth) (from) heat* can be excluded because the noun is not the direct object of the verb. Uhrig *et al.* (2018) systematically studied various dependency parsers and schemes for the extraction of standard collocations. Huang *et al.* (2018) demonstrated that parsers built from standard English may introduce bias while being applied to learner English. Berzak *et al.* (2016) proposed a Treebank of Learner English and measured the impact of grammatical errors on parsing. Yet, to the best of our knowledge, the impact of dependency parsing on erroneous collocation extraction, such as **create [construct] a taller building*, has not been studied.

This research presents a preliminary study about extracting erroneous VN collocations from the National University of Singapore Corpus of Learner English (NUCLE) (Dahlmeier *et al.*, 2013) where grammatical errors are annotated by professional instructors. The NLP libraries spaCy (Honnibal and Johnson, 2015) and Natural Language Toolkit (NLTK) (Bird and Loper, 2004) are used for dependency parsing and collocation extraction, respectively. The performance is evaluated with the manual annotation.

References

- Yevgeni Berzak, Jessica Kenney, Carolyn Spadine, Jing Xian Wang, Lucia Lam, Keiko Sophie Mori, Sebastian Garza and Boris Katz. 2016. Universal Dependencies for Learner English. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 737–746, Berlin, Germany, August.
- Steven Bird and Edward Loper. 2004. NLTK: The Natural Language Toolkit. In The Companion Volume to the Proceedings of 42th Annual Meeting of the Association for Computational Linguistics, pp. 214–217, Barcelona, Spain, July. Association for Computational Linguistics.
- Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu. 2013. Building a Large Annotated Corpus of Learner English: The NUS Corpus of Learner English. In Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications, pp. 22–31, Atlanta, Georgia, June. Association for Computational Linguistics.
- James Garner, Scott Crossley, and Kristopher Kyle. 2020. Beginning and intermediate L2 writer’s use of N-grams: an association measures study. *International Review of Applied Linguistics in Language Teaching*, 58(1):51-74.
- Sylviane Granger and Tove Larsson. 2021. Is core vocabulary a friend or foe of academic writing? Single-word vs multi-word uses of THING. *Journal of English for Academic Purposes*, 52, [100999]
- Yan Huang, Akira Murakami, Theodora Alexopoulou, and Anna Korhonen. 2018. Dependency parsing of learner English. *International Journal of Corpus Linguistics*, 23(1):28-54, January.

- Yingying Liu and Xiaofei Lu. 2020. N1 of N2 constructions in academic written discourse: A pattern grammar analysis. *Journal of English for Academic Purposes*, 47, [100893].
- Fanny Meunier and Sylviane Granger, editors. 2008. *Phraseology in foreign language learning and teaching*. John Benjamins Pub. Co, Amsterdam; Philadelphia.
- Matthew Honnibal and Johnson, Mark. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1373-1378, Lisbon, Portugal, September. Association for Computational Linguistics.
- Nadja Nesselhauf. 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, 24(2):223–242, June.
- Takumi Uchihara, Masaki Eguchi, Jon Clenton, Kristopher Kyle, and Kazuya Saito. 2021. To what extent is collocation knowledge associated with oral proficiency? A Corpus-based approach to word association. *Language and Speech*, accepted.
- Peter Uhrig and Thomas Proisl. 2012. Less hay, more needles – using dependency-annotated corpora to provide lexicographers with more accurate lists of collocation candidates. *Lexicographica* 28:141–179.
- Peter Uhrig, Stefan Evert, and Thomas Proisl. 2018. Collocation Candidate Extraction from Dependency-Annotated Corpora: Exploring Differences across Parsers and Dependency Annotation Schemes. In: Cantos-Gómez P., Almela-Sánchez M. (eds) *Lexical Collocation Analysis. Quantitative Methods in the Humanities and Social Sciences*. Springer, pp 111-140.

The process of Spanish article acquisition by undergraduate students of Spanish Language and Literature

Virginia Rapún Mombiola

University of Tartu

virginia.rapun@ut.ee

The ultimate aims of my research are 1) to describe students' competencies related to the use of articles and 2) to ascertain at what point Estonian and English languages influence this use. In order to attain these goals, I will compile the *Corpus Longitudinal de Español de Aprendices Estonios* (Estonian Longitudinal Learner Corpus of Spanish) at the University of Tartu. I will base my research on theories related to L3 acquisition (Williams and Hammarberg, 1998; De Angelis, 2007; Hammarberg, 2010; Cenoz, 2013).

To lay the groundwork of this project, I have carried out preliminary research during the academic year 2020-2021, during which I have compiled a pilot corpus with written and oral samples produced by 15 voluntary first-year students. In the first semester, all students who had previously studied Spanish took a test to determine their level. Each sample was annotated for a large amount of metadata with the characteristics of the student and the text. Through a background questionnaire extra information on the learners' experiences in Spanish was obtained.

The pilot corpus contains 157 written texts and 186 spoken interviews. The written tasks have been completed in the computer room to get digital versions of students' output. In the future, this process will benefit from recent research by Mari Kruse (2021): I will record the elaboration process of the texts with the keylogger Inputlog and the screen recorder Panopto to know in detail which external sources students use to produce their texts. The oral samples were collected by recording conversations held in Microsoft Teams and manually transcribed following CEDEL2 conventions (Lozano, 2020).

To be able to pin down the cross-linguistic influence of the students' utterances there will be necessary to gather a control corpus of learner natives. Furthermore, in order to carry out a parallel analysis with a corpus integrated by samples from English-speaking students, I am currently collaborating in the Project CEDEL2, led by Cristóbal Lozano from the University of Granada.

References

- Cenoz J., Hufeisen B. & Jessner U.
(2001) Towards Trilingual Education, *International Journal of Bilingual Education and Bilingualism*, 4 (1), 1-10, DOI: 10.1080/13670050108667714
- Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching*, 46 (1), 71-86.
- De Angelis, G. and L. Selinker (2001) Interlanguage transfer and competing linguistic systems, in J. Cenoz, B. Hufeisen and U. Jessner (eds), *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters, pp. 42–58.
- De Angelis, G. (2007). *Third or Additional Language Learning*. Clevedon: Multilingual Matters.
- Hammarberg, B. (2001) Roles of L1 and L2 in L3 production and acquisition, in J. Cenoz, B. Hufeisen and U. Jessner (eds), *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters, pp. 21–41.
- Hammarberg, B. (2010). The languages of the multilingual: Some conceptual and terminological issues. *International Review of Applied Linguistics in Language Teaching*, 51, 299–322. DOI: 10.1515/iral.2010.005
- Jessner, U. (2003) 'On the nature of crosslinguistic interaction in multilinguals', in J. Cenoz, B. Hufeisen and U. Jessner (eds), *The Multilingual Lexicon*. Dordrecht: Kluwer, pp. 45–55.

- Jessner, U. (2006). *Linguistic Awareness in Multilinguals*. Edinburgh: Edinburgh University Press.
- Leonet, O. & Cenoz, J. & Gorter, D. (2019). Developing morphological awareness across languages: translanguaging pedagogies in third language acquisition. *Language Awareness*, 29 (2). 1-19. DOI: 10.1080/09658416.2019.1688338.
- García Mayo, M.P. & Hawkins R. (eds.). (2009). *Second Language Acquisition articles. Empirical findings and theoretical implications*. Amsterdam: John Benjamins Publishing Company.
- Granger, S., Dagneaux, E.; Meunier, F. & Paquot M. (2009). *International Corpus of Learner English (ICLE) v2*. UC Louvain. ISBN: 978-2-87463-143-6.
- Granger, S., Gilquin G. & Meunier F. (eds.) (2005). *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.
- Lozano, C. & Mendikoetxea, A. (2013). Learner corpora and SLA: the design and collection of CEDEL2. In Díaz-Negrillo, Ballier & Thompson (eds.), *Automatic Treatment and Analysis of Learner Corpus Data*, 65-100. Amsterdam: John Benjamins.
- Lozano, C. (submitted 2020, revised 2021, under 2nd review). *CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research*.
- Lozano, C. (2021). Corpus textuales de aprendices para investigar sobre la adquisición del español LE/L2. In Cruz Piñol (Ed.), *e-Research y español LE/L2. Investigar en la era digital*, 138-163. Londres y Nueva York: Routledge.
- Pajusalu, R. (1997). Is there an article in (spoken) Estonian? In Erelt, M. (Ed.), *Estonian: Typological Studies II*, 146-177.
- Santiago Alonso, G. (2017). Los efectos de la instrucción de procesamiento con input/output enriquecido para la adquisición del artículo en español. *Colindancias: Revista de la Red de Hispanistas de Europa Central*, 8, 221-249.
- Valverde, P. & Ohtani, A. (2014). Annotating article errors in Spanish learner texts: design and evaluation of an annotation scheme. Presented at the 28th Pacific Asia Conference on Language, Information and Computation, 234-243. Phuket, Thailand: Wirote Aroonmanakun, Prachya Boonkwan, and Thepchai Supnithi.
- Williams, S., & Hammarberg, B. (1998). Language Switches in L3 Production: Implications for a Polyglot Speaking Model. *Applied Linguistics*, 19, 295-333.
DOI: <http://dx.doi.org/10.1093/applin/19.3.295>

Native vs. Non-native Uses of Transition Phrases: Evidence from the International Corpus of Japanese as a Second Language

Ryoko Osada

Temple University, Japan Campus
osada@temple.edu

Transition phrases connect sentences or clauses. This study investigated how learners whose native language is Chinese/Korean/English/Japanese use 28 transition phrases in dialogues using the International Corpus of Japanese as a Second Language (the I-JAS corpus) that is the largest Japanese learners' corpus completed in 2020. These two questions were examined; if there were any frequency differences of transition expressions usage, and if there were any transition expressions that are used more/less by comparing the 4 groups.

We found from the data of pivot table and ANOVA that native speakers used 'de' frequently in many contexts unlike learners. Learners used 'soshite (and)', 'sorekara (then)', and 'sonoato (after that)' unlike native speakers. In addition, learners of Chinese, Korean, and native speakers of Japanese used 'tada (just)' more than English native learners. Furthermore, correspondence analysis result showed that native speakers used various types of transition phrases; whereas learners used transition phrases limitedly and the variety of transition phrases genre were less than native speakers.

Native speakers used 'de' in many meanings; sequential, additive, comparative, supplementary, and contradictory. This is why the number of usages 'de' by native speakers is high. If learners can use 'de' in many meanings in the same way by native speakers, sentences would sound similar to native speakers. In addition, learners used 'soshite' as a sequential transition phrase; however, native speakers did not. It is likely that native speakers use not 'soshite' but 'de' for sequential meaning. Native speakers did not use 'soshite' in dialogues because it is more used in writing. It could be assumed that native speakers distinguish transition phrases for writing and speaking.

Learners use transition phrases differently from native speakers. Instructors could advise learners to use various genres of transition phrases and instruct to distinguish transition phrases for writing and speaking.

References

Asai, M.

(2003). Ronriteki bunshoniokeru setsuzokushinitsuite: Nihongo bogowashato jyokyunihongoga kushushano sakubunhikaku [Conjunctions in critical writings: Comparison of essays written by native speakers of Japanese and advanced Japanese learners]. *Kotobato bunka*, 4, 87-98. <http://doi.org/10.18999/isslc.4.87>

Ichikawa, T. (1978). *Kokugo kyoikunotameno bunshoron gaisetsu* [General statement for Japanese language education]. Tokyo, Japan: Kyoiku Shuppan.

Ishiguro, K.

(2008). *Bunsho wa setsuzokushi de kimaru* [Conjunctions are keys for sentences]. Kobunsha.

Kim, R.

(2017). YUN kakikotoba kopasuni mirareru nihongo gakushushano setsuzokushino siyounitsuite: Kankokugo bogowashano gyakusetsu kankeino setzusokushinichuumokushite [Usage of conjunctions by learners of Japanese: focused on contradictory conjunctions by Korean learners]. *Yokohama Kokudai Kokugo Kenkyu*, 35, 79-93.

- Kuramochi, M., & Suzuki, H.
(2007). Nihongo gakushushaniokeru setsuzokushino hsutoku-ryugakuseino setsuzokushi siyoujyokyou [Conjunction acquisition of Japanese learners- how exchange students learn conjunctions]. Kanda Gaigodaigaku Kiyou, 19, 211-234.
- Kobayashi, Y. (2019). Kotobano deta saiensu [Data science of languages]. Asakura shoten.
- Sakuma, M. (1990). Setsuzoku hyogen (1) (2) [Transition expression (1) (2)], Bunsho/danwano shikumi [Structure of sentences/dialogues], 12-33, Oufu.
- Sakoda, K., Ishikawa, S., & Lee, J. (2020), Nihongo gakushusha ko-pasu I-JAS nyumon [The gateway to Japanese learner corpus, the I-JAS corpus]. Kuroshio publisher.
- Sakoda, K. Ying, I., & Zhang, P.
(2017). Chugokujin nihongo gakushushano 'nenoshi' hyogeni miru bogono eikyo: I-JAS no rorupureiniokeru iraihyougenni motoduite [the influence of learners' native language: based on the request expression in roll-play tasks of the I-JAS corpus]. Kokuritsu kokugo kenkyujo gengosigen katsuyou wakushoppu happyo ronbunshu 2, 212-215.
- Tanaka, A. (1984). Setsuzokushono shomondai: sonoseiritstu to kino [Problems of conjunctions: the establishment and function]. Kenkyusiryō nihon bunpou dai 4 kan shushokuku/ dokuritoku hen fukushi rentaiishi setzuokushi kandoushi, 81-123, Meiji shoin.

Fluency markers in the speech of advanced learners of English before and after a study stay in an English-speaking country

Simona Sobolevská

Department of English Language and ELT Methodology, Faculty of Arts, Charles University
simona.sobol@gmail.com

The paper analyses productive fluency of advanced learners of English and the aim is to find out if study abroad causes changes in the productive fluency of advanced learners of English. The introduction defines fluency based on Housen et al. (2012) and Götz (2013) and explains the operationalization of productive fluency by performance phenomena following Lennon (1990) and Götz (2013). In this paper, productive fluency was operationalized by quantitative measurements of repeats, false starts, and self-corrections.

Fourteen interviews with seven advanced learners of English conducted before and after studying abroad in an English-speaking country for one semester were analysed. 1,464 instances of repeats, false starts and self-corrections were identified, tagged and subsequently quantitatively analysed based on their frequency, word length and word class before and after studying abroad.

The research revealed that the use of repeats and false starts has not changed after studying abroad, while the use of self-corrections dropped significantly after studying abroad. One-word repeats were found to be the most common with pronouns being the most repeated word class. One or two abandoned words were most frequent when it came to false starts; in the case of self-corrections, the speakers typically retraced one or two words. It was also found that certain speakers produced more performance phenomena before studying abroad and maintained the higher rates after studying abroad, suggesting that performance phenomena are employed as a speech management strategy by advanced learners of English.

The results of this paper may be considered in the context of the Common European Framework of Reference for Languages guidelines, which state that advanced learners produce fewer performance phenomena, such as repeats.

References

- Götz, Sandra. *Fluency in Native and Nonnative Speech*. John Benjamins Publishing Company, 2013.
- Housen, Alex, et al. 'Complexity, Accuracy and Fluency : Definitions, Measurement and Research.' *Dimensions of L2 Performance and Proficiency : Complexity, Accuracy and Fluency in SLA*, John Benjamins Publishing Company, 2012, pp. 1-20.
- Lennon, Paul. 'Investigating Fluency in EFL : A Quantitative Approach.' *Language Reading*, vol. 40, no. 3, 1990, pp. 387-417.

The Acquisition of Russian Agreement and Case Government by French-speaking Students

Ekaterina Uetova

Unaffiliated

euetova@gmail.com

Our study is focused on the error analysis of agreement and case government in Russian as a second language in the written works of French-speaking university students. The working corpus includes 61 students aged 17 to 26. This corpus is collected in the Russian Learner Corpus (RLC, <http://www.web-corpora.net/RLC/>) and the study relies on RLC's error annotation system.

Using quantitative analysis, the goal of the study was to observe how the number and quality of errors change with the improvement of proficiency and what are the "weak" points of Russian grammar for our experimental group. The analysis was conducted in two ways, both on erroneous and on correct forms.

The data attest that errors on gender are more frequent in adjectives, and errors on number and case in nouns, for example, the most difficult case was dative as in other studies (Rubinstein, 1995a, 1995b; Cherepovskaia al. 2021): these are the "weak" points of Russian grammar to take into account while teaching. Moreover, errors on the case are caused by incorrectly chosen gender or declension of the noun, like in the case of Russian-speaking monolingual children, (Gvozdev 1961, Cejtlin, 1982, 2003, 2009a, 2009b), which probably hints to similar acquisitional mechanisms of L1 and L2 acquisition. The results also reveal that the acquisition of Russian agreement and case government goes together with the increasing of the language proficiency level, as the distribution of correct forms gradually approaches what one finds in the standard Russian corpus - The Russian National Corpus (Slioussar & Samojlova 2015).

The results of our study might be of interest in teacher training, as well as in L2 acquisitional studies.

References

Rubinstein 1995a — Rubinstein, G. On Acquisition of Russian Cases by American Classroom Learners, IRAL: International Review of Applied Linguistics in Language Teaching, Feb 1995, Vol. 33 Issue 1. Retrieved May 9, 2003 from Academic Search Premier, Periodical Abstracts

Rubinstein 1995b — Rubinstein, G. On case errors made in oral speech by American learners of Russian. Slavic and East European Journal, 408-429, 1995

Gvozdev 1961 — Gvozdev A.N. Voprosy izučeniya detskoj reči [Issues in child language acquisition]. Moscow: Academy of Pedagogical Sciences of the RSFSR

Cejtlin 1982 — Cejtlin S.N. Rečevye ošibki i ih preduprezhdenie [Speech errors and their prevention]. Moscow: Prosveshchenie.

Cejtlin 2003

— Cejtlin S.N. Rečevye ošibki pri osvoenii russkogo jazyka kak rodnogo i inostrannogo: Opyt so postovitel'nogo analiza [Speech errors in the acquisition of Russian as a native and a foreign language: A comparative analysis]. In: Proceedings of the X Congress of the International Association of Teachers of Russian Language and Literature: Volume 2. Saint Petersburg: Politechnica, pp. 426–32.

Cejtlin 2009a — Cejtlin S.N. Grammatičeskie ošibki v osvoenii russkogo jazyka kak pervogo i kak vtorigo [Grammatical errors in the acquisition of Russian as a first and a second language]. Voprosy psixolingvistiki 9: 43–53.

Cejtlin 2009b — Cejtlin S.N. Očerki po slovoobrazovaniju v detskoj reči [Sketches on word formation in child language]. Moscow: Znak.

Posters



Cherepovskaia al. 2021 — Cherepovskaia N., Slioussar N., and Denissenko A. Acquisition of the nominal case system in Russian as a second language. *Second Language Research*.

Slioussar & Samojlova 2015 — Slioussar N.

and Samojlova M. Častotnosti različnyx grammatičeskix xarakteristik i okončanij u suščestvitel'nyx russkogo jazyka [Frequencies of different grammatical features and inflectional affixes in Russian nouns]. In: Proceedings of the conference 'Dialogue'. Available at: <http://www.dialog-21.ru/digests/dialog2015/materials/pdf/SlioussarNASamoilovaMV.pdf> (accessed January 2021).

The impact of communicative purpose on the discourse functions of personal pronouns in dialogic assessment tasks

Seda Acikara

Northern Arizona University

sa2585@nau.edu

Personal pronouns are found to be occurring frequently in spoken interaction (Biber, 1988). As they are related to interactivity (Friginal et al., 2017) and high collaboration (Crawford et al., 2019), researchers investigated the use of personal pronouns in academic discourse settings. However, such studies mainly focused on the instructor language (Rounds, 1987; Fortanet, 2004; Yeo & Ting, 2014; Okamura, 2009). Additionally, most of them examined the frequency counts of the pronouns and provided the discourse functions that the personal pronouns repeatedly serve.

By taking a register approach (see Biber, 1988), this study analyzes the use of first- and second-person pronouns in the learner corpus. Additionally, unlike the previous studies, the main focus of the study is the discourse functions of the personal pronouns and whether the communicative purpose of the task impacts the functions. The research question that guides the investigation is “Does the communicative purpose of the task affect the discourse functions of first and second-person pronouns in dialogic assessment tasks?”. To answer this question, a framework for the discourse functions (e.g., *to present an opinion/idea/argument vs. to keep the flow of the conversation*) was developed and applied to two assessment tasks with *persuasion* and *decision-making* purposes. The tasks were taken from the Corpus of Collaboration Oral Tasks (Crawford and McDonough, 2014) and include 27 pair interactions that were administered in an intensive English program. The preliminary results indicated that the frequency of personal pronouns and their discourse functions are impacted by the communicative purpose of the task. The variation between the tasks is explained in relation to the situational characteristics. The pedagogical considerations are provided.

References

- Biber, D. (1988). *Variation across speech and writing*. Cambridge: CUP.
- Crawford, W., & McDonough, K. (2014, March). *The language of tasks: A corpus-based approach*. Paper presented at the American Association of Applied Linguists Conference, Portland, OR.
- Crawford, W. J., McDonough, K., & Brun-Mercer, N. (2019). Identifying linguistic markers of collaboration in second language peer interaction: A Lexico-grammatical approach. *TESOL Quarterly*, 53(1), 180-207.
- Fortanet, I. (2004). The use of 'we' in university lectures: reference and function. *English for Specific Purposes*, 23(1), 45-66
- Friginal, E., Lee, J. J., Polat, B., & Roberson, A. (2017). You, I, and We: Personal Pronouns in EAP Classroom Discourse. In *Exploring spoken English learner language using corpora* (pp. 95-113). Palgrave Macmillan, Cham.
- Okamura, A. (2009). Use of personal pronouns in two types of monologic academic speech. *The Economic Journal of Takasaki City University of Economics*, 52(1), 17-26.
- Rounds, P. (1987). Multifunctional personal pronoun use in educational setting. *English for Specific Purposes*, 6(1), 13–29.
- Yeo, J. Y., & Ting, S. H. (2014). Personal pronouns for student engagement in arts and science lecture introductions. *English for Specific Purposes*, 34, 26-37.



Investigation of the Use of English Relative Clauses in Learner Corpora

Golpar Bahar

University of Siegen, Germany
golparbahar.et@gmail.com

This talk reports on a study that investigated the patterns of use of different types of relative clauses (RCs) in learner corpora, the argumentative essays written by learners of English with three different first languages: Azerbaijani, German, and Kurdish-Sorani. It is assumed that the frequency of RCs used in essays written by L2 learners of English may reflect the ease or difficulty of the RCs for language learners of English. Hence, 144 English argumentative essays written by 72 learners of English whose mother tongues were either Azerbaijani, German, or Kurdish-Sorani were collected. All the RCs found in the essays were extracted and analyzed. Furthermore, the instances of inaccurate forms of RCs were counted and the types of errors and the avoided RC types were identified.

The study aimed to answer the following questions: (i) How often and how correctly do the learners of English with the Azerbaijani, German, and Kurdish Sorani L1s use different types of RCs in their argumentative essays? (ii) Does the frequency of use of different RC types in each learner corpus conform to the hierarchies proposed by the noun phrase accessibility hierarchy (NPAH) and the perceptual difficulty hierarchy (PDH)?

Based on the results attained from each learner corpus, a difficulty hierarchy was set up. The hierarchies obtained from the analysis of each learner corpus were compared to the hierarchies proposed by the NPAH and the PDH to see to what extent the findings of the present study support or refute the above-mentioned hypotheses. Furthermore, to investigate whether the L1(s) of the learners affect their performance in the formation and use of English RCs, a close analysis of the errors in the data was conducted. The analysis of the results obtained is in progress.



Thai Learner English: Capturing the interaction between grammatical accuracy and syntactic complexity

Kanittarat Boottawong

Université Saint-Louis, Belgium

kanittarat.boottawong@student.usaintlouis.be

The current study presents a PhD project undertaken at the Université Saint-Louis and on which I would appreciate expert feedback. It focusses on two SLA constructs, namely grammatical accuracy and syntactic complexity. Rather than studying them in isolation as has been done in learner corpus research (see for instance Norris & Ortega, 2009; Thewissen, 2015), or examining their concomitant development within only one participant (Rosmawati, 2014; Spoelman & Verspoor, 2010) their interaction is investigated across proficiency levels and curriculum of 318 students.

The study aims to:

- Track the development of grammatical accuracy and syntactic complexity across proficiency levels: do one develop at the expense of the other as shown in Rosmawati (2014) and Spoelman & Verspoor (2010)?
- Observe the combined impact of several variables such as students' curriculum (the percentage of content and language integrated learning lessons or CLIL in the curriculum): do the essays of students with more than 80% of their lessons given in English (Full CLIL) show a higher rate of accuracy and complexity than those of other groups (Light CLIL and No-CLIL)?

The EFL written corpus consists of 318 students from the Faculty of Human Sciences in years 3 to 4 at UBR-University (Thailand). The participants were required to:

- Write an argumentative essay about their study choice.
- Take the writing and structures assessments of the DIALANG test to provide their proficiency levels.
- Provide meta-data (age, sex).

The essays will be manually annotated for grammatical errors according to the Louvain error tagging system (Dagneaux et al., 2008) and submitted to an automatic program, the L2 Syntactic Complexity Analyzer (Lu, 2010), to generate syntactic complexity scores for (1) length of production units, (2) subordination, (3) degree of phrasal sophistication. The correlation coefficient between the complexity and accuracy scores will be computed to verify their interaction across proficiency levels and curriculum. The data-processing and analysis phase are currently being prepared.

References

- Dagneaux, E., Denness, S., Granger, S., Meunier, F., Neff, J., & Thewissen, J. (2008). *Error Tagging Manual Version 1.3*. Centre for English Corpus Linguistics. Université catholique de Louvain, Louvain-la-Neuve. Unpublished manual.
- De Bot, K.; Lowie, W. & Verspoor, M. (2007). "A Dynamic Systems Theory approach to second language acquisition". *Bilingualism: Language and Cognition*. 10 (1): 7–21. doi:10.1017/S1366728906002732.
- Lu, Xiaofei (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4):474-496.
- Norris, J.M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics*, 30(4), 555-578.
- Rosmawati, R. (2014). Dynamic development of complexity and accuracy: A case study in second language academic writing. January 2014. *Australian Review of Applied Linguistics*. 37(2):75-100. DOI:10.1075/ara1.37.2.01ros
- Spoelman, M. & Verspoor, M. (2010). Dynamic Patterns in Development of Accuracy and



Full papers



Complexity: A Longitudinal Case Study in the Acquisition of Finnish. *Applied Linguistics*, 31(4):532-553.

Thewissen, J. (2015). *Accuracy across proficiency levels: A learner corpus approach*. *Corpora and Language in Use*. Presses Universitaires de Louvain. Louvain-la-Neuve.

<http://www.i6doc.com/fr/livre/?GCOI=28001100217910>



[Back to paper overview](#)

Bidirectional Language Transfer: The influence of L2 English on L1 French

Dorra Chaabane

Paris Nanterre University
dora.chaabene@gmail.com

The increasing number of multilinguals around the world urged psycholinguists to investigate the interaction of languages and its impact on the acquisition and development of the different linguistic systems. One of the most frequent forms of this interaction is Language Transfer. Jarvis and Pavlenko define as “the influence of a person’s knowledge of one language on that person’s knowledge or use of another language” (Jarvis and Pavlenko 2008:1). Language transfer also referred to as Crosslinguistic Influence has long been studied as a unidirectional phenomenon occurring from the native to a foreign language. However, in the last two decades, linguists have started investigating the bidirectionality of transfer and exploring different ways learners transfer from L1 to L2 and L2 to L1 since it can provide a more comprehensive view of learners’ interlanguage. This study takes a learner-centered approach to investigate backward transfer: the effect of L2 English on L1 French in oral production. It hypothesizes a relation between learners’ transfer errors and their language development. The aim of this study is to establish a link between backward transfer, learners’ cognitive strategies and their inter language development. Data was gathered from 50 EFL learners and 10 English native speakers aged between 18 and 30 years old. These participants filled a bio-data questionnaire for more precisions about their profiles. The sampling procedure used is the deliberate sampling procedure which involves a purposive selection of the participants. The proficiency level of the cohorts of learners is B2, C1 and C2. They are asked to watch a three-minutes silent video and describe it; first in English then in French. The description is recorded, transcribed, and then analyzed. Preliminary analysis of the corpus has shown that there are not enough instances of backward transfer. (Example: “il est confident” => lexical backward transfer/ “diable” => morphological backward transfer).

References:

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. London: Routledge.



The use of *ni* by beginner-level Japanese as a Foreign Language learners: An exploratory study of phraseological tendencies and semantic preferences

Nicole C. De Los Reyes
 Georgia State University
 ndelosreyes1@student.gsu.edu

The Japanese suffix/particle *ni* is associated with an array of semantic sense types at various levels of abstraction (Kabata, 2000), such as a marker of stative location, dative case, and reason, respectively. Only one known study to date (Kabata, 2016), however, has used learner corpus data to investigate the breadth of Japanese as a Foreign Language (JFL) learners' conceptual understanding of the semantic network of *ni* in usage events and has related these findings to the particle's phraseological tendencies.

This study further explores notions of learner corpora as valuable tools to inform second language acquisition research and pedagogy (see, e.g., Granger, 2009; Ozeki, 2016). Adopting a phraseological framework, it analyzes transcripts of semi-structured oral interview data (78,448 words) from 38 American-English speaking JFL learners extracted from the International Corpus of Japanese as a Second Language (Sakoda et al., 2016) to examine the extent to which the semantic senses of *ni* (Kabata and Rice, 1997) are represented in the spontaneous speech of beginner-level JFL learners and the collocations of *ni* within each represented sense type. A content analysis of two introductory Japanese textbooks is included to explain the learner corpus findings and to illustrate monolithic depictions of *ni* in pedagogical materials. The results reveal that learners' uses of *ni* are skewed toward concrete spatial and temporal sense types and its collocations display limited lexical variation in a NP-*ni*-VP construction, a pattern partially explained by the particle's narrow representation in Japanese textbooks. These findings suggest a need for JFL instructors to raise learners' consciousness of *ni*'s sense types through exemplars in input and schematic aids from the outset of instruction, which may extend to other polysemous particles. In so doing, they may build learners' conceptual understanding of particles' semantic network and extend dominant realizations in use from concrete to more abstract domains.

References

- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 13-32). John Benjamins. <https://doi.org/10.1075/scl.33.04gra>
- Kabata, K. (2016). A usage-based account of learner acquisition of Japanese particles *ni* and *de*. In K. Kabata, & K. Toratani (Eds.), *Cognitive-functional approaches to the study of Japanese as a second language* (pp. 89-112). De Gruyter Mouton. <https://doi.org/10.1515/9781614515029-007>
- Kabata, K. (2000). Speakers' judgements about the lexical complexity of *ni*. In A.K. Melby & A.R. Lommel (Eds.), *The Twenty-Sixth LACUS Forum* (pp. 467-478). LACUS.
- Kabata, K., & Rice, S. (1997). Japanese *ni*: The particulars of a somewhat contradictory particle. In M.H. Verspoor, K.D. Lee & E. Sweetser (Eds.), *Lexical and syntactical constructions and the construction of meaning* (pp. 107-127). John Benjamins. <https://doi.org/10.1075/cilt.150>
- Ozeki, H. (2016). Corpus-based second language acquisition research. In K. Kabata & K. Toratani (Eds.), *Handbook of Japanese Applied Linguistics* (pp. 313-334). De Gruyter. <https://doi.org/10.1515/9781614515029>
- Sakoda, K., Konisi, M., Sasaki, A., Suga, W., & Hosoi, Y. (2016). Tagengo bogo no nihongo gakusyuusya oodan koopasu [International Corpus of Japanese as a Second Language]. *Kokugokyyu Purojekuto Rebyuu [NINJAL Project Review]*, 6(3), 93-110. <http://doi.org/10.15084/00000831>



A Learner Corpus Study of Indonesian EFL Learners' Connector Usage: Comparing Written and Spoken Argumentative Texts

Nida Dusturia

University of Bremen, Germany
dusturia@uni-bremen.de

This paper compares the use of connectors in argumentative writing and speech by Indonesian EFL learners at different proficiency levels (A.2. and B.1.2 of the Common European Framework of Reference for Languages (Council of Europe, 2001) and English native speakers. The study adopts Biber et al.'s (1999) semantic categorization of connectors and Granger's (2015) Contrastive Interlanguage Analysis framework. The data come from the *International Corpus Network of Asian Learners of English* (Ishikawa, 2014) which provides spoken and written data produced in response to the same argumentative tasks and topics.

The following research question is addressed:

Do Indonesian EFL learners differ from English native speakers in the use of connectors in their argumentative essays and spoken monologues in terms of:

- a. the frequency of use and representation of semantic types (with a view to over-/underrepresentation)
- b. the contextual misuse of connectors
- c. the positioning of connectors with a sentence

Various studies have found that learners may under or overuse certain lexical items in their writing in comparison with native speakers (e.g. Milton & Tsang 1993; Granger & Tyson 1996; Aijmer & Strensöm 2004). The present study confirms such findings in that the Indonesian EFL learners tend to use more connectors in their argumentative texts than the native speakers. As for the positioning of connectors, the learners prefer to use connectors in clause-initial position, while it is more varied for the native speakers. The study thus provides further evidence for the assumption that there is a general tendency for learners to place connectors in initial position irrespective of their L1 (Van Vuuren & Berns, 2018). Drawing on this first result, this study sets out to examine whether similar observations can be made for the spoken monologues.

References

- Aijmer, K. & Strensöm, A.B. (eds.). (2004). *Discourse Patterns in Spoken and Written Corpora*. Amsterdam: Benjamins.
- Biber, D. et al. (1999). *Longman Grammar of Spoken and Written English*. Essex: Pearson Education.
- Council of Europe (2001). *The Common European Framework of Reference for Languages: Learning Teaching, Assessment*. Cambridge: Cambridge University Press.
- Granger, S. and Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English, *World Englishes*, 15(1), 17-27.
- Granger S. (2015). Contrastive Interlanguage Analysis: A reappraisal, *International Journal of Learner Corpus Research*, 1(1), 7-24.
- Ishikawa, S. (2014). The International Corpus Network of Asian Learners of English. <http://language.sakura.ne.jp/icnale/index.html>
- Milton, John and Tsang, Elza S. C. (1993). A Corpus-based study of logical connectors in EFL students' writing: directions for future research. In *Studies in Lexis*. Edited by R. Pemberton and E.S.C. Tsang. Hong Kong: The Hong Kong University of Science and Technology. pp. 2 15-246.
- Van Vuuren, S. & Berns, Janine. 2018. Same difference? L1 influence in the use of initial adverbials in English novice writing. *International Review of Applied Linguistics in Language Teaching*, 56(4): 427–461.



A Corpus-Based Network Approach to Second Language Acquisition and Teaching

Linda Gerisch

Inland Norway University of Applied Sciences

linda.gerisch@inn.no

In this project, I am investigating second language acquisition in the light of usage-based construction grammar (cf. Ellis and Cadierno 2009, Goldberg 2003, Goldberg et al. 2004). My main goal is to bring linguistic theory and classroom approaches closer together (cf. Llopis-García 2010) and to use insights from cognitive linguistics and learner corpora for both the explicit and the implicit teaching of grammatical structures. I am building on previous research in first and second language acquisition, as well as Höder's proposed "Diasystematic Construction Grammar" (cf. Höder 2018, Höder et al. Forthcoming). My overarching research question is in what way a network approach (cf. Diessel 2019, 2020) can be beneficial for the understanding of second language acquisition and, in turn, useful for L2 teaching. I intend on supplementing a theoretical analysis with results from a small learner corpus from German L2 learners, as well as available learner (at the moment EFCAMDAT) and native speaker corpora. Theoretically, the goal is to combine the notions of transfer and chunk learning, which have been mostly investigated in isolation (cf. Abbot-Smith and Behrens 2006: 996). In addition to combining the two in theory, I also want to analyse learner language taking both processes into account. Learner corpora are a useful tool for gaining knowledge about the acquisition of additional languages, but I believe that their potential is not being realised at the moment. Moreover, the knowledge obtained from them can be used to improve language teaching. So, another important goal of this project is to find out how the creation and analysis of a learner corpus can be beneficial for teaching English constructions and their fillers (esp. verbs) in the classroom, with a focus on the influence of transfer for learners with specific L1s.

References

- Abbot-Smith, Kirsten and Heike Behrens (2006). How Known Constructions Influence the Acquisition of Other Constructions: The German Passive and Future Constructions. *Cognitive Science*, 30(6), 995–1026. https://doi.org/10.1207/s15516709cog0000_61.
- Diessel, Holger (2020). A Dynamic Network Approach to the Study of Syntax. *Frontiers in Psychology*, 11, 1-15. <https://doi.org/10.3389/fpsyg.2020.604853>
- Diessel, Holger (2019). *The Grammar Network: How Linguistic Structure is Shaped by Language Use*. Cambridge University Press.
- Diessel, Holger (2009). On the Role of Frequency and Similarity in the Acquisition of Subject and Non-Subject Relative Clauses. In Talmy Givón and Masayoshi Shibatani (Eds), *Syntactic Complexity* (p. 251-276). John Benjamins.
- Ellis, Nick C. and Teresa Cadierno (2009). Constructing a Second Language. *Annual Review of Cognitive Linguistics*, 7(1), 111-139. <https://doi.org/10.1075/arcl.7.05ell>
- Goldberg, Adele E. (2003). Constructions: A New Theoretical Approach to Language. *Trends in Cognitive Sciences*, 7(5), 219–24.
- Goldberg, Adele E., Devin M. Cashenhiser, and Nitya Sethuraman (2004). Learning Argument Structure Generalizations. *Cognitive Linguistics*, (15)3, 289–316.
- Höder, Steffen, Julia Prentice, and Sofia Tingsell (Forthcoming). Acquisition of Additional Languages as Reorganization in the Multilingual Constructicon. In Hans C. Boas and Steffen Höder (Eds), *Constructions in Contact 2. Language Change, Multilingual practices, and Additional Language Acquisition*. John Benjamins.
- Höder, Steffen (2018). Grammar is Community Specific: Background and Basic Concepts of Diasystematic Construction Grammar. In Hans C. Boas and Steffen Höder (Eds.), *Constructions in Contact: Constructional Perspectives on Contact Phenomena in Germanic Languages* (p. 37-70). John Benjamins.



Full papers

Llopis-García, Reyes (2010). Why Cognitive Grammar Works in the L2 Classroom. *AILA Review*, 23(2010), 72-94. <https://doi.org/10.1075/aila.23.05llo>



[Back to paper overview](#)

Working Towards a Gold Standard in Writing Revision Analysis

Helena Hanneder

Philipps-Universität Marburg

helena.hanneder@staff.uni-marburg.de

So far, learner corpus research on writing revisions has focused on advanced learners and very specific topics (cf. Chanquoy 2001, Lindgren 2005, Razak & Saeed 2015), resulting in the absence of a widely applicable taxonomy and generalisable findings. The present paper will discuss methodological difficulties in writing revision analysis, specifically asking: how can we consistently distinguish between conceptual and formal revisions of non-native speakers? The data is taken from the *Marburg Corpus of Intermediate Learner English*, which consists of written high school exams by German learners of English. The longitudinal data includes the school years nine to twelve of the same 91 secondary school students and cumulates in 15,302 revisions. This data allows for a process-based approach that adds to the research findings of the usually product-based investigations in corpus research. However, delineating the specific change present in a revision can be difficult, which we want to explore. This paper will present a possible taxonomy and rules to guide writing revision analysis generally as well as annotation of the MILE specifically. The observations that will be discussed range from uncomplicated ones to more problematic cases, such as:

(1a) Link left ~~his family~~ Bradford (0004-1-09-00109)

(1b) [...] so it is easier for Iago to ~~get~~ make Othello jealous than Desdemo. (0061-2-11-00052)

The addition in (1a) is a straightforward observation concerning content. (1b) could be a content revision or a change for the sake of a collocation, for example: uncertainty between *get angry* and *make angry*, uncertainty between *get jealous* and *make jealous* or a conceptual change. This paper will discuss similar pitfalls and problems before proposing a possible solution, thereby starting a conversation that hopefully brings us one step closer to a gold standard in writing revision analysis.

References

- Chanquoy, Lucile. (2001). How to make it easier for children to revise their writing: A study of text revision from 3rd to 5th grades. *The British Journal of Educational Psychology*, 71, 15-41.
- Lindgren, E. (2005). *Writing and Revising: Didactical and Methodological Implications of Keystroke Logging*. Umeå Universitet.
- Razak, N. A., & Saeed, M. A. (2015). EFL Arab Learners' Peer Revision of Writing in a Facebook Group: Contributions to Written Texts and Sense of Online Community. *English Language Teaching*, 8(12), 11– 26.



Multilingual lexical transfer in a longitudinal learner corpus

Yevheniia Hasai

Hamburg University

yevheniia.hasai@uni-hamburg.de

Cross-linguistic influence (CLI) at the lexical level or lexical transfer (LT) is a widely studied phenomenon (Bardel, 2015; Jarvis, 2009; Ringbom, 2001; Ringbom, 2006). It allows to draw conclusions on the role of the first language (L1) or any other language (Ln) in the vocabulary learning process. Few studies have explored the presence of multilingual LT in longitudinal corpora and the influence of additional factors (e.g., teaching style, age, socio-economic status) on it.

This research investigates the use of LT (3,447 tokens) in short English texts written by monolingual (German) and bilingual (Russian/Turkish-German) pupils (initially attending grades 7 and 9) from a longitudinal perspective. The data were taken from the research project *Mehrsprachigkeitsentwicklung im Zeitverlauf (Multilingual development: a longitudinal perspective) (MEZ)* that was carried out at the University of Hamburg from 2014 to 2019 (Brandt et al. 2017). In total 449 texts that were produced by 120 students during four measuring points (over 2.5 years) were collected and analyzed.

This study assesses i) whether the bilingual learners transfer from the majority language (German) and the heritage language (Russian/Turkish) or just from one, ii) whether there are differences in the use of lexical transfer between the groups, and iii) whether the variables such as type of school (higher vs. lower academic track), gender, age, socio-economic status, and motivation towards learning English influence the use of lexical transfer.

The results demonstrate that the bilingual learners use LT (predominantly) from the majority language (German). This can be explained by typological similarity between English and German, the dominant language transfer, and the use of German in the English language classroom. Grade, measuring point, and school type impact lexical transfer in a statistically significant way.

References

- Bardel, Camilla. 2015. Lexical cross-linguistic influence in third language development. In Hagen Peukert (ed.), *Transfer Effects in Multilingual Language Development*, 111–128. Amsterdam: John Benjamins.
- Brandt, Hanne, Marina Lagemann & Sharareh Rahbari. 2017. Multilingual development: A longitudinal perspective – Mehrsprachigkeitsentwicklung im Zeitverlauf (MEZ). *European Journal of Applied Linguistics* 5(2). 347–357.
- Jarvis, Scott. 2009. Lexical transfer. In Aneta Pavlenko (ed.), *Bilingual mental lexicon: Interdisciplinary approaches*, 99–124. Clevedon: Multilingual Matters.
- Ringbom, Håkan. 2001. Lexical transfer in L3 production. In Jasone Cenoz, Britta Hufeisen & Ulrike Jessner (eds.), *Cross-linguistic influence in third language acquisition: psycholinguistic perspectives*, 59–68. Clevedon: Multilingual Matters.
- Ringbom, Håkan. 2006. The importance of different types of similarity in transfer studies. In J. Arabski (ed.), *Cross-linguistic influences in the second language lexicon*, 36–45. Clevedon: Multilingual Matters.



The acquisition of the Chinese 是shì...的de construction by L1 Italian learners: A preliminary analysis based on the compilation of a learner corpus and experimental data

Alessia Iurato

Ca' Foscari University of Venice – Bremen University
alessia.iurato@unive.it

The Chinese 是shì...的de construction, often treated as a cleft construction, poses a challenge for L2 learners because it differs from *it*-clefts in languages like English (Xu 2014). This construction has been widely analysed but has never been studied within the framework of Learner Corpus Research (LCR) in which L2 Chinese is generally understudied. I will focus on the acquisition of the *shì...de* pattern proper (Paul & Whitman 2008) which consists of a positionally determined contrastive focus and a presupposition, and has a past-tense reading only (Simpson & Wu 2002), as e.g.:

(1) 他是昨天来的

tā shì zuótiān lái de

3SG COP yesterday come DE

'It was yesterday that he came' (Jing-Schmidt 2017: 213).

I adopt a multi-method triangulated approach, grounded in LCR (Granger 2008), consisting in the analysis of both corpus data and experimental data to provide different insights into the phenomenon under study (Callies 2013, Gilquin 2021). I collected written and spoken data of 103 L1 Italian university students through open-ended tasks to compile an (error) annotated learner corpus, supplemented by experimental data elicited through experimental tasks to counterbalance potential avoidance phenomena and construct underrepresentation (Tracy-Ventura & Myles 2015). Moreover, I collected comparable spoken and written data of 30 L1 Chinese speakers.

The study will address the following research questions:

- Do learners across different proficiency levels have (explicit) knowledge of the pragmatic meaning of the *shì...de* construction?
- Are there any differences in the use of the *shì... de* construction by L1 Chinese speakers and L2 Italian learners to highlight information and produce contrastive focus?

Preliminary results indicate that learners are not aware of the pragmatic function of the *shì...de* construction. Learner errors primarily involve the omission of the construction due to the overgeneralization of the perfective aspect marker 了 *le*.

References

- Callies, Marcus (2013). 'Triangulation'. S.J. Schierholz & H.E. Wiegand (eds.), *Wörterbücher zur Sprach- und Kommunikationswissenschaft [WSK] online*. Berlin: De Gruyter Mouton.
- Callies, Marcus (2015). 'Learner Corpus Methodology'. In Granger, Gilquin & Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 35-55.
- Cheng, Lisa Lai-Shen (2008). 'Deconstructing the *shì...de* construction'. *The Linguistic Review*, 25: 235–266.
- Diez-Bedmar, María Belén (2015). 'Dealing with Errors in Learner Corpora to Describe, Teach and Assess EFL Writing: Focus on Article Use'. In Castello, Ackerley & Coccetta (eds.), *Studies in Learner Corpus Linguistics. Research and Applications for Foreign Language Teaching and Assessment*. Bern: Peter Lang AG, 37-69.



- Granger, S. (2008). 'Learner Corpora'. A. Lüdeling & M. Kytö, *Corpus Linguistics. An International Handbook*. Berlin/New York: Walter de Gruyter, 259-275.
- Gilquin, Gaëtanelle (2021). 'Combining learner corpora and experimental methods'. N.Tracy-Ventura & M. Paquot (eds.), *The Routledge Handbook of Second Language Acquisition and Corpora*. New York: Routledge, 133-144.
- Jing-Schmidt, Z. (2017). 'Grammatical construction and Chinese Discourse'. Shei, C. (ed.), *Routledge Handbook of Chinese Discourse Analysis*. London: Routledge, 209-230.
- Lüdeling & Hirschmann (2015). 'Error annotation systems'. In Granger, Gilquin & Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 135-157.
- Myles, Florence (2015). 'Second language acquisition theory and learner corpus research'. Granger, Gilquin & Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 309-331.
- Paul, Waltraud & Whitman, John (2008). 'Shi...de focus clefts in Mandarin Chinese'. *The Linguistic Review* 25 (3/4): 413-451.
- Shyu, Shu-ing (2015). 'Shì 是.....de 的 sentences'. Sybesma Rint (edited by) *Encyclopedia of Chinese language and linguistics*. Consulted online on 03 April 2020 <http://dx.doi.org/10.1163/2210-7363_ecll_COM_00000092> . First published online: 2015.
- Simpson, Andrew & Wu, Zoe (2002). 'From D to T- Determiner incorporation and the creation of tense'. *Journal of East Asian Linguistics* 11: 169-209.
- Tracy-Ventura, Nicole & Myles, Florence (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. Amsterdam: John Benjamins.
- Xu, Yi (2014). 'A corpus-based functional study of shi...de constructions'. *Chinese Language and Discourse* 5 (2): 146-184.



The use of synonymous adjectives of Finnish as a second language learners: Applying the MuPDAR(F) approach

Niina Kekki

Centre for Language and Communication Studies, University of Turku
niina.kekki@utu.fi

In this study, we apply Gries and Deshors' (2014, 2016) MuPDAR(F) approach to explore the use of synonymous adjectives *tärkeä/keskeinen* 'important, central' in written, academic native and advanced learner Finnish. The data of this study were extracted from two corpora: The Corpus of Advanced Learner Finnish (LAS2), and its reference corpus The Corpus of Academic Finnish (LAS1). We seek to answer the following research questions: i) What kind of similarities and differences are there between NS and NNS Finnish in the use and linguistic aspects of constructions where near-synonymous adjectives occur, and ii) What does the MuPDAR(F) approach reveal concerning the use of near-synonymous adjectives in academic NS and NNS of Finnish?

The MuPDAR(F) method applied in the study was able to yield results to confidently model the differences between the use of the near-synonyms in the native data and to distinguish between the native-like and non-native-like uses in the learner data. Crucially, the model distinguished between the contexts in which one of the synonyms is clearly favoured, and those in which either one could be considered acceptable, in accordance with Gries and Deshors' (2020) recent suggestion. The results suggest that Finnish learners follow fairly coherently the tendencies of native speakers, but that several variables differentiate learners' use of the synonyms from native speakers. We interpret the differences to reflect complexity- and prototypicality-related phenomena: on the one hand learners use more often the more common option. On the other hand, non-nativelike adjective uses only take place in contexts where the context of use is structurally in its most prototypical and least complex form, suggesting that learners employ complexity-related structural alternations, e.g. non-prototypical grammatical subjects or use of degree modifiers, after lexical alternations.

References

- Deshors, S. C., & Gries, T. Th. (2016). Profiling verb complementation constructions across New Englishes. A two-step random forests analysis of *ing* vs. *to* complements. *International Journal of Corpus Linguistics* 21:2, 192–218.
- Gries, S. Th., & Deshors, S. (2014). Using regressions to explore deviation between corpus data and a standard/target: Two suggestions. *Corpora* 9(1), 109–136.
- Gries, S. Th., & Deshors, S. (2020). There's more to alternations than the main diagonal of a 2x2 confusion matrix: Improvements of MuPDAR and other classificatory alternation studies. *ICAME Journal* 44(1), 69–96.
- LAS1 = Akateemisen suomen korpus. School of Languages and Translation Studies, University of Turku. Available at: <https://digilang.utu.fi/>
- LAS2 = Edistyneiden suomenoppijoiden korpus. School of Languages and Translation Studies, University of Turku. Available at: <https://digilang.utu.fi/>



Challenges in Building Comparable L1 and L2 Corpora

Lin Li

Massey University, New Zealand

blculyn@gmail.com

This paper illuminates some challenges encountered in the development of comparable L1 and L2 corpora. Spoken L1 corpora have vital roles in L1–L2 contrastive studies on L2 production. In such contrastive research design, corpus creation considerations have mainly focused on L2 corpora, while considerations on their L1 counterparts have usually been described in passing. It however is challenging to construct an L1 corpus which is expected to match an L2 corpus as much as possible. Firstly, representativeness and comparability are ultimately incompatible ways of looking at corpus design (Leech, 2007). So, the comparability of L1 and L2 corpora may be achieved at the expense of the representativeness of each corpus. Aside from this dilemma, another challenge is the possibility of achieving comparability. It is understandable that the L1 corpus are expected to differ from the L2 corpus in terms of only one variable (i.e., being native/non-native speakers), but be similar in other respects. However, corpus practices have proved that even with identical design criteria, L1 and L2 corpora which are claimed to be comparable, can differ considerably in some respects other than the native/non-native distinction (e.g., the degree of interactivity between interlocutors).

Based on a new corpus resource, the Spoken Chinese Corpus, which is made up of an L2 corpus and a comparable L1 corpus, this paper reflects on the above challenges by reviewing the methodological decisions taken during the corpus compilation. In doing so, it seeks to (i) suggest what lessons can be learned about how to design and build comparable L1 and L2 corpora, and (ii) encourage corpus users to pay attention to the differences between the two corpora when they use them to conduct L1–L2 contrastive studies.

References

- Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus Linguistics and the Web* (pp. 133-149). Rodopi.



Predicting the difficulty level of learner-corpus-based grammar questions

Nikita Login

National Research University Higher School of Economics
nvlogin@edu.hse.ru

Anna Viklova

National Research University Higher School of Economics, Learner Corpora Lab
aviklova@hse.ru

Determining the level of difficulty for testing items is an important problem in computer-assisted language testing - it is crucial both for developing scoring rules and for creating adaptive tests. Difficulty scoring may become an issue especially while working with a big pool of questions generated from text corpora, and in this case difficulty can be estimated automatically - using language models and features obtained from natural language processing tools, as seen in (Ha & Yaneva, 2019).

In our work we tried to create a supervised scorer based on a selection of questions sourced from REALEC (Kuzmenko & Kutuzov, 2014). The questions were correct-an-error exercises extracted from the error annotation of the learner corpus with the help of the Testmaker tool (Vinogradova, 2019). We developed a set of 21 rules for determining difficulty level and manually annotated two sets of generated questions (with and without the error tag information), while also implementing these rules in the form of a computer program. However, due to the incompleteness of corpus data and restrictions of the underlying syntactic parser model (Honnibal et al., 2021), the rule-based algorithm achieved only 67% accuracy of manual annotation.

In order to boost the performance of our difficulty scorer, we added machine learning for cases where rule-based solution failed to determine difficulty level. We tried different approaches of feature extraction (Lyashevskaya et al., 2021, Wolf et al, 2020), and classification (Pedregosa et al., 2011) on a tagless subset and compared their performance in terms of the F1-score. The best overall F1-score was achieved by a Bernoulli Naive Bayesian Classifier trained on averaged GPT-2 embeddings (Radford et al., 2019), while a DecisionTreeClassifier trained on averaged BERT embeddings (Devlin et al., 2019) showed the highest performance in identifying the most difficult questions. With the BERT-based solution added to the rule-based difficulty scoring pipeline, the accuracy improved by 10%, reaching 77%. The developed difficulty classifier is designed to be integrated in the grammar quiz editor of our online language assessment testing system LangExBank (Login, 2020).

References

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019) Bert: Pre-training of deep bidirectional transformers for language understanding. *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)*, 4171-4186. <https://doi.org/10.18653/v1/N19-1423>
- Ha, L., Yaneva, V., Baldwin, P., & Mee, J. (2020). Predicting the Difficulty of Multiple Choice Questions in a High-stakes Medical Exam. *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, 11-20. <https://doi.org/10.18653/v1/W19-4402>
- Honnibal, M., Montani, I, Van Landeghen, S., & Boyd, A. (2021). spaCy: Industrial-strength Natural Language Processing in Python. <https://doi.org/10.5281/zenodo.1212303>
- Kuzmenko, E., & Kutuzov, A. (2014). Russian error-annotated learner English corpus: A tool for computer-assisted language learning. *Proceedings of the third workshop on NLP for computer-assisted language learning*, 22, 87-97. <https://www.aclweb.org/anthology/W14-3507>
- Login, N. (2020). LangExBank. Github repository <https://github.com/lcl-hse/LangExBank>



Full papers



- Lyashevskaya, O., Panteleeva, I., & Vinogradova, O. (2021). Automated assessment of learner text complexity. *Assessing Writing*, 49. <https://doi.org/10.1016/j.asw.2021.100529>
- Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://arxiv.org/abs/1201.0490v4>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. <https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe>
- Vinogradova, O. (2019). To automated generation of test questions on the basis of error annotations in EFL essays: a time-saving tool? *Learner Corpora and Language Teaching*, 92, 29–48. <https://doi.org/10.1075/scl.92.03vin>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q., & Rush, A.M. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>

Appendix

The source code for this paper is available at <https://github.com/nicklogin/langexbank-difficulty-scoring>



[Back to paper overview](#)

Cross-linguistic influence in the acquisition of progressive aspect: a corpus-based study on multilingual learners of English as Third Language

Olga Lopopolo

Eurac Research – University for Foreigners Perugia
olga.lopopolo@eurac.edu

There are almost no corpus-based studies of cross-linguistic influence (CLI) in Third language acquisition. An exception is Martins and Pinharandas Nunes' (2013) corpus-based study of Chinese learners of L3 Portuguese, which showed preliminary evidence for instances of lexical, morphological, and semantic transfer from the learner's English L2 to their Portuguese L3 [1].

As part of my PhD project, it is my aim to investigate CLI occurring in the acquisition of Tense/Aspect (TA), specifically the use of the progressive aspect of English as L3. The investigation is guided by the following research questions:

- How do multilingual learners convey the progressive aspect in English as L3?
- Do the background languages involved (German and Italian) affect the subsequent use of progressive aspect in English?
- Which lexical verbs and semantic domains are associated with the progressive construction?

In my study, I focus on German and Italian as learners' background languages in order to trace their possible activation and influence as they differ typologically with respect to TA [2]. I will test the prevalent theories concerning CLI in Third Language Acquisition (*L2 Status Factor Model* [3], *Cumulative Enhancement Model* [4], ...) in combination with the theories concerning the acquisition of the TA system (the most widely discussed *Aspect Hypothesis* [5], *Default Past Tense Hypothesis* [6], ...).

The analysis is conducted on a sub-corpus of English texts taken from LEONIDE [7], a trilingual data collection coming from lower secondary school learners in the multilingual region of South Tyrol. A multi-layer annotation scheme using EXMARaLDA Partitur-Editor has been created which allows more levels of analysis structured in a tier tagging system. The presentation aims to discuss my multi-layer annotation scheme for a comprehensive interpretation of CLI in TA at different linguistic levels (semantics, morphology) as well as methodological issues about the interpretation of learner errors.

References

- [1] Martins, C., & Nunes, Pinharanda (2013). L2 to L3 transfer: Learner corpora analyses. *Learner Corpus Studies in Asia and the World*, 1, 271-281.
- [2] Fuchs, R., & Werner, V. (2018) Tense and aspect in Second Language Acquisition and learner corpus research: Introduction to the special issue. *International Journal of Learner Corpus Research*, 4(2), 143-163.
- [3] Bardel, C., & Falk, Y. (2007) The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, 23, 459–484.
- [4] Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *The International Journal of Multilingualism*, 1(1), 3-16.
- [5] Andersen, R. W. & Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition*, 16(2), 133-156.
- [6] Salaberry, M.R. (2008). *Marking past tense in Second Language Acquisition. A theoretical model.* London: Continuum.
- [7] Glaznieks, A., Frey, J.-C., Stopfner, M., Zanasi, L. & Nicolas, L. (submitted): LEONIDE: A longitudinal trilingual corpus of young learners of Italian, German and English.



Specialized corpora for primary school young learners

Ana Lúcia S. Pitanguy Marques

Universidade Federal de Minas Gerais – UFMG – Belo Horizonte (Minas Gerais) - Brazil
alspm2019@ufmg.br

My aim is to share the initial stages of a study which investigates possible results of building and making a level-appropriate specialized pedagogic corpora in English (L2) in Geography and Science topics to be made available to teachers to work with young learners in cross-curricular work. My ultimate aim is that this 'condensed exposure' (GABRIELATOS, 2005, p. 10) to projects in L2 can contribute to vocabulary expansion and heightened awareness of language patterns (GRANGER, 1998). For that matter, the corpora will need to be hosted in a dedicated platform to be accessible to teachers and learners:

- Can corpus-informed L2 most frequent vocabulary and concordances lines¹ be of help to teachers to create engaging and meaningful activities?
- Will learners resort to the vocabulary they had been exposed to when writing short texts in L2?

The material is being compiled from printed workbooks published to support young learners in their home work, from age-appropriate materials available on websites such as National Geographic and also from educational videos in Youtube. Nine topics have been selected from both Brazilian and foreign curricula to be assembled as pedagogic corpora. Aston (1997, p. 13) states "that work with small specialised corpora can be not only a valuable activity in its own right, [but] a means of discovering the characteristics of a particular area of language use."

The vocabulary will be tagged by a software (#LancsBox) whose different tools show the most frequent lexis and n-grams that can be used in awareness-raising DDL² tasks. The communicative activities should motivate learners and compel them to carry out an investigative work which, in time, should yield the learners corpus - written evidence of learning and development.

References

- ASTON, G. 1997. Small and large corpora in language learning. In: LEWANDOWSKA-TOMASZCZYK, B.; MELIA, P. (Eds.), 1997, *PALC '97 Proceedings of the first annual conference*. Łódź: Łódź University Press. p. 51-62.
- GABRIELATOS, C. 2005. Corpora and Language Teaching: Just a fling or wedding bells? *TESL – EJ*, 8.4:1-37.
- GRANGER, S. 1998. The computer learner corpus: a versatile new source of data for SLA research. In: GRANGER, S. (Ed.), *Learner English on Computer*. London; New York: Addison Wesley Longman. pp. 3-18.
- JOHNS, Tim. Should you be persuaded: Two examples of data driven learning. In: JOHNS, Tim; KING, Philip (Eds.). *Classroom concordancing*. ELR Journal, v. 4. Birmingham: University of Birmingham, 1991. p. 1-16.



Preterit-Imperfect Development in Three Written Spanish Learner Corpora

Sophia Minnillo

University of California, Davis
smminnillo@ucdavis.edu

Despite the extensive body of research on tense-aspect acquisition (Bardovi-Harlig & Comajoan-Colomé, 2020), few studies have considered Spanish preterit-imperfect development using corpora of L2 writing. Prior investigations have shown a preference for analyzing data from forced-choice tasks (Salaberry, 2011), which cannot fully reflect learners' productive knowledge of tense-aspect. The preterit-imperfect distinction is especially challenging for L1 English students given the differences in morphological marking between English and Spanish. English does not mark perfective and imperfective aspect for past events in the same way as Spanish; in many contexts, the English simple past can be used for both aspectual cases. In the context of U.S. Spanish language programs, it is important to deepen understanding of how students acquire the preterit and imperfect given the difficulty of this construction for the largely English-dominant student population. The present study furthered knowledge of tense-aspect acquisition through the analysis of three learner corpora featuring writing samples of L1 English, L2 Spanish learners. The corpora included 1) the Corpus of Written Spanish of L2 and Heritage Speakers (COWS-L2H; Yamada et al., 2020), 2) *Corpus Escrito del Español L2* (CEDEL2; Lozano, 2009), and 3) *Corpus de Aprendices de Español* (CAES; Instituto Cervantes, 2020). The longitudinal data of beginner learners ($N = 26$) in the COWS-L2H corpus demonstrated that the preterit generally emerged before the imperfect. The cross-sectional data of beginner and intermediate learners in CEDEL2 ($N = 611$) and CAES ($N = 123$) highlighted the role of lexical aspect in preterit-imperfect production. Comparison of the L1 Spanish reference corpus in CEDEL2 ($N = 820$) and the L2 corpus showed an association between L1 and L2 token and contingent frequency, the extent to which a verb form is a frequent member of a tense-aspect category. Findings will be discussed in terms of the advantages for investigating preterit-imperfect development conferred by each corpus.

References

- Bardovi-Harlig, K., & Comajoan-Colomé, L. (2020). The aspect hypothesis and the acquisition of L2 past morphology in the last 20 years: A state-of-the-scholarship review. *Studies in Second Language Acquisition*, 42(5), 1137–1167.
- Corpus de aprendices de español (CAES)* (1.3). (2020). [Computer software]. Instituto Cervantes. <https://galvan.usc.es/caes/>
- Lozano, C. (2009). CEDEL2: Corpus escrito del español L2. In C. Bretones Callejas & et al. (Eds.), *Applied Linguistics Now: Understanding Language and Mind/La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente* (pp. 197–212). Universidad de Almería.
- Salaberry, M. R. (2011). Assessing the effect of lexical aspect and grounding on the acquisition of L2 Spanish past tense morphology among L1 English speakers. *Bilingualism: Language and Cognition*, 14(2), 184–202.
- 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.



A Phraseological Analysis of Korean EFL Learners' Use of *This/These* in Academic Writing

Yoon Namkung

Georgia State University
ynamkung1@student.gsu.edu

Writers have long been advised to avoid unattended *this/these* (*this/these* used without a following NP) in academic writing because they can confuse readers and disrupt cohesion (Swales, 2005). However, studies show that skilled L1 English speakers use unattended *this/these* for textual cohesion in academic writing (Gray, 2010; Römer & Wulff, 2010; Gray & Cortes, 2011; Wulff et al., 2012; Crossley et al., 2017; Jiang & Wang, 2018). To date, few studies examined L2 learners' use of (un)attended *this/these* (Oh, 2009; Jin, 2019; Lee et al., 2021). Also, *this/these* and their following words were examined in isolation rather than as part of multi-word units. Because formulaicity is ubiquitous in language and learners find formulaicity challenging (Meunier, 2012), analyzing L2 learners' use of *this/these* from a phraseological perspective is necessary.

This study examines Korean EFL learners' use of sentence-initial (un)attended *this/these* in academic writing by adopting a cluster analysis. It uses both a learner corpus and a native English speaker corpus to identify learners' errors and differences in their stylistic features (Granger, 2002). To examine the frequencies of sentence-initial (un)attended *this/these* in low-, intermediate- and high-level Korean EFL learners' and native English speaker students' academic writing (RQ1) and their sentence-initial *this/these* clusters (RQ2), argumentative essays from the Yonsei English Learner Corpus (Rhee & Jung, 2014) and Michigan Corpus of Upper-Level Student Papers (Römer & O'Donnell, 2011) were used. To answer RQ1, the frequencies of sentence-initial (un)attended *this/these* were calculated. To answer RQ2, the most frequent 2-, 3-, 4- and 5-word *this/these* clusters were extracted. Findings indicated that attended *this/these* were generally preferred over unattended *this/these*. As L2 proficiency increased, learners' use of unattended *this/these* increased. Clusters supporting arguments with evidence (e.g. *This is because*) and those elaborating on preceding discourse (e.g. *This means that*) were also absent from the low-level learners' data.

References

- Crossley, S. A., Rose, D. F., Danekes, C., Rose, C. W., & McNamara, D. S. (2017). That noun phrase may be beneficial and this may not be: Discourse cohesion in reading and writing. *Reading and Writing, 30*(3), 569-589.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.). *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins.
- Gray, B. (2010). On the use of demonstrative pronouns and determiners as cohesive devices: A focus on sentence-initial *this/these* in academic prose. *Journal of English for Academic Purposes, 9*, 167-183.
- Gray, B., & Cortes, V. (2011). Perception vs. evidence. An analysis of *this* and *these* in academic prose. *English for Specific Purposes, 30*, 31-43.
- Jiang, F., & Wang, F. (2018). 'This is because...': Authorial practice of (un)attending *this* in academic propose across disciplines. *Australian Journal of Linguistics, 38*(2), 162-182.
- Jin, H. (2019). On the anaphoric use of demonstratives *this/these* in L2 academic writing. *Journal of English for Academic Purposes, 38*, 62-74.
- Lee, J. J., & Tytko, T., & Larkin, R. (2021). (Un)attended *this/these* in undergraduate student writing: A corpus analysis of high- and low-rated L2 writers. *Journal of English for Academic Purposes, 50*, <http://doi.org/10.1016/j.jeap.2021.100967>
- Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics, 32*, 111-129.



Full papers



- Oh, S. Y. (2009). Korean college students' use of English demonstratives in argumentative essays. *English Teaching*, 64(1), 51-78.
- Rhee, S. -C., & Jung, C. K. (2014). Compilation of the Yonsei English Learner Corpus (YELC) 2011 and its use for understanding current usage of English by Korean pre-university students. *The Journal of the Korea Contents Association*, 14(11), 1019-1029.
- Römer, U., & Wulff, S. (2010). Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing Research*, 2(2), 99-127.
- Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): The design, compilation, and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 6(2), 159-177.
- Swales, J. (2005). Attended and unattended "this" in academic writing: A long and unfinished story. *ESP Malaysia*, 11, 1-15.
- Wulff, S., Römer, U., & Swales, J. (2012). Attended/unattended *this* in academic student writing: Quantitative and qualitative perspectives. *Corpus Linguistics and Linguistic Theory*, 8(1), 129-157.



A Corpus-Based Case Analysis on Syntactic Complexity in Russian ESL Learners' Writing

Anna Scherbakova

NRU HSE, Moscow

aniezka.sherbakova@gmail.com

Measuring the complexity of a learner text is considered to be a significant factor in assessing the level of foreign language proficiency. We aim to study syntactic complexity (SC), which is usually interpreted as the variety and degree of complexity of the syntactic structures that are present in a text.

The research was carried out based on 984 learner texts written in English by Russian speakers, which were collected in the corpus REALEC (Kuzmenko & Kututzov, 2014). Each text has a grade given by independent experts and information on the number of 7 types of syntactic errors identified by annotators.

This study examines methods of SC evaluation via automated tools for analysis of SC: TAASSC (Kyle, 2016), L2SCA (Lu, 2010), and Inspector (Lyashevskaya et al., 2021). It has not yet been established which SC constructions or errors in their use are often found among Russian learners of English. We hypothesize that there is a correlation between the level of language proficiency and the number of syntactic errors and values of SC parameters. Hence, the objective of our study is to answer the research questions: Which parameters of SC most accurately reflect the level of English proficiency among Russian speakers? How can we explain the results of SC evaluation? Is there a correlation between the level of language proficiency and the number of syntactic errors and SC? For the analysis we used rank correlation coefficients.

Consequently, the SC parameters of learner texts which correlate most with the essay grade or the number of syntactic errors were identified. We can't report a strong correlation (the maximum value of Spearman's correlation coefficient is 0.439). The correlation between the SC parameters and the number of syntactic errors was found to be much weaker than the correlation between the same parameters and the grade.

References

- Kuzmenko, E., & Kutuzov, A. (2014). Russian error-annotated learner English corpus: a tool for Computer-Assisted Language Learning. In Proceedings of the third workshop on NLP for computer-assisted language learning. NEALT Proceedings Series 22 / Linköping Electronic Conference Proceedings 107, 87–97.
- Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication (Doctoral Dissertation).
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International journal of corpus linguistics*, 15(4), 474-496.
- Lyashevskaya, O., Panteleeva, I., & Vinogradova, O. (2021). Automated assessment of learner text complexity. *Assessing Writing*, 49.



A longitudinal study of teenage learners of German as a second language in intensive classes in Germany

Julia Schlauch

Justus-Liebig-Universität, Gießen, Germany
julia.s.schlauch@germanistik.uni-giessen.de

Syntactic structures are a well-established topic in SLA research. For German as a second language mainly structuralist studies (e.g. Clahsen 1995, Czinglar 2017, 2018), aimed for a description of *natural* phases in the acquisition of word order. Focusing mainly on the question of an *age factor* in SLA, influences of learning environment and the *untutored vs. tutored* learning process have been of little interest. Therefore there are specific groups of learners, like lateral entry students (LES) in so-called *intensive classes*, that have yet been described by SLA- in general or L2-learner corpus research in particular.

Intensive classes are organized to enhance the language competence of recently migrated students to allow the participation in a regular class after a limited period. Thereby those classes are of immense importance for the educational success of the LES. Aim of this study is the creation of a two-year longitudinal corpus of spoken language of LES learners. Individual monthly sessions contain two parts of data collection: The first part consists of one elicited picture story in which participants have to retell the story to the interviewer and a non-present addressee. The second part comprises a free talk with the interviewer. Data transcription and annotation uses the EXMARaLDA partitur editor (s. Schmidt & Wörner 2014) and aims for a multi-level annotation (s. Reznicek et al. 2013).

The presentation focuses on first results on the acquisition of basic syntactic structures by LES undergoing a highly tutored learning process in comparison to the existing studies mentioned above. In addition, conceptual questions on the subsequent research interest of the project, the development of more complex syntactic structures preparing for the participation in the education system shall be discussed.

References

- Clahsen, Harald (1985): Profiling Second Language Development. A Procedure for Assessing L2 Proficiency. In: Kenneth Hyldenstam und Manfred Pienemann (ed.): Modelling and assessing second language acquisition. Clevedon u.a.: Multilingual Matters (Multilingual matters, 18), 283–332.
- Czinglar, Christine (2017): Finiteness and V2 in Second Language Acquisition: Longitudinal Evidence from Two Late Learners of German. In: Wiener Linguistische Gazette (WLG) 82/2017. Themenheft 11-11-17 Festschrift für Martin Prinzhorn. Herausgegeben von Clemens Mayr und Edwin Williams, 51–60.
- Czinglar, Christine (2018): Zweitspracherwerb im Jugendalter. Die Bedeutung des Alters und literaler Kompetenzen von neu zugewanderten Jugendlichen. In: Nora von Dewitz, Henrike Terhart und Mona Massumi (Hg.): Neuzuwanderung und Bildung. Eine interdisziplinäre Perspektive auf Übergänge in das deutsche Bildungssystem. Weinheim, Basel: Beltz Juventa, 158–173.
- Reznicek, Marc; Lüdeling, Anke; Hirschmann, Hagen (2013): Competing Target Hypotheses in the Falko Corpus. A Flexible Multi-Layer Corpus Architecture. Available online: https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/korpuslinguistik/mitarbeiterinnen-en/marc/ReznicekEA_toAppear.pdf (29.04.2021).
- Schmidt, Thomas; Wörner, Kai (2014): EXMARaLDA. In: Durand, Jacques (ed.) Handbook on Corpus Phonology. Oxford: Oxford University Press, 402–419.



A longitudinal study of Swedish upper secondary school students' vocabulary development

Christian Holmberg Sjöling
Mid Sweden University
Chsj1600@student.miun.se

The importance of vocabulary is continuously stressed as a central aspect of language learning and a recent study conducted in Sweden, where this study is also situated, concluded that there is a discrepancy between practice and research in terms of how vocabulary is acquired efficiently (Bergström, Norberg and Nordlund 2021: 1–13). The aim of this investigation was to investigate how lexical input from a textbook and use of individual vocabulary notebooks affected Swedish upper secondary school students' vocabulary acquisition by answering the four following research questions:

- What frequency and CEFR levels are represented in the vocabulary taught in the textbook?
- To what extent does the receptive vocabulary knowledge of the group as measured by the VLT change after exposure to the textbook?
- To what extent does the vocabulary production of the group change following exposure to the textbook?
- To what extent did the students benefit from the vocabulary notebooks?

A learner corpus was created with the help of 34 upper secondary school students from a northern region in Sweden in order to answer these questions and the research questions were, firstly, answered by analysing the vocabulary represented in the textbook in terms of frequency and CEFR levels. Then, Nation's Vocabulary Levels Test (e.g., Laufer and Nation 1999) was used to establish the group's receptive vocabulary knowledge while the group's vocabulary production was also analysed in terms of frequency and CEFR levels. The efficiency of the vocabulary notebook was, lastly, examined. The results indicate that the particular textbook was well-suited in terms of frequency and CEFR levels, the students' knowledge of K3 words improved and so did their vocabulary production. Unfortunately, the efficiency of the vocabulary notebook was not positive.

References

Primary sources

Christian's Learner English Corpus. 2019. Compiled by Christian Holmberg Sjöling.
Larsson, Gun-Marie, and Catrin Norrby. 2005. *Masterplan 2*. Stockholm: Liber AB.

Secondary Sources

Bergström, Denise, Norberg, Cathrine, and Marie Nordlund. 2021. "Words are picked up along the way' – Swedish EFL teachers' conceptualizations of vocabulary knowledge and learning." *Language Awareness*: 1–17.
Laufer, Batia and I. S. Paul Nation. 1999. "A vocabulary-size test of controlled productive ability." *Language Testing* 16: 36–55.



Once upon a time...A tale of learner corpus research

Sylviane Granger

List of references

- Alonso Ramos, M., García Salido, M. & Vincze, O. (2014). Towards a collocation writing assistant for learners of Spanish. <http://www.grupolys.org/biblioteca/AloGarVin2014a.pdf>
- Belz, J.A. & Vyatkin, N. (2005). Learner corpus analysis and the development of L2 pragmatic competence in networked intercultural language study: The case of German modal particles. *The Canadian Modern Language Review*, 62(1), 17-48.
- Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33/1: 1-17.
- Brezina, V. (2018). *Statistics in Corpus Linguistics. A Practical Guide*. Cambridge: Cambridge University Press.
- Callies, M. & Götz, S. (2015). Learner corpora in language testing and assessment. Prospects and challenges. In M. Callies & S. Götz (eds) *Learner Corpora in Language Testing and Assessment*. Amsterdam & Philadelphia: Benjamins, 1-9.
- Chuang, F. & Nesi, H. (2007). Grammar Talk: Developing computer-based materials for Chinese EAP students. In O. Alexander (ed.) *New Approaches to Materials Development for Language Learning*. Peter Lang, 315-330.
- Dagneaux E., Denness S. & Granger S. (1998) Computer-aided Error Analysis. *System: An International Journal of Educational Technology and Applied Linguistics* 26(2), 163-174.
- Díez-Bedmar, B. (2021). Error Analysis. In Tracy-Ventura, N. & Paquot, M. (eds) *The Routledge Handbook of Second Language Acquisition and Corpora*. New York & London: Routledge, 90-104.
- Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL - International Review of Applied Linguistics in Language Teaching*, 47(2), 157-177.
- Ebeling, S.O. & Hasselgård, H. (2015). Learner corpora and phraseology. In Granger, S., Gilquin, G. & Meunier, F. (eds). *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 207-230.
- Ebeling, S.O. & Hasselgård, H. (2021). The functions of n-grams in bilingual and learner corpora: An integrated contrastive approach. In Granger, S. (ed.) *Perspectives on the L2 Phrasicon: The view from learner corpora*. Bristol: Multilingual Matters, 25-48.
- Ellis, N. C., Simpson-Vlach, R., Römer, U., Brook O'Donnell, M. & Wulff, S. (2015). Learner corpora and formulaic language in second language acquisition research. In Granger, S., Gilquin, G. & Meunier, F. (eds) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 357-378.
- Forsberg Lundell, F. (2021). Formulaicity. In Tracy-Ventura, N. & Paquot, M. (eds) *The Routledge Handbook of Second Language Acquisition and Corpora*. New York & London: Routledge, 370-381.
- Gillard, P. & Gadsby, A. (1998). Using a learners' corpus in compiling ELT dictionaries. In Granger, S. (ed.) *Learner English on Computer*. London & New York: Addison Wesley Longman, 59-71.
- Gilquin, G., Papp, S. & Díez-Bedmar, M.B. (eds) (2008). *Linking Contrastive and Learner Corpus Research*. Amsterdam & New York: Rodopi.
- Golden, A., Jarvis, S. & Tenfjord, K. (eds) (2017). *Crosslinguistic Influence and Distinctive Patterns of Language Learning. Findings and Insights from a Learner Corpus*. Bristol: Multilingual Matters.
- Götz, S. (2015). Tense and aspect errors in spoken learner English: Implications for language testing and assessment. In Callies, M. & Götz, S. (eds) *Learner Corpora in Language Testing and Assessment*. Amsterdam & Philadelphia: Benjamins, 191-215.

Useful LCR references



- Granger, S. (1993). The International Corpus of Learner English. In J. Aarts, P. de Haan & N. Oostdijk (eds) *English Language Corpora: Design, Analysis and Exploitation*. Rodopi: Amsterdam & Atlanta, 57-69.
- Granger, S. (1996). From CA to CIA and back: an integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (eds) *Languages in Contrast. Text-based cross-linguistic studies*. Lund University Press: Lund, 37-51.
- Granger, S. (1998). The computerized learner corpus: a versatile new source of data for SLA research. In Granger S. (ed.) *Learner English on Computer*. London & New York: Addison Wesley Longman, 3-18.
- Granger, S. (ed.) (1998). *Learner English on Computer*. London & New York: Addison Wesley Longman.
- Granger S. (2003) Error-tagged learner corpora and CALL: a promising synergy. *CALICO* 20(3), 465-480.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research* 1(1), 7-24.
- Granger, S. (2019). Formulaic sequences in learner corpora: Collocations and lexical bundles. In Siyanova-Chanturia, A. & Pellicer-Sanchez, A. (eds) *Understanding Formulaic Language: A Second Language Acquisition Perspective*. Routledge, 228-247.
- Granger, S. (ed.) (2021). *Perspectives on the L2 Phrasicon: The View from Learner Corpora*. Bristol: MultilingualMatters.
- Granger, S., Dupont, M., Meunier, F., Naets, H. & Paquot, M. (2020). *The International Corpus of Learner English. Version 3*. Louvain-la-Neuve: Presses universitaires de Louvain. Trial version: <https://corpora.uclouvain.be/cecl/icle/trial/>
- Granger, S., Gilquin, G. & Meunier, F. (eds) (2015). *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.
- Granger, S. & Larsson, T. (2021). Is core vocabulary a friend or foe of academic writing? Single-word vs multi-word uses of THING. *Journal of English for Academic Purposes* 52, 1-13.
- Gyllstad, H. & Snoder, P. (2021). Exploring learner corpus data for language testing and assessment purposes: The case of verb + noun collocations. In Granger, S. (ed.) *Perspectives on the L2 Phrasicon: The view from learner corpora*. Bristol: Multilingual Matters, 49-71.
- Helland Gujord, A.-K. (2021). Crosslinguistic influence. In N. Tracy-Ventura & M. Paquot (eds) *The Routledge Handbook of Second Language Acquisition and Corpora*. New York & London: Routledge, 345-357.
- Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning* 50 (2): 245-309.
- Larsson, T., Egbert, J., & Biber, D. (Forthcoming). On the status of statistical reporting versus linguistic description in corpus linguistics: A ten-year perspective. *Corpora*, 17(1).
- Lozano, C. & Díaz-Negrillo, A. (2019). Using learner corpus methods in L2 acquisition research: The morpheme order studies revisited with Interlanguage Annotation. *Revista Española de Lingüística Aplicada* 32(1): 82-124.
- Lüdeling, A. & Hirschmann, H. (2015). Error annotation systems. In Granger, S., Gilquin, G. & Meunier, F. (eds). *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 135-158.
- Milton, J. & Chowdhury, N. (1994). Tagging the interlanguage of Chinese learners of English. In Flowerdew, L. & Tong, A.K.K. (eds) *Entering Text*. Language Centre, Hong Kong University of Science and Technology, 127-43.
- Milton, J. & Tsang, E. S.-C. (1993). A corpus-based study of logical connectors in EFL students' writing. In R. Pemberton & E. S.-C. Tsang (eds), *Studies in Lexis*. Language Centre, Hong Kong University of Science and Technology, 215-246.
- Möller, V. (2017). *Language Acquisition in CLIL and Non-CLIL Settings: Learner corpus and experimental evidence on passive constructions*. Amsterdam & Philadelphia: Benjamins.
- Myles, F. (2015). Second language acquisition theory and learner corpus research. In Granger, S., Gilquin, G. & Meunier (eds) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 309-332.

Useful LCR references



- O'Donnell, M. (2012). Using learner corpora to redesign university-level EFL grammar education. *Revista Española de Lingüística Aplicada* 1, 145–60.
- Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*, 18(3), 391-417.
- Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, 35(1), 121–145.
- Paquot, M. & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research* 3(1): 61-94.
- Richards, J.C. (2006). *Communicative Language Teaching Today*. Cambridge: Cambridge University Press.
- Swan, M. & Smith, B. (2001). *Learner English. A teacher's guide to interference and other problems*. Second edition. Cambridge: Cambridge University Press.
- Thewissen, Jennifer (2015). *Accuracy across Proficiency Levels: A Learner Corpus Approach*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Turton, N. D. & Heaton, J. B. (1996). *Longman Dictionary of Common Errors*. Harlow: Addison-Wesley Longman.