



Summer school “Learner Corpus Research - Theory and practical applications”

University of Bremen/Germany – August 27-31, 2018

organized under the aegis of the Learner Corpus Association

Day 1 – Colloquium (work-in-progress presentations by participants)

Schedule

13.30 - 15.30		SFG 1040
13.30 - 13:45	Christoph Beeh	
13:45 -14:00	Antonin Brunet	
14:00 - 14:15	Shujun Wan	
14:15 - 14:30	Richard Bonnie	
14:30 - 14:45	Youcef Bencheikh	
14:45 - 15:00	Rachel Rubin	
15:00 - 15:15	Abdelhakim Boubekri	
15:15 - 15:30	Kathrin Kircili	
15.30 - 16.00	<i>Coffee break</i>	
16.00 - 18.00		SFG 1040
16.00 - 16:15	Barry Kavanagh	
16:15 -16:30	Ingo Kleiber	
16:30 - 16:45	Benjamin Siegmund	
16:45 - 17:00	Maria Rudneva	
17:00 - 17:15	Eliane Lorenz	
17:15 - 17:30	Fernando Martín-Villena	
17:30 - 17:45	Teresa Quesada	

Abstracts

Christoph Beeh (University of Szeged/Hungary, Institute of German Studies)

Dulko: Towards a German-Hungarian Learner Corpus

Within a Research Group Linkage Programme funded by the Alexander von Humboldt Foundation launched in July 2017, an annotated German-Hungarian learner corpus is being built at the German Linguistics chair of the University of Szeged/Hungary. Data from learner (L1 = HU) essays and translations are collected in a controlled environment without use of dictionaries, meta data are given according to a proposed standard (Granger & Paquot 2017). This corpus aims at being compatible with the Falko learner corpora at the Humboldt University of Berlin, while still substantially extending and generalising the conception of target hypotheses and error annotation in Falko.

For the annotation scheme used in the Dulko project ("Deutsch-ungarisches Lernerkorpus"/German-Hungarian learner corpus) various transformation scenarios for the EXMARaLDA partitur-editor have been developed. These apply XSLT stylesheets to the EXMARaLDA XML data format in order to help the annotator tag learner texts in a semi-automatic way. These tools are actively maintained in a public repository on bitbucket (<https://bitbucket.org/nolda/exmaralda-dulko>). Files generated in EXMARaLDA (Dulko) shall be convertible to the format of the query tool ANNIS (<http://corpus-tools.org/annis/>). In the long term, it could be made accessible to the query tool KorAP (<https://korap.ids-mannheim.de>).

This contribution introduces the general set-up and aims of the project, its innovations and a concrete sample.

Hirschmann, Hagen & Andreas Nolda (forthc.). Dulko – auf dem Weg zu einem deutsch-ungarischen Lernerkorpus. In *Jahrbuch des Instituts für Deutsche Sprache* 2018, Berlin: de Gruyter.

Granger, Sylviane & Magali Paquot (2017). *Core metadata for learner corpora*. Draft 1.0. Unpublished manuscript, Louvain-la-Neuve: Université catholique de Louvain.

Youcef Bencheikh (Kasdi Merbah University of Ouargla/Algeria)

A Corpus Based Study of Lexical Collocations in the Writings of Algerian Undergraduate EFL Learners

The present paper is a PhD research in progress that endeavours to investigate lexical collocations use in the writing of Algerian undergraduate EFL learners (L1 Arabic). The compiled corpus consists of 163 untimed essays: Types: 5282 and Tokens: 56833. An in-depth analysis was conducted using Laurence Anthony AntConc (3.4.4w2014) programme to extract collocations. The types investigated in the present study are based on Benson et al (1997): L1 (verb +noun), L2 (adjective +noun), L3 (noun+ verb), L4 (noun1 of noun2), L5 (adverb+ adjective), L 6 (verb+ adverb) and L7 (noun+ noun). The extracted lexical collocations were evaluated for their frequency, types, and collocational strength against two reference corpora: British National Corpus (henceforth, BNC) and Louvain Corpus of Native English Essays (henceforth LOCNESS).

The results showed that learners tend to overuse (adjective + noun) collocational construction and underuse (adverb + adjective), and (verb + adverb) ones. The misuse of

collocations i.e. collocational errors were found at the level of all the constructions investigated but with varied frequencies. The collocational errors can be attributed to: (a) False concept hypothesized (b) Overgeneralisation (c) The use of synonymy, (d) Ignorance of rule restrictions (e) Negative transfer (f) Word coinage and (g) Approximation.

Abdelhakim Boubekri (Mohamed V University Rabat/Morocco)

The Acquisition of English *If*-Conditionals by Moroccan Learners of English

If-conditionals are among the most difficult constructions that challenge L2 learners and exhibit some variation among native speakers. In this respect, this dissertation examines the factors that influence the acquisition process of *if*-conditionals by L2 Moroccan learners of English with two different proficiency levels, advanced and intermediate. Brown's (1973) Cumulative Complexity principle is applied to determine the grammatical complexity of ten *if*-conditionals: generic, non-distanced predictive, weak-distanced predictive, strong-distanced predictive, epistemic, speech act, meta-metaphorical, meta-linguistic, elliptical, and meta-spatial. O'Grady's (1997) Developmental Law is used as the theoretical framework for predicting the acquisition order of the *if*-clause and the main clause of English conditionals. Hwang's (1979) and Fulcher's (1991) frequency rankings are also adopted to predict the expected order of the different types of *if*-conditionals based on input frequency. Adopting the new typology of conditional constructions suggested by Dancygier (1998) and Dancygier & Sweetser (2005), two written-gap filling tasks were given to 180 students in order to examine the influence that level of proficiency, grammatical complexity, input frequency, and the influence of L1 have on the acquisition of *If*-conditionals by L2 learners. Preliminary results suggest that none of the mentioned factors above can alone best explain the acquisition process of the different types of *if*-conditionals by Moroccan learners of English. In fact, they all interact in the acquisition process. Therefore, the challenge that L2 learners face when acquiring *if*-conditionals is due to the interaction of many factors in the process of acquisition.

Antonin Brunet (University of Poitiers/France)

A Corpus for Advanced Learners of French: Why and How?

The presentation that we propose focusses on the importance of building a corpus of advanced learners of French as a foreign language. So far this specific type of learners – with a B2 proficiency level or above – has been quite neglected by both researchers and methodologists. As consequences, their abilities and competences are often wrongly evaluated and only a few handbooks are available on the market leading to a lack of material support for teachers. In the field of language teaching though, corpus based approach – especially for the teaching of English as a foreign language, has already proven its efficiency. This revolutionary approach allows on the one hand to provide the learners with more authentic language, but it also allows to identify the learners' competences and difficulties with empirical data and angle our teaching methods without solely relying on our intuition. Beside the linguistic data contained in the corpus itself, metadata that can be collected while building the corpus which are also crucial for the analysis of some linguistic and sociolinguistic features.

In this talk we will present how we reinvested the methodological framework “Contrastive Interlanguage Analysis” initiated by Sylviane Granger and her team at the University of Louvain in order to build our own corpus of advanced learners of French for our thesis. We will show from our experience that despite all the advantages of a corpus based approach, this method can also be tricky to use since there are many difficulties that any researcher should be aware of.

Eliane Lorenz (University of Hamburg/Germany)

Acquisition and Use of Tense and Aspect in a Foreign Language. A Study on Monolingual and Bilingual Learners of English

This dissertation project is part of the highly researched areas of L2 and L3 acquisition, bilingualism, multilingualism, and cross-linguistic influence. The focus here lies on young, unbalanced bilingual heritage speakers of Russian, Turkish, and Vietnamese that grow up in Germany and learn English as L3 in school. Their performance in the use of tense and aspect of English based on a written and an oral picture description task is compared to the performance of monolingual German peers that learn English as their L2.

The study uses data from E-LiPS, a subproject of the Linguistic Diversity Management in Urban Areas (LiMA) Panel Study (LiPS) that was conducted at the University of Hamburg from 2009 until 2013 (Linguistic Diversity Management in Urban Areas, 2009-2013, directed by Peter Siemund and Ingrid Gogolin). The main research question is to find out whether both languages, the majority language German and the heritage language, are sources of CLI in third language acquisition or whether only one of the two languages is being transferred.

First results clearly show that (i) there is a difference between monolingual and bilingual learners of English; (ii) there seems to be a clear difference between formal correctness and correct meaning when producing English and therefore, the analysis of learner language needs to differentiate between form and meaning; and (iii) both previously acquired languages serve as sources for CLI, but German, as the majority language of the bilingual heritage speakers and the typologically closer language to English, mainly influences the English.

LiMA, Linguistic Diversity Management in Urban Areas-LiPS, LiMA Panel Study. 2009-2013. Project co-ordination LiPS: Prof. Dr. Dr. H. C. Ingrid Gogolin; ©LiMA-LiPS 2013. Hamburg: LiMA.

Barry Kavanagh (Inland Norway University of Applied Sciences)

Can Corpora Be Useful in English Language Teaching in Norway?

How familiar are teachers of English in Norway with corpus linguistics? I released a questionnaire for teachers on Qualtrics.com (for computer or phone), and there were 210 respondents, with a spread of grades taught, geographical coverage, and variation in teaching experience and teacher age. When asked about materials for English usage, the teachers do not mention corpora. However, most teachers at least sometimes base teaching on common pupil mistakes and get an overview from a collection of pupil texts. Teachers seem to use corpora in this basic sense. 193 responded to a question about corpus linguistics: 15% never heard of it, 39% had little or no idea what it is, 28% were fairly familiar with it, and 18% had already done some work with it. The gap between exposure to corpus concepts and corpus use in teaching

showed itself: only 12 teachers used corpora in teaching (setting aside the abovementioned pupil texts). The teachers who use corpora in teaching use it most for vocabulary and idiom, which may be useful to know for future integration of corpus work into education. My next step is to approach four teachers who use corpora in teaching and volunteered to be interviewed – this data will also be for the first article. Then I will provide training for some in-service teachers in autumn 2018. For this, I need to work out what tools and methods are best to show in-service teachers.

Kathrin Kircili (Justus Liebig University, Giessen/Germany)

Non-Canonical Syntax in Learner Languages: A Contrastive Interlanguage Analysis

In my presentation, I would like to provide an overview of my dissertation project on the use of so-called non-canonical syntactic patterns, i.e. those patterns that deviate from the standard S-V-X core structure in English (cf. Ward and Birner 2004; Biber et al. 1999), by learners from four different L1 backgrounds, namely German, Spanish, Turkish and Japanese. The phenomena considered include fronting, inversion, cleft-sentences, introductory-*it*, existential-*there*, right- and left dislocation (cf. Quirk 1985; Birner & Ward 1998; Biber et al. 1999). For the analysis, 13 different syntactic as well as pragmatic variables (e.g. information status, form and function of the sentence-initial element, form and function of the sentence-final element, type of non-canonical phenomenon, etc.) will be annotated for a total number of 25,000 sentences from the International Corpus of Learner English (ICLE; Granger et al. 2009) as well as the British and American student essays from the Louvain Corpus of Native English Essays (LOCNESS; <http://www.learnercorpusassociation.org/resources/tools/locness-corpus>) On the basis of these annotations and the subsequent in-depth quantitative (statistical) and qualitative analysis, the study aims at determining differences, similarities as well as preferences with regard to the employment of the considered syntactic structures in the individual varieties and is also interested in whether the investigated syntactic and pragmatic variables are robust predictors of the use of certain non-canonical patterns over others and/or whether there are correlations between the individual variables.

Apart from new findings in field of interlanguage syntax and pragmatics, this study is expected to provide some new insights into whether EFL performances can be traced back to L1 transfer or whether the employment of non-canonical structures is rather influenced by so-called language universals (cf. Lado 1957; Gass 1979; Comrie 1884; Gass 1984).

- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999): *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Birner, Betty J. & Gregory Ward. 1998. *Information Status and Noncanonical Word-Order in English*. Amsterdam: John Benjamins.
- Comrie, B. (1984). "Why Linguists Need Language Acquirers." In: W. E. Rutherford (Ed.). *Language Universals and Second Language Acquisition*.
- Gass, S. (1979): "Language Transfer and Universal Grammatical Relations." *Language Learning* 29(2), 327-344.
- Gass, S. (1984): "A Review of Interlanguage Syntax: Language Transfer and Language Universals." *Language Learning* 34(2), 115-132.
- Granger, S., E. Dagneaux and F. Meunier (2009). *International Corpus of Learner English*, Louvain: UCL.
- Lado, R. (1957): *Linguistics across Cultures*. Ann Arbor: University of Michigan Press.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985): *A Comprehensive Grammar of the English Language*. Harlow: Longman Group Limited.
- Ward, G. and B. Birner (2004): "Information Structure and Non-canonical Syntax." In: Horn, Laurence; Gregory Ward (Eds.). *The Handbook of Pragmatics*. Malden: Blackwell, 153-174.

Ingo Kleiber (Heidelberg School of Education, University of Heidelberg/Germany)
Constructing and Analyzing Locally Compiled Corpora for English Language Teaching

Locally compiled (learner) corpora, i.e. corpora compiled by the learners and teachers themselves, have a huge potential for language teaching, learning, and the analysis of learner language in general. However, there is still a relatively large gap between linguistic research, didactic research, and classroom practices with regard to the adoption of corpus linguistic approaches to foreign language teaching.

My current research project aims at closing this gap by bringing together state-of-the-art (corpus) linguistic methods and approaches, second language acquisition and didactic research as well as computational linguistics and machine learning.

In this work-in-progress presentation, I will present a new construction-grammar based approach to fruitfully integrating locally compiled corpora into foreign language teaching and learning. Furthermore, I will present some early work regarding a new software platform for constructing and analyzing such locally compiled corpora.

Fernando Martín-Villena (Universidad de Granada/Spain)
Analysis of the Interlanguage of L1 English-L2 Spanish Learners: Anaphora Resolution in Topic Continuity

Previous studies have shown deficits with anaphora resolution at the syntax-discourse interface in the production of L1 English-L2 Spanish learners. In particular, in topic-continuity contexts, learners produce infelicitous overt pronouns and noun phrases where a null pronoun is expected (Lozano 2009; Lozano 2016; Montrul & Rodríguez-Louro 2016; Rothman 2009). This research explores the instances of overproduction in topic continuity in L2 Spanish written compositions using a developmental corpus of L2 Spanish (beginner, intermediate, and advanced) learners compared against a native Spanish subcorpus from CEDEL2 (<http://cedel2.learnercorpora.com/>). Each anaphoric form was tagged using UAM Corpus Tool and was assigned a number of different tags regarding 1) their pragmatic (in)felicity; 2) the syntactic patterns in which they occur; and 3) the distance between the anaphoric form and their textual potential antecedent(s). The overall results reveal that learners overuse overt anaphoric forms where null pronouns are expected (inter alia Lozano 2009; Lozano 2016). In particular, native-like behaviour cannot be eventually attained at very advanced levels, thus supporting the Interface Hypothesis (Sorace & Filiaci 2006; Sorace 2011). In addition, the results also reveal that the L1 is a modulating factor: learners use felicitous null pronouns mostly with syntactic coordination from early stages, simply because this is possible in their L1 English. By contrast, in non-coordinate contexts there is a gradual increase of null pronouns, which eventually approaches native levels in very advanced learners. This indicates that learners gradually acquire the pragmatic constraints of AR at the syntax-discourse interface. Finally, results show that the anaphoric chains produced by learners reach native-like levels at advanced stages.

- Lozano, C. (2009). Selective deficits at the syntax-discourse interface: Evidence from the CEDEL2 corpus. In: N. Snape, Y-k. I. Leung & M. Sharwood-Smith (eds). *Representational Deficits in Second Language Acquisition* (pp. 127–166). Amsterdam: John Benjamins.
- Lozano, C. (2016). Pragmatic principles in anaphora resolution at the syntax-discourse interface: advanced English learners of Spanish in the CEDEL2 corpus. In: M. Alonso Ramos (ed.), *Spanish Learner Corpus Research: Current Trends and Future Perspectives* (pp. 236–265). Amsterdam: John Benjamins.
- Montrul, S., & Rodríguez-Louro, C. (2006). Beyond the syntax of the Null Subject Parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. In: V. Torrens & L. Escobar (eds.), *The Acquisition of Syntax in Romance Languages* [Language Acquisition and Language Disorders 41] (pp. 401–418). Amsterdam: John Benjamins.
- Rothman, J. (2009). Pragmatic deficits with syntactic consequences? L2 pronominal subjects and the syntax-pragmatics interface. *Journal of Pragmatics*, 41, 951–973.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1–33.
- Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339–368.

Rachel Rubin (Université catholique de Louvain, Louvain-la-Neuve/Belgium)
Lexicogrammatical Complexity in L2 Dutch

My PhD research is a response to a recent call to widen the scope of L2 complexity research to the lexis-grammar interface (Paquot 2017; Housen et al, in press), building on Paquot (2017) who looked at the phraseological dimension of language in EFL learner writing. Paquot (2017) operationalized phraseological complexity in terms of diversity and sophistication, adopting measures such as type-token ratios (TTR) and mutual information (MI), and found that measures of phraseological sophistication are better suited to index proficiency than measures of syntactic and lexical complexity, particularly at the B2 to C2 levels of the CEFR. To extend the cross-linguistic validity of metrics of phraseological complexity established in Paquot (2017), I explore whether phraseological, and lexicogrammatical complexity, more generally, can also contribute to the description of L2 Dutch performance and development. Lexicogrammatical complexity will be investigated via an analysis of high-frequency verbs and passive constructions in two corpora of L2 Dutch writing (under compilation), one cross-sectional and one longitudinal, and one native Dutch reference corpus. This study will expand on the metrics of sophistication used in Paquot (2017), i.e. MI and ratio of academic to total collocations, by assessing various alternatives, including collocation analysis. The focus of the project to date has been on data collection and corpus annotation, including the evaluation of corpus annotation tools on our dataset. The presentation will focus more particularly on the transition from the corpus annotation phase to the corpus analysis phase of the project.

- Paquot, M. (2017). The phraseological dimension in interlanguage complexity research. *Second Language Research*. <https://doi.org/10.1177/0267658317694221>.
- Housen, A., De Clercq, B., Kuiken, F. & Vedder, I. (in press). Multiple approaches to L2 complexity. *Second Language Research*.

Richard Bonnie (University of Hamburg/Germany)
The Production of Twi Distinct Vowels and Consonants by German learners: Towards the Creation of a Twi Learner Corpus

While the effects of first languages on second-language forms of European languages have been studied extensively, research on the process of acquisition and learning of African languages is still limited. The present study therefore investigates the production and perception of certain vowels and consonants by German-speaking learners of Twi. Twi phonology displays a number of distinct consonants and vowels (Dolphyne 1988, 1996) that are not present in German (Koenig & Gast, 2009), and based on an experiment conducted with ten German learners of Twi, this study first reports on how the learners produce these distinct Twi sounds. The productions of the learners are juxtaposed with those of Twi native speakers and subjected to auditory and acoustic analyses. Then the participants are taken through a perception experiment that investigates the perceptual robustness of these distinct Twi sounds to test whether learners can distinguish them from closely related sounds. The findings suggest that the students' perception of the sounds considerably determines their production. Since corpora of African languages are significantly lacking, the broader aim of this project is to compile a larger spoken corpus of learner Twi.

Dolphyne, F. A. (1996). *A Comprehensive Course in Twi (Asante) for the Non-Twi Learner*. Accra: Ghana University Press.

Dolphyne, F. A. (1988). *The Akan (Twi-Fante) Language: Its Sound Systems and Tonal Structure*. Accra: Ghana University Press.

Koenig, E., & Gast, V. (2009). *Understanding English-German Contrasts*. Berlin: Erich Schmidt.

Maria Rudneva (RUDN University, Moscow/Russia)

Academic Writing Within a University Setting: A Corpus-Based Perspective

The main objective of the present study is to examine academic writing development, compare and contrast L1 and L2 key aspects, provide meaningful resources for grammar and vocabulary instruction within a university setting. This research is intended as a case study of BSc and MSc students in the Environmental Faculty. They were given the same assignment – to present their research at an international students' conference and to prepare a paper for the conference proceedings. The papers were prepared in Russian, discussed with their research advisors and then translated into English by students themselves. I collected the following corpora for further investigation:

	BSc Students	MSc Students	Native Speakers PhDs
Russian	100 000 words	100 000 words	100 000 words
English	100 000 words	100 000 words	100 000 words

All students have the same L1 background (Russian), their L2 competence varies from A2/B1 in BSc students to B2/C1 in MSc students. I also introduced the NS corpus to investigate key points of improvement that should be targeted throughout the university course of English as well as to model developmental picture of learning process.

I intend to look into the following aspects:

- 1) I will look into general EAP and specific professional terminology and analyze their development towards NS norms.
- 2) I will conduct a comparative study of 1-st person pronouns in English and Russian research articles. I expect Russian students to use fewer 1st-person pronouns in their English articles

than native speakers due to the traditional collectivist approach to scholarly work that informs Russian academic discourse. I would also like to see if there is substantial overuse of passive voice constructions in English research papers by Russian students due to the same reasons.

3) On the pragmatic level I would like to target discourse markers and investigate the development of discourse marker competence.

Finally, I would like to consider the implications of these findings for EAP/ESP pedagogies.

Shujun Wan (Department of German language and linguistics, Humboldt University of Berlin/Germany)

Discourse structure in German argumentative essays: L1 German and Chinese learner German compared - Stage report: the challenges of annotation and evaluation

Corpus-based discourse research faces many difficulties due to its inevitable subjectivity and complexity. May as a result, among literature on L2 German, few analyze at the text and discourse level, although linguistic competence at the text level is very important for learners. This study is going to investigate the variety of discourse structure and connectives in L1 and L2 German argumentative writings, using 40 written argumentative essays from the learner corpus Kobalt-DaF (Zinsmeister et al., 2012). All the essays are on the same topic, i.e. "Is the situation of young people today better than of previous generations?" Among them, twenty are from advanced L2 learners of German (Chinese L1), and twenty from German native speakers, around 25,000 tokens in total.

Following the guidelines of the Potsdam Commentary Corpus (Stede, 2016), the author and a second annotator manually segment and annotate the 40 argumentative essays, based on the Rhetorical Structure Theory (Mann & Thompson, 1988) and the Content Zones (Stede, 2016). Given that the author is now at the stage of annotating, the presentation will mainly focus on the annotation procedure, specifically on the challenges, and the measurement of an interannotator agreement study at discourse level.

The author plans to finish the annotating works by the end of August, so that she can present more and discuss with teachers and classmates about future work during the summer school.

Mann, William C., and Sandra A. Thompson. 1988. "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization." *Text* 8 (3): 243–81.

Stede, Manfred. 2016. *Handbuch Textannotation. Potsdamer Kommentarkorpus 2.0*. Edited by Manfred; Stede. Potsdam: Universitätsverlag Potsdam. <https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/8276/file/pcss8.pdf>.

Zinsmeister, Heike, Marc Reznicek, Julia Ricart Brede, Christina Rosén, and Dirk Skiba. 2012. "Das Wissenschaftliche Netzwerk „Kobalt-DaF“." *Zeitschrift für Germanistische Linguistik* 40 (3): 457–58.

Benjamin Siegmund (University of Tübingen/Germany)

Building Academic Language in Science Classes at the Transition from Primary to Secondary School – An Intervention Study Using Scaffolding Techniques and Focus-on-Form Instruction

Language proficiency in the language of schooling is a major factor in the explanation of the difficulties children with migration background and/or low SES are faced with in the German educational system. While most children do not lack oral competencies in German, many struggle with the (conceptually) written academic language register (German 'Bildungssprache') (Eckhardt, 2008; Haberzettl, 2009) that is the most important medium of learning and performance assessment in school (Feilke, 2012; Halliday, 1978; Koch & Oesterreicher, 1985; Schleppegrell, 2004). As the mastery of this register is important for learning in all subjects, academic language teaching must be integrated into all subjects (Benholz & Lipkowski, 2000). It is the aim of my PhD Project to develop teaching strategies and materials for an integrated language support within primary school science teaching (4th grade) using focus-on-form strategies (e.g., Ellis 2016) and to investigate their impacts in a small intervention study in six real world classrooms (three intervention classes, three control classes, n=115). In order to test if children start using the focused academic language structures (conditionals, passive and others), they were asked to write small texts about experiments they observed in the pre- and in the posttest. In total, about 220 texts were elicited from fourth graders with heterogeneous language biography (with German as L1 or L2). I now want to annotate, analyze, and compare these texts using learner corpus methods.

- Benholz, C., & Lipkowski, E. (2000). Förderung in der deutschen Sprache als Aufgabe des Unterrichts in allen Fächern. *Deutsch Lernen*, Heft 1, 3–11.
- Ellis, R. (2016). Focus on Form: A Critical Review. *Language Teaching Research*, 20(3), 405–428.
- Eckhardt, A. G. (2008). *Sprache als Barriere für den schulischen Erfolg: potentielle Schwierigkeiten beim Erwerb schulbezogener Sprache für Kinder mit Migrationshintergrund* (1. ed.). Münster: Waxmann.
- Feilke, H. (2012). Bildungssprachliche Kompetenzen - fördern und entwickeln. *Praxis Deutsch*, 233, 4–13.
- Haberzettl, S. (2009). Förderziel: Komplexe Grammatik. *Zeitschrift Für Literaturwissenschaft Und Linguistik*, 39(153), 80–95.
- Halliday, M. A. K. (1978). *Language as social semiotic: the social interpretation of language and meaning*. London: E. Arnold.
- Koch, P., & Oesterreicher, W. (1985). Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch*, 36, 15–43.
- Schleppegrell, M. J. (2004). *The language of schooling. A functional linguistics perspective*. Mahwah, N.J.: Lawrence Erlbaum.

Teresa Quesada (Universidad de Granada/Spain)

Using Learner Corpora to investigate Anaphora Resolution: A comparison of L1 Spanish – L2 English learners and L1 English – L2 Spanish learners

Anaphora Resolution (AR) has been recently investigated in Second Language Acquisition (SLA). AR relates to how overt and null pronominal subject anaphors corefer with their antecedents in discourse. A particular instance of AR is the Position of Antecedent Strategy (PAS) (Carminati, 2002), which postulates that the syntactic position of the antecedent determines the anaphoric form: null pronouns bias towards an antecedent in subject position, while overt pronouns bias towards an antecedent in non-subject position as in (1).

(1) Lolai vio a Evaj mientras ellaj/Øi jugaba en el parque.

AR has been studied in L2 English and L2 Spanish using corpus and experimental methods. In particular, the PAS has been mostly studied experimentally in L2 Spanish (e.g., Alonso-Ovalle et al. 2002; Bel et al. 2016b; Jegerski et al. 2011). Importantly, these studies have investigated the PAS in rather artificial and have overlooked the use of anaphoric forms other than null or overt.

This study explores additional factors that may affect AR in general and the PAS in particular using learner corpora (LC). The production of L1 Spanish – L2 English learners and L1 English – L2 Spanish learners plus a control group of English and Spanish natives in the COREFL and CEDEL2 (respectively) is analysed. UAM Corpus Tool was used to create an annotation scheme and tag the data. This study tries to ascertain whether i) AR is affected by additional factors in L2 English and L2 Spanish; ii) the PAS occurs in natural corpus production; iii) the PAS behaves in native (and L2) as originally predicted by Carminati.

- Alonso-Ovalle, L., Fernández-Solera, S., Frazier, L., & Clifton, C. (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. *Journal of Italian Linguistics*, 14(2), 151–169.
- Carminati, M. N. (2002). *The Processing of Italian Subject Pronouns*. PhD thesis: University of Massachusetts at Amherst.
- Bel, A., & García-Alcaraz, E. (2015). Subject pronouns in the L2 Spanish of Moroccan Arabic speakers. In T. Judy & S. Perpiñán (Eds.), *The Acquisition of Spanish in Understudied Language Pairings* (pp. 201–232). Amsterdam: John Benjamins.
- Bel, A., García-Alcaraz, E., & Rosado, E. (2016a). Reference comprehension and production in bilingual Spanish: The view from null subject languages. In A. Alba de la Fuente, E. Valenzuela, & C. Martínez Sanz (Eds.), *Studies in Bilingualism* (Vol. 51, pp. 37–70). Amsterdam: John Benjamins.
- Bel, A., Sagarra, N., Comínguez, J. P., & García-Alcaraz, E. (2016b). Transfer and proficiency effects in L2 processing of subject anaphora. *Lingua*, 184, 134–159.
- Jegerski, J., VanPatten, B., & Keating, G. D. (2011). Cross-linguistic variation and the acquisition of pronominal reference in L2 Spanish. *Second Language Research*, 27(4), 481–507.
- Lozano, C., & Mendikoetxea, A. (2013). Learner corpora and second language acquisition: the design and collection of CEDEL2. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*. (pp. 65–100). Amsterdam: John Benjamins.