FLaRN

FORMULAIC LANGUAGE RESEARCH NETWORK

7th conference

BOOK OF ABSTRACTS

Vilnius University

28-30 June 2016

WELCOME TO FLARN 2016 IN VILNIUS

Dear Conference Participants,

Welcome to the Seventh International Conference of the Formulaic Language Research Network hosted by Vilnius University!

We hope that our Conference will become an interesting and inspiring continuation of the previous FLaRN conferences in Cardiff, UK (2004), Twickenham, UK (2005), Nottingham, UK (2008), Paderborn, Germany (2010), Tilburg, Netherlands (2012), and Swansea, UK (2014). What started in the United Kingdom as PhD student meetings has developed over the past years into a full international conference. FLaRN 2016 is attended by over 60 participants from 17 countries whereas papers to be presented cover a broad range of topics from different areas of linguistics. We would like to keep up the spirit of the previous FLaRN events and encourage both young and advanced researchers to share their ideas in open and stimulating discussions. We also hope that first-time FLaRNsters will make the most of the Conference and become ambassadors of the FLaRN community (www.flarn.org.uk) at their home institutions.

The organizing committee are particularly happy to welcome you at the Faculty of Philology, the venue of the Conference, which is situated in the Old Town of Vilnius. We suggest that you take a minute in your free time to explore the town outside the campus.

Finally, we would like to thank our colleagues and students at the Faculty of Philology for their support and help with organization of the Conference.

Let the Conference begin!

The FLaRN 2016 Organizing Committee

Jonė Grigaliūnienė
Rita Juknevičienė
Jurga Kasteckienė
Birutė Ryvitytė
Inesa Šeškauskienė
Laura Vilkaitė

Table of Contents

PROGRAMME

Tuesday, 28 June

12.00-13.00 REGISTRATION (Writers' Hall)

13.00-13.30 OPENING (Krėvės' Auditorium). Chair: Rita Juknevičienė

13.30-14.30 KEYNOTE (Krėvės' Auditorium)

Magali Paquot, University of Louvain (Belgium). Statistical collocations: a corpus-based construct and technique with theoretical implications and practical applications in second language research.

Chair**:** Jonė Grigaliūnienė

14.30-15.00 TEA/COFFEE (Writers' Hall)

15.00-16.30 PARALLEL SESSIONS:

	Krėvės' Auditorium Chair: Rita Juknevičienė	**German Room** Chair: Laura Vilkaitė	**Canada Room** Chair: Birutė Ryvitytė
15.00-15.30	**Benjamin Kremmel, Tineke Brunfaut, J. Charles Alderson** How important is knowledge of multi-word expressions for reading comprehension?	**Gareth Carrol, Kathy Conklin** The Queen and King are out of order: Effects of frequency on the processing of binomial phrases	**Jurga Cibulskienė** Rhetorical peculiarities of conventional and dead metaphors in live basketball commentaries
15.30-16.00	**Hiroyuki Matsumoto, Neil Heffernan** Can utilization of lexical units facilitate EFL reading proficiency?	**Cordula Glass** Collocations, corpora, and cognition: Measuring dynamic processes in phraseological research	**Svetlana Schukina** Formulaicity in English musical vocabulary
16.00-16.30	**Stuart Webb, Nguyen Thi My Hang** Examining second language receptive knowledge of collocation and factors that affect learning	**Saskia Lensink, Niels O. Schiller, Arie Verhagen** Parts and wholes - on how to empirically test for the cognitive reality of multi-word units	**Inesa Šeškauskienė, Justina Urbonaitė** From freedom to liberty to custodial sentence in the Criminal Code of the Republic of Lithuania, or some strategies of translating Lithuanian collocations with *laisvė* 'freedom' into English

16.30-18.00 WELCOME RECEPTION (Writers' Hall)

Wednesday, 29 June

9.00-10.30 PARALLEL SESSIONS

	Krėvės' Auditorium Chair: Laura Vilkaitė	**German Room** Chair: Justina Urbonaitė	**Canada Room** Chair: Inesa Šeškauskienė
9.00-9.30	**Diana Van Lancker Sidtis, John Sidtis** Characterization and classification of formulaic expressions: A view from language disorders	**Hana Al-Mutairi, Michael Rodgers** Learning polysemy of phrasal verbs and high frequency single-word verbs through spoken output	**Jelena Čolovic-Marković** To teach or not to teach explicitly: A qualitative examination of ESL writers' production of academic formulas in controlled and uncontrolled situations
9.30-10.00	**John Sidtis, Diana Van Lancker Sidtis** Formulaic expressions may have a special home in the brain: Evidence from functional imaging	**Ferdy Hubers, Wendy van Ginkel, Catia Cucchiarini, Helmer Strik, Ton Dijkstra** How native speakers see the light: A study on the role of literal word meanings in non-transparent idioms	**Rita Juknevičienė, Łukasz Grabowski** Comparing formulaicity of learner writing through phrase-frames: A corpus-driven study of Lithuanian and Polish EFL student writing
10.00-10.30	**Yuriy Kovalyuk** Finding the common ground: Phraseological theory and terminology revisited	**Ewa Guz** The relationship between fluency and the use of formulaic language in learner speech	**Agnieszka Leńko-Szymańska** A corpus-based analysis of the development of phraseological competence in EFL learners using the CollGram profile

10.30-11.00 TEA/COFFEE (Writers' Hall)

11.00-12.30 PARALLEL SESSIONS

	Krėvės' Auditorium Chair: Jūratė Ruzaitė	**German Room** Chair: Agnieszka Leńko-Szymańska	**Canada Room** Chair: Maria Fernandez-Parra
11.00-11.30	**Lenore Grenoble** Formulaic language and contact-induced change: Evidence from Russian	**Laura Vilkaitė** Reading collocations in a second language: An eye-tracking study	**Nihal Çalışkan** Bigramming a small Turkish corpus: the path to a contrastive analysis of genetically different languages
11.30-12.00	**Batia Laufer, Liubov Baladzhaeva** Attrition of collocations in the native language of bilingual and monolingual immigrants	**Beatriz González-Fernández** The acquisition of collocations in relation to other components of vocabulary knowledge	**Alexander Wahl** MERGE: An algorithm for the corpus extraction of continuous and discontinuous formulaic sequences
12.00-12.30	**Narcisse Torshizi, Magda Stroińska** Mechanism of transition from patriotism to nationalism: Some insights from the views of homeland in second generation Persians in Canada	**Natsumi Okuwaki** The processing of spoken and written formulaic sequences by L2 speakers	**Jolanta Kovalevskaitė, Loïc Boizou, Erika Rimkutė** Morphological realisations of two-word syntactic relations in the *Dictionary of the Lithuanian Nominal Phrases*

12.30-14.00 LUNCH

14.00-15.00 POSTER SESSION (Writers' Hall)

Jelena Čolovic-Marković: Does explicit teaching of topic-induced word combinations have an effect on ESL writers?

Danguolė Kalinauskaitė: Identifying clichés and their impact on the headlines of the Lithuanian news portals

Seval Kömürcü: Phrasal verb usage by Turkish learners of English: A corpus-based study

Justina Mandravickaitė: Formulaic language in the speech of politicians: Dimension of political position and political opposition in the Lithuanian Parliament (2008-2012)

Diana Van Lancker Sidtis & Seung-yun Yang: Formulaic language performance in left- and right-hemisphere damaged patients: Formal testing

Irina Vaynshteyn: Effects of age of arrival on use and understanding of formulaic expressions in Russian bilingual speakers

Rūta Zukienė: Alliterative and formulaic properties of poetic compounds in *Beowulf*

15.00-15.30 TEA/COFFEE (Writers' Hall) and GROUP PHOTO

15.30-16.30 PARALLEL SESSIONS:

	Krėvės' Auditorium Chair: Łukasz Grabowski	**German Room** Chair: Yuriy Kovalyuk	**Canada Room** Chair: Jurga Cibulskienė
15.30-16.00	**Geraint Paul Rees** The selection of EAP vocabulary: A formulaic solution	**Jūratė Ruzaitė** Formulaic vagueness markers in spoken Lithuanian	**Birutė Ryvitytė** Evaluative meaning and disciplinary values: A corpus-based study of the adjective *svarbus* 'important' in the Corpus Academicum Lithuanicum
16.00-16.30	**Keith Bateson** Fauxpology in public discourse: Manipulating formulaic sequences	**Mariusz Kamiński** A quantitative study of dictionary definitions	**Teresė Ringailienė** Collocations in popular scientific and academic discourse in English and Lithuanian

16.30-17.30 KEYNOTE (Krėvės' Auditorium)
Rūta Petrauskaitė, Vytautas Magnus University in Kaunas (Lithuania). Is formulaic language catching in translation?
Chair: Birutė Ryvitytė

18.00-20.00 CONFERENCE DINNER (University Café, the Grand Courtyard on the University Campus)

Thursday, 30 June

9.30-10.30 KEYNOTE (Krėvės' Auditorium)
Alison Wray, Cardiff University (the United Kingdom).
Getting a shoehorn in: How we work out the meaning of unknown formulaic expressions
Chair: Inesa Šeškauskienė

10.30-11.00 TEA/COFFEE (Writers' Hall)

11.00-12.30 PARALLEL SESSIONS

	Krėvės' Auditorium Chair: Jonė Grigaliūnienė	**German Room** Chair: Benjamin Kremmel	**Canada Room** Chair: Jurga Kasteckienė
11.00-11.30	**Jennifer Boutz, Claudia Brugman, Alia Lancaster** Formulaicity and creativity: case studies from online citations of Islamic religious texts	**Wendy van Ginkel, Ferdy Hubers , Catia Cucchiarini, Ton Dijkstra, Helmer Strik** Norming studies for idiom processing: native and non-native benchmarks	**Donata Berūkštienė** Lexical bundles in court judgments: A corpus-driven analysis
11.30-12.00	**Seung-yun Yang, Diana Van Lancker Sidtis** Listeners' identification and evaluation of Korean idiomatic utterances produced by persons with left- or right-hemisphere damage	**Glenn Hadikin** Building a community of practice? The role of language in citizen science	**Jiaoyue Chen** Reconceptualising formulaic language through students' perception
12.00-12.30	**Parvin Gheitasi** Formulaicity in the oral language production of young foreign language learners	**Carlos Prado-Alonso** On the register variation of idiomatic *Do So*	

12.30-13.00 Conference Closing (Krėvės' Auditorium)

KEYNOTE SPEAKERS

Magali Paquot (University of Louvain)

Statistical collocations: a corpus-based construct and technique with theoretical implications and practical applications in second language research

Usage-based research in corpus linguistics, psycholinguistics and cognitive linguistics has provided convergent evidence that lexis and grammar are inextricably intertwined and that word combinations, be they framed in terms of phraseological units, formulaic sequences or constructions, play crucial roles in language acquisition, processing, fluency, idiomaticity and change (e.g. Ellis, 1996; Sinclair, 1991; Wray, 2002; Schmitt, 2004; Goldberg, 2006). Phraseology is now at the forefront of theoretical and empirical developments in second/foreign language learning and teaching.

In this talk, I will focus on EFL learners' use of one specific type of phraseological units, i.e. statistical collocations. Statistical collocations involve "the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance" (Gries, 2008: 6). I will show how association measures such as the Mutual Information score, coupled with NLP techniques (lemmatization, part-of-speech tagging and parsing), can be used to describe L2 performance at the B2, C1 and C2 levels of the Common European Framework of Reference for Languages (Council of Europe, 2001). Collocational indices also prove to be better discriminators of proficiency level for academic texts written by L2 writers in content courses as compiled in the VESPA corpus when compared to traditional measures of syntactic complexity and lexical diversity.

I will round off this talk by discussing the theoretical implications and practical applications in second language research, language teaching and language testing.

Rūta Petrauskaitė (Vytautas Magnus University)

Is formulaic language catching in translation?

If we assume that formulaic language is distinctive for its repetitive, reusable and ready-made linguistic items, and if we believe that translations are made to be as adequate as possible, then we may expect that the language of translations is as formulaic as the language of the original texts.

The topic of investigation of this presentation is to what extent parallel corpora are made up of formulaic sequences or phrases and whether they have phrasal counterparts in translation. It draws on a previous research (Daudaravičius et al. 2004) on the detection of collocational boundaries and continues with the phrase-based models of statistical machine translation system Moses (Koehn et al. 2003). Moses was used to translate from Lithuanian to English languages and trained on the 750 million running words corpus of parallel texts translated both from and to Lithuanian. As the training process in Moses uses phrases to infer translation correspondences between the two languages it is supplied with the so-called LT-EN language model, .i.e. an extracted list of phrase to phrase translations with their estimated probabilities based on corpus-wide statistics.

The methodology of this pilot research focuses on the most frequent 100 phrases in each of the languages that were manually analysed in detail trying to find out whether a) the statistical chunking of the texts into phrases and b) the mapping of translation correspondences was accurate and intuitively appropriate. Since the top frequent phrases make up a considerable part of the language model, the results of their analyses can tentatively signal a comparable density of formulaic sequences.

References

Daudaravičius Vidas, Marcinkevičienė Rūta (2004). Gravity Counts for the boundaries of collocations. *International Journal of Corpus Linguistics*, 9(2): 321–348.

Koehn Philip, Och Franz Josef Och, Marcu Daniel (2003). Statistical phrase-based translation. *Proceedings of the HLT-NAACL 2003 Main Papers*. – Edmonton. 48–54.

Alison Wray (Cardiff University)

Getting a shoehorn in: how we work out the meaning of unknown formulaic expressions

> 'Took it off of a fat old gager a couple o' years back,' he explained, with engaging frankness. 'Prigged his tattler, too, but I sold that. I'm a great one for a pinch o' merry-go-up, and this little box just happened to take my fancy, and I've kept it. I daresay I'd get a double finnup for it, too,' he added, sighing over his own prodigality. 'It's worth more, but when it comes to tipping over the dibs there ain't a lock as isn't a hob-grubber.' (The Toll-gate, Georgette Heyer, 1954).

When we first encounter a new expression, how do we work out what it means? Although there has been research into how L2 learners approach unknown formulaic expressions, it has been difficult to make direct comparisons with what native speakers do, because of the ceiling effect that would arise in giving them ecologically valid expressions in their L1 – in other words, if an expression is formulaic in the L1, they will tend to know it already. In this talk, I describe an experiment (co-researched with Huw Bell and Katie Jones) that was able to present both L1 and L2 speakers with genuine, contextualised, formulaic expressions of English that were not known to either group. They were historical phrases researched and used by the British novelist Georgette Heyer in her works set in the Georgian and Regency periods (c.1800-1837). Through a think-aloud approach, participants gave commentaries on what they thought the expressions meant, and why. The results showed some important differences between the approach taken by L1 and L2 speakers, and suggest that increased knowledge of, and/or confidence with, an L2 enables a learner to get increasingly closer to behaving like an L1 speaker when encountering new expressions.

Jelena Čolovic-Marković (West Chester University)

Does explicit teaching of topic-induced word combinations have an effect on ESL writers?

The research on formulaic language in second-language writing emphasizes that in order to discuss a particular topic adequately, learners need to employ topic-induced word combinations (Erman, 2009). To illustrate, in an effective examination of the topic of international adoptions, word combinations such as adoptive parents or place a child for adoption are likely to be found.

The aim of the present study is to investigate whether or not there is a significant difference in ESL students' abilities to produce, in an essay, topic-induced word combinations between the students who receive explicit instruction and those who do not. In addition, through post-treatment interviews with a group of students from the treatment group, the study attempts to glean insights into the approaches ESL writers use for the production of the target structures.

Participants (N=54), ESL students enrolled in five high-intermediate writing classes in an IEP in the US, were assigned to the control (N=19) and experimental (N=35) groups based on their class enrolment. Over a period of four days, the experimental group received a 60-minute training on 15 topic-induced word combinations. The control group received no explicit instruction on the target word combinations but focused instead on other writing-oriented activities (e.g., journaling). The data included the scores participants received on the production of the target word combinations in an in-class impromptu timed essay.

One-way ANOVA revealed a statistically significant difference between the groups with the treatment outperforming the control group. The findings from the interviews indicated the students' perceptions of the usefulness of the target structures may influence whether or not students

produce them in writing. The findings suggest that explicit instruction of topic-indicted word combinations may be helpful for the students' learning of formulaic language for the purposes of writing and should be the focus of future experimental research.

References

Erman, B. (2009). Formulaic language from the learner perspective. In R. Corrigan, E. Moravcsik, H. Ouali & K..Wheatley (eds) *Formulaic language: Acquisition, loss, psychological reality and functional explanations*. Amsterdam: John Benjamins. 323–346.

Folse, K. (2008). Myth 1: Teaching vocabulary is not the writing teacher's job. In J. Reid (ed.) *Writing Myths: Applying Second Language Research to Classroom Teaching*. Ann Arbor, MI: University of Michigan Press.

Danguolė Kalinauskaitė (Vytautas Magnus University)

Identifying clichés and their impact on the headlines of the Lithuanian news portals

Formulaic language (FL) and its components (Kuiper, 2009), formulaic sequences including both fixed and flexible lexical items (LIs) are characterized by a stereotyped form, and conventionalized meaning (Kitzinger, 2000). Cliché is a part of FL that is conceived as a linguistic stamp due to its overuse. Cliché expressions preserve their formal, structural and semantic features in different contexts and discourses.

FL is taken here as a methodological framework for the investigation of clichés used in Lithuanian headlines. Clichés are considered from the point of view of their rigidity, with the aim to specify the methods for identifying clichés based on their frequent use in headlines.

Frequency, structure, and semantics of LIs were employed to answer research question, namely, what part of headlines is novel in comparison with formulaic? 164 possible clichés (formulaic sequences frequently used in the headlines) were selected for identification. A frequency list derived from the corpus of 4858 headlines was used as a basis to detect the most

frequent words of the selected clichés. Then the words were searched in the Corpus of the Contemporary Lithuanian Language (http://tekstynas.vdu.lt) for their most frequent collocates. The method allowed identification of well-structured and frequently used expressions as candidates to journalistic clichés that meet preassembled models (following Kuiper, 2000). Further analysis revealed that most clichés contain a) idioms used in different contexts but most frequently in journalistic discourse and b) other non-idiomatic fixed expressions typical of press genres and easily identifiable as journalistic clichés.

Corpus-based evidence points to the LIs that have become clichés due to both factors: their overuse and formulaic nature. The Lithuanian news headlines due to frequently exploited clichés could be ascribed to FL. In general, headlines are easily identifiable and predictable because of a big number of clichés that in their own right are made out of idioms. Therefore headlines as one more type could be added to the taxonomy of FL.

References

Kitzinger, C. (2000). How to resist an idiom. *Research on Language and Social Interaction 33(2)*: 121–154.

Kuiper, K. (2000). On the Linguistic Properties of Formulaic Speech. *Oral Tradition 15(2)*: 279–305.

Kuiper, K. (2009). *Formulaic genres*. Basingstoke: Palgrave Macmillan.

Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (ed.) *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins. 37–54.

Schmitt, N. and R. Carter. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (ed.) *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam / Philadelphia: John Benjamins Publishing Company. 1–22.

Van Lancker-Sidtis, D. (2012). Two-Track Mind: Formulaic and Novel Language Support a Dual-Process Model. In M. Faust (ed.) *The Handbook of the Neuropsychology of Language*. Oxford: Blackwell Publishing Ltd. 342–367.

Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction*. London: New York: Bloomsbury.

Wood, D. (2010). *Perspectives on Formulaic Language: Acquisition and Communication*. London / New York: Continuum International Publishing Group.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Wray, A. and M. R. Perkins. (2000). The functions of formulaic language: An integrated model. *Language & Communication 20*: 1–28.

Seval Kömürcü (Freiburg University)

Phrasal verb usage by Turkish learners of English: A corpus-based study

One of the phenomena that cause great difficulty even for advanced learners of English on the lexical level is the acquisition and active usage of phrasal verbs. They are complex both in terms of their grammatical form and their lexical meaning. The phenomena of idiomaticity and polysemy are two other features of some phrasal verbs that add to their difficulty. As a result of their difficulty, most English learners avoid using them for fear of making mistakes and have a tendency to use their one-word counterparts instead –regardless of the learners' L1 background (e.g. Dagut and Laufer 1985; Hulstijn & Marchena 1989; Laufer & Eliasson 1993; Liao and Fukuya 2002; Schmitt & Redwood 2011). However, although there are mostly one-word counterparts for phrasal verbs, "there is usually no total congruence, as the whole concept of synonymy is indeed always a question of degree" (Waibel 2007: 37): phrasal verbs "are often more specific in meaning than their "equivalents" and often carry connotations which their potential users must be aware of" (Cornell 1985: 275). As "the use of phrasal verbs is extremely common and a standard feature of good idiomatic English" (Alexander 1988: 153), phrasal verbs are of paramount importance for attaining a native-like command of English, and avoidance or underuse of phrasal verbs by learners will therefore result in unidiomatic language and stylistic deficits.

Taking its cue from corpus-based studies on the use of English phrasal verbs by foreign learners of various L1 backgrounds (e.g., Waibel 2007; Mazaherylaghab 2013; Chen 2013), the present study focusses on Turkish learner English, which has not previously been considered in this regard.

The Turkish sub-corpus (TICLE) of the International Corpus of Learner English (ICLE) is the basis for this study, and the Louvain Corpus of Native English Essays (LOCNESS) functions as the native speaker control corpus. Findings related to the most frequently used phrasal verbs found in the learner corpus will be presented – in terms of frequency of occurrence and also with respect to semantic and stylistic considerations.

References

Alexander, Louis G. (1988). *Longman English Grammar*. London: Longman.

Chen, Meilin (2013). Overuse or underuse. A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics 18(3)*: 418–442.

Cornell, Alan (1985). Realistic goals in teaching and learning phrasal verbs. *IRAL 23(4)*: 269–280.

Dagut, Menachem & Batia Laufer (1985). Avoidance of phrasal verbs – a case for contrastive analysis. *Studies in Second Language Acquisition 7*: 73–79.

Hulstijn, Jan H. & Elaine Marchena (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition 11*: 241–255.

Laufer, Batia & Stig Eliasson (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity. *Studies in Second Language Acquisition 15*: 35–48.

Liao, Yan & Yoshinori J. Fukuya (2002). Avoidance of phrasal verbs: the case of Chinese learners of English. *Second Language Studies 20(2)*: 71–106.

Mazaherylaghab, Hamzeh (2013). *Iranian Learner English: A corpus-based study of phrasal verb usage*. Ph.D. Thesis. Freiburg: University of Freiburg.

Schmitt, Norbert & Stephen Redwood (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin and Magali Paquot (eds) *A Taste for Corpora: In honour of Sylviane Granger*. Amsterdam: John Benjamins. 173–208.

Waibel, Birgit (2007). *Phrasal Verbs in Learner English: A corpus-based study of German and Italian students*. Ph.D. Thesis. Freiburg: University of Freiburg.

Justina Mandravickaitė (Vilnius University)

Formulaic language in the speech of politicians: dimension of political position and political opposition in the Lithuanian Parliament (2008-2012)

This paper explores the use of formulaic sequences (Wray 2013) in the decision making speeches of politicians in the Lithuanian Parliament focusing on political position and political opposition via corpus-led analysis of transcribed Lithuanian parliamentary speeches (term 2008-2012). Legislation at the Parliament consists of concrete acts that are typical for the political domain (eg., passing laws, voting, debating) where a relatively formal style as well as legal jargon and dialogical conventions for different interactions prevail (van Dijk 1997). Thus it promises significant coverage of formulaic language. As political position and opposition have different intentions and purposes, it indicates differences in language use as well.

The data for analysis is a part of corpus created for the project "Automatic Authorship Attribution and Author Profiling for the Lithuanian Language" (acronym ASTRA) (No. LIT-8-69). This part covers speeches only of members of the Parliament with at least 200 speeches. The minimum number of words in an individual speech is 100. All in all there are 2.15 million words for political position and 1.95 million words – for political opposition.

Analysis of formulaic parliamentary speeches was performed with the tool Formulib (Forsyth 2015). Firstly, 'formulexicon' was compiled out of most frequent 2-5-grams. Then coverage by sequences in this lexicon was calculated for each category (political position and political opposition) as well as separate documents. Afterwards highly typical and highly atypical documents for the aforementioned categories based on their coverage of the sequences in the 'formulexicon' were identified. The results are discussed in terms of their implications for the language of political decision making via perspective of political position and opposition.

References

Forsyth, Richard S. (2015). *FORMULIB: Formulaic Language Software Library*.

van Dijk, Teun A. (1997), What is political discourse analysis. *Belgian Journal of Linguistics 11(1)*: 11–52.

Wray, Alison (2013). Formulaic language. *Language Teaching 46(3)*: 316–334.

Diana Van Lancker Sidtis (New York University/Nathan Kline Institute)

Seung-yun Yang (Touro College/Nathan Kline Institute)

Formulaic language performance in left- and right-hemisphere damaged patients: Formal testing

Purpose: Formulaic expressions - conversational speech formulas, idioms, pause fillers, conventional phrases, proverbs, expletives, and so on, constitute a considerable proportion of everyday communication. Having its presence in language use, formulaic language presents a challenge to study under laboratory conditions. While some clinical and experimental studies have implicated a right hemisphere involvement, controversy remains surrounding cerebral processing with respect to comprehension and production of formulaic expressions. The purpose of this study was to further investigate hemispheric specialization for production and comprehension of formulaic and matched literal expressions by examining performance of individuals with unilateral brain damage compared with matched healthy controls (HC) utilizing structured tasks involving repetition, elicited speech, and comprehension.

Methods: Individuals with unilateral brain damage (RHD or LHD) due to cerebro-vascular accident and matched HC participants completed five different tasks (Repetition, Multiple choice sentence completion, Free form sentence completion, Picture-matching completion, and Written multiple choice) involving conversational speech formulas, idioms & proverbs, and matched literal expressions.

Results: The LHD and RHD groups showed significant differences between two types of sentences (formulaic versus literal expressions) in

the free form sentence completion and picture matching tasks, but not in the repetition or multiple choice (spoken or written) tasks.

Conclusions: The findings lend support to the dual process model of formulaic and literal expressions, implying a right hemisphere involvement in processing of formulaic expressions when an elicitation task for production and picture-matching for comprehension were utilized. Despite fine-granular measures applied to the repeated exemplars, subjects did not differ in articulation or prosody in the repetition of formulas, idioms, and matched novel expressions. The results suggest the important role of task and reveal the limitations of formal testing of formulaic language.

Irina Vaynshteyn (Touro College/Nathan Kline Institute)

Effects of age of arrival on use and understanding of formulaic expressions in Russian bilingual speakers

Background: Little is known about the effects of age on formulaic language acquisition in second language (L2) learners. Existing L2 literature has examined the effects of age on the acquisition of grammatical language, and not on formulaic language, despite difficulties associated with formulaic language acquisition in L2. The purpose of this research was to investigate the effects of age of arrival (AoA) to the USA on use and comprehension of formulaic expressions (FEs) in English and Russian by Russian bilingual speakers. It was hypothesized that L2 participants in both age groups would use smaller proportions of English FEs in conversation, and show lower scores on structured tasks as compared to the control group. It was predicted that the early group would demonstrate better performance than the later group in English FEs, whereas the later group would perform better in Russian.

Method: Participants were two groups of Russian bilingual speakers of English, the early-arriving group (AoA: 9-13 years) and the later-arriving group (AoA: 18-22 years), and a control group (monolingual speakers of

English). Performance data were acquired using spontaneous speech and structured testing in order to assess knowledge and use of FEs in both English and Russian.

Results: As expected, the control group performed significantly better on all English tasks than both bilingual groups. The later arriving group scored significantly higher than the early group on all formulaic tasks in Russian, and performed significantly better in Russian than English. Both bilingual groups scored higher on comprehension than production for English. Contrary to our prediction, the early arriving group did not perform significantly better than the later arriving group on the English formulaic tasks. They did perform better on the English than the Russian formulaic tasks.

Discussion: Linguistic input and brain maturation likely both play important roles in formulaic language acquisition.

Rūta Zukienė (Vilnius University)

Alliterative and formulaic properties of poetic compounds in Beowulf

The formulaic nature of Old English poetry and its predilection for decorative compounds have long been known for scholars. In Beowulf, unique among Anglo-Saxon poems because of its sheer length and heroic thematics, the frequency of compounds is remarkable: a compound of one or another form is found in every three lines on average. The present research is an attempt to discuss the relative weight of poetic or low-frequency compounds in the alliterative line in relation to their behaviour in or as formulas. Although in previous studies low-frequency poetic compounds were considered to contribute mainly to the formal aspects of poetic diction (i.e. metre, alliteration, repetition and variation), recent research in the field indicates that even in the case of pleonastic (or redundant) compounds, the complexity of their structure creates emphasis, which in turn highlights the contextual meaning of the

compounds. My aim is to investigate the syntactic types of poetic compounds in Beowulf, sieve them through the metrical system of alliteration and to correlate the results with their appearance in the formulaic sequences of the poem. The hypothesis is twofold: first, that the content-matter of the pleonastic elements of compounds is purposeful and therefore enhanced by their alliterative and formulaic patterns; second, that formulas aid the understanding of rare and original compounds.

References

Brodeur, Arthur Gilchrist (1959). *The Art of 'Beowulf.'* Berkeley: University of California Press.

Chapman, Don and Ryan Christensen (2007). Noun-Adjective Compounds as a Poetic Type in Old English. *English Studies*, Vol. 88(4): 447–464.

Davis-Secord, Jonathan (2016). *Joinings: Compound Words in Old English Literature*. Toronto: Toronto University Press.

Orchard, Andy (2005). *A Critical Companion to 'Beowulf'*. Cambridge: Boydell & Brewer.

Hana Al-Mutairi (University of Nottingham)

Michael Rodgers (University of Nottingham)

Learning polysemy of phrasal verbs and high frequency single-word verbs through spoken output

The development of a language learner's depth of knowledge of high frequency vocabulary and formulaic language has long been considered an essential but challenging endeavour (Nation, 1990). Much of the learning burden stems from the fact that the more frequently a word occurs, the more meaning senses it is likely to have (Schmitt, 2000). This is particularly problematic for the learning of verbs as they have been shown to be even more polysemous than nouns (Gentner, 1981; Miller & Fellbaum, 1991). Moreover, acquiring knowledge of the spoken form of polysemous vocabulary has been shown to be particularly problematic (Thornbury, 2002). Pushed Output (Swain, 1985) has been promoted as a potential method for learning the productive form of vocabulary. Previous research has empirically examined Pushed Output in relation to vocabulary gains; however, target items have predominantly been concrete nouns e.g. furniture (e.g. De La Fuente, 2002; Holster & de Lint, 2012). By contrast, the effect of production tasks on the acquisition of single-word and multiword verbs is an under-researched area (Nation & Webb, 2011). This study examines 100 Saudi EFL learners' acquisition of multiple meaning senses of high-frequency verbs and phrasal verbs through two treatment conditions: spoken output-based tasks and written vocabulary exercises. Acquisition of the meanings of the verbs was measured through receptive multiple-choice pre- and post-tests, and an oral post-test. A control group completed only pre- and post-tests. The results were analysed for the differences in vocabulary gains between the treatment groups and the difference in acquisition rates between single-word verbs and phrasal verbs. Results indicated that meaning senses of high-frequency verbs can be learnt at similar rate to that of phrasal verbs, although the depth of vocabulary acquisition differs depending on the method of learning. Explanations for the findings and pedagogical applications of this research are offered.

References

De La Fuente, M. J. (2002). Negotiation and oral acquisition of Spanish L2 vocabulary: the roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition 24(1)*: 81–112.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory 4*: 161–178.

Holster, T., & de Lint, D. (2012). Output tasks and vocabulary gains. *The Language Teacher 36(2)*: 3–10.

Miller, G., & Fellbaum, C. (1991). Semantic networks of English. *Cognition* 41: 197–229.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Boston, Massachusetts: Heinle & Heinle.

Nation, I. S. P., & Webb, S. (2011) *Researching and Analysing Vocabulary*. Boston, MA: Heinle Cengage Learning.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. G. Madden *Input in second language acquisition*. Roleway, MA: Newbury House Publishers. 235–253

Thornbury, S. (2002). *How to Teach Vocabulary*. Harlow: Longman.

Keith Bateson (McMaster University)

Fauxpology in public discourse: Manipulating formulaic sequences

Media discourse in the West is full of what seems to be official apologies offered by public figures as atonement for their public misdeeds, and now even the private. When one analyzes these 'apologies,' however, it becomes clear that instead of taking responsibility, expressing remorse, and promising that it will never happen again – as the speech act of apology requires in order to be felicitous – these apparent apologies obviously have different agendas. The widespread existence of such false apologies in public discourse is evident with the popularization of the term "fauxpology," which is "[w]hen a person makes it sound like they are apologizing when, in fact, they are just shifting the blame or using twisted logic to argue their way out of responsibility for their actions". This strategy is "[o]ften used by pro athletes, movie stars, and politicians when they get caught cheating, making absurd comments, or get arrested for drugs" (Urban Dictionary). It is a speech act that exploits the form of an apology, but does not

fulfil its mandate, serving, instead, the purpose of protecting the speaker from liability or other forms of litigation (Lazare 2005, 173).

Austin categorizes apology as a performative speech act and provides rules for evaluating the success (felicity) of the performative. Yet, the ease with which fauxpologies are deployed, and the extent to which they resemble felicitous apologies, suggests that the unique formulaic sequences structuring an apology (Schmitt 2004, 7) are modified in order to customize the original constraints, and thus dictating which set of words are allowed in the open slots of the sequences. In this paper I will specifically look at the Canadian government's manipulation of these formulaic sequences in its apology to the First Nations about residential schools and compare this historical apology to other international cases involving past atrocities.

References

Urban Dictionary: fauxpology. Web. 9 Feb. 2014.<http:// www.urbandictionary. com/define.php?term=fauxpology>.

Lazare, Aaron (2005). *On Apology*. Oxford: Oxford University Press.

Schmitt, Norbert (2004), ed. *Formulaic Sequences in Action: An Introduction. Formulaic Sequences: Acquisition, Processing and Use*. Introduction by Norbert Schmitt and Ronald Carter. Amsterdam: John Benjamins.

Turner, Victor W. (1969) *The ritual process: structure and anti-structure*. Chicago: Aldine Publishing.

Donata Berūkštienė (Vytautas Magnus University)

Lexical bundles in court judgments: a corpus-driven analysis

One of the characteristic features of legal language is its formulaicity. However, there has been little research available on the nature of frequently occurring sequences of words, e.g. collocations, idioms, lexical bundles, in different genres of legal texts. What is more, some descriptions of legal language simply repeat traditional observations or provide only intuitive lists of expressions typical of legal discourse. Thus there is a need for a corpus-driven, inductive methodology in the research of multi-word sequences in legal language since intuitively based investigations cannot tell what exact patters are fixed and how frequently they are

repeated in legal texts. Taking into consideration the above mentioned points, the aim of this paper is to identify typical lexical bundles frequently used in court judgments of the Court of Justice of the European Union. Lexical bundles are defined as commonly recurrent sequences of word combinations which are often semantically regular, structurally incomplete, and not idiomatic (Biber et al., 1999). In the course of investigation of lexical bundles, the methodological guidelines of corpus linguistics are followed. To explore the textual behaviour of fixed strings in court judgments, this paper concentrates on the frequency, structure and function of lexical bundles. For the purpose of this study, a corpus of 500, 000 words of court judgments in the English language has been created and examined. Lexical bundles in this research have been identified by n-gram extraction method using the corpus analysis toolkit AntConc 3.4.4.

Jennifer Boutz (University of Maryland)

Claudia Brugman (University of Maryland)

Alia Lancaster (University of Maryland)

Formulaicity and creativity: case studies from online citations of Islamic religious texts

A prominent style of formulaic language use in Arabic is the quotation or citation of hadith, the collected sayings and actions attributed to the Prophet Muhammad. For Muslims, hadith provide guidance on all matters of daily life and are often quoted by modern speakers in diverse contexts (Brown, 2014). These citations can vary widely in length. Sometimes a single phrase is enough to evoke the complete hadith; in other cases, long passages may be quoted. All are formulaic in being prefabricated sequences from the same source, stored and accessed as a whole, rather than combined in real time (Wray, 2002). As used in modern Arabic discourse, this language may be metaphorical and the connection to the discourse context implicit. In this work, we analyze hadith citations from Arabic-language online message boards in terms of metafunctions of language: ideational, interpersonal, and textual (Halliday, 1994). In the interpersonal realm, these citations act as markers of shared cultural identity and reliably demonstrate a shared social context (they may perform all of the first-order interpersonal

functions noted in Wray & Perkins, 2000, Table 2). However, like proverbs, their contributions to the rhetorical structure of the larger text are highly variant. They can be found in different discourse contexts with contradictory rhetorical meanings. Within the ideational function, we find instances of argumentation from authority and self-promotion. Often, a single citation demonstrates stable interpersonal and varying ideational metafunctions. As memorized chunks, the genre of hadith falls on the more formulaic end of a formulaicity continuum; however, we find creativity in the form as well as the function of the citations. Taking hadith citation as a style of rhetorical contribution, we describe similarities to and differences from the metafunctions of formulaic language enumerated in the literature.

References

Brown, J. A. C. (2014). *Misquoting Muhammad: The Challenge and Choices of Interpreting the Prophet's Legacy*. London: Oneworld Publications.

Halliday, M.A.K. (1994). *An Introduction to Functional Grammar*. New York: Routledge.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A., & M. R. Perkins (2000). The functions of formulaic language: an integrated model. *Language & Communication 20*: 1–28.

Nihal Çalışkan (Yıldırım Beyazıt University)

Bigramming a small Turkish corpus the path to a contrastive analysis of genetically different language

Turkish, as an agglutinative language, has a rich derivational and inflectional morphology. As a result, the equivalent of lexical bundles in Indo-European languages (Salem 1987, Butler 1997, Altenberg 1998, Biber et al. 1999, Biber et al. 2003, Biber et al. 2004, Cortes 2004, Biber and Barbieri 2007, Biber 2009) are generally observed on the level of free and bound morpheme combinations in Turkish. This study, following Kaneyasu's (2002) work on morphemic bundles in Japanese and Durrant's (2012) study on inflectional morpheme collocations in Turkish, aims to discuss the nature of Turkish data obtained from bigram analysis of a two million METU-Turkish Corpus, the first of its kind, (Say et al. 2002), and to make a contrastive analysis of Turkish and previously studied languages,

especially English. So far, we have extracted all bigrams using AntConc 3.4.3, and classified them into two groups: lexical and grammatical. The bigrams, which do not correspond to any category, generally due to the fact that there is no semantic relationship between the components, are excluded. A rough impression of the data is that the lexical bigrams are equal to collocations, and grammatical bigrams are to lexical/morphemic bundles. A detailed analysis of these grammatical bigrams will be likely to answer these questions:

- What are the structural types of lexical bundles in Turkish? Are there any differences between Turkish and previously studied languages, especially English?

- Which of the free morphemes are dominant in constructing bundles? To which grammatical elements of lexical bundles in Indo-European languages do these free morphemes correspond?

The study will largely depend on the manual analysis. This will enable us to answer a methodological problem: What are the methodological difficulties of working with an agglutinative language? What are the main problematic areas and how can they be treated?

References

Altenberg B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations, In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications*. Oxford: OUP. 99–122.

Biber D. (2009). A corpus-driven approach to formulaic language in English: Extending the construct of lexical bundle. In L. Eckstein and C. Reinfandt (eds) *Anglistentag 2008 Proceedings*. Berlin: Wissenschaftlicher Verlag Trier. 367–377

Biber D, Johansson S., Leech G., Conrad S., Finegan E. (1999). *Longman Grammar of Spoken and Written English*, London: Longman.

Biber D., Conrad S., Cortes V. (2003). Lexical bundles in speech and writing: an initial taxonomy, In A. Wilson, P. Rayson, T. McEnery (eds) *Corpus Linguistics by the Lune: a Festschrift for Geoffrey Leech*, Frankfurt/Main: Peter Lang. 71-92.

Biber D., Conrad S., Cortes V. (2004). If you look at…: lexical bundles in university teaching and textbooks. *Applied Linguistics 25(3)*: 371–405.

Biber D., Barbieri F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes 26:* 263–286.

Butler C. S. (1997). Repeated word combinations in spoken and written text: Some implications for Functional Grammar. In C. S. Butler, J. H. Connolly, R. A.

Gatward, and R. M. Vismans (eds): *A Fund of Ideas: Recent Developments in Functional Grammar*. Amsterdam: IFOTT, University of Amsterdam. 60–77.

Cortes V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes 23:* 397–423.

Durrant, P. (2013).Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory 9(1)*: 1–38.

Kaneyasu M., *From Frequency to Formulaicity: Morphemic Bundles and Semi-Fixed Constructions in Japanese Discourse*, (PhD Thesis), University of California: Los Angeles.

Salem, A. (1987). *Pratique des Segments Répétés*. Paris: Institut National de la Langue Française.

Say B., Zeyrek D., Oflazer K., Özge U. (2002). Development of a corpus and a TreeBank for present-day written Turkish. In K. İmer ve G. Doğan (ed.) *Proceedings of the Eleventh International Conference of Turkish Linguistics*, Eastern Mediterranean University, Cyprus, August 2002. 183–192.

Gareth Carrol (University of Birmingham)

Kathy Conklin (University of Nottingham)

The Queen and King are out of order: Effects of frequency on the processing of binomial phrases

Of the vast array of formulaic subtypes, binomials remain relatively unexplored, at least compared to units such as idioms and collocations. Binomials are sequences of *noun and noun* or *verb and verb*, such as *salt and pepper* or *king and queen*. Such units may be considered to be "statistical idioms (Baldwin and Nan Kim, 2010), in that their formulaicity derives from their highly conventionalised word order rather than any semantic idiosyncrasies. Hence, despite having an identical surface meaning, *salt and pepper* is much more frequent than *pepper and salt* (in English), and native speakers show a clear processing advantage for the "correct" form (Carrol, 2015; Siyanova et al., 2011).

Corpus-based accounts of binomial ordering principles (e.g. Lohmann, 2012; Mollin, 2012) propose a complex set of linguistic and cultural factors. In this study we explore how far processing patterns are simply due to frequency effects. We presented readers with a series of invented binomials: logical *X and X* sequences which are not frequent based on corpus data and which do not have a conventional ordering, e.g. *grass and leaves* or *wires and plugs*. We embedded such phrases within short story contexts multiple times to see whether:

i) Overall processing time for the phrase decreases as a function of frequency of encounter

ii) Encountering the reversed form leads to a processing cost, once the "correct" form has been encountered multiple time.

We use eye-tracking to compare reading patterns for invented and attested binomials, considering factors such as number of encounters, corpus frequency, semantic association strength and predictability (based on cloze test scores). We compare behaviour for native and non-native speakers to establish whether both sets of speakers are sensitive to frequency effects for multiword combinations. We discuss the results in terms of usage based and lexical networks views of the mental lexicon.

References

Baldwin, T., & Nam Kim, S. (2010). Multiword Expressions. In N. Indurkhya & F. Damerau (eds) *Handbook of Natural Language Processing (2nd Edition)*. Boca Raton/London: CRC Press. 267–292.

Carrol, G. (2015). *Found in Translation: A Psycholinguistic Investigation of Idiom Processing in Native and Non-native Speakers*. Unpublished PhD thesis. University of Nottingham.

Lohmann, A. (2012). A processing view on order in reversible and irreversible binomials. *Vienna English Working Papers 21(1)*: 25–50.

Mollin, S. (2012). Revisiting binomial order in English: ordering constraints and reversibility. *English Language and Linguistics 16(1)*: 81–103.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition 37(3)*: 776–784.

Jiaoyue Chen (University of Southampton)

Reconceptualise formulaic language through students' perception

Formulaic language is used as an umbrella term to describe the formulaic aspect of languages, including English (Wray, 2002). This language phenomenon has been studied extensively in the areas of corpus linguistics, pragmatics, psycholinguistics and second language learning research.

Myles (2016) points out that there are two main approaches in existing formulaic language research: speaker-internal and speaker-external approach. The speaker-internal approach refers to "what is formulaic language in the language outside the speaker", with focus on formal properties, frequency and pragmatic aspects, whereas the speaker-external approach refers to "psycholinguistic units for a given speaker", that is the language is stored holistically and/or processing advantage.

However, there is consensus that what is formulaic in the language is not necessarily so in the mind of speakers (Wray, 2002, 2008; Schmitt, 2004; Myles, 2016), the linguistics clusters and processing units are not always overlapped. On the other hand, there are very few studies focusing on psycholinguistic formulaic language in advance learners with speaker-internal approach.

This paper, as part of my PhD research, aims to reconceptualise formulaic language through speaker-internal approach, that is, students' perspective, by investigating the students' understanding towards formulaic language, and their learning, use and comments on the teaching of formulaic language.

Two rounds of interviews were conducted at the beginning and end of one semester (five months). 12 interviewees were willing to take part in the study. Through the bottom-up and top-down approach, the interview transcripts were analysed, which provides insightful information on the students' perception on formulaic language. This study, at the end, also offers some implications in formulaic language teaching and learning in the context of the classroom, especially for advance learners in EFL settings.

References

Myles, F. (2016). *The pyscholinguistic construct of formulaicity in second language learners.* Plenary speech at FLaRN 2016, Swansea University, Swansea, UK.

Schmitt, N. (2004). *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam: John Benjamins.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Jurga Cibulskienė (Lithuanian University of Educational Sciences)

Rhetorical peculiarities of conventional and dead metaphors in live basketball commentaries

Steen's three-dimensional model of metaphor (2008, 2013) has stimulated a heated discussion about rhetorical impact or the lack of rhetorical impact of conventional and dead metaphors. According to Steen, metaphor is not just a matter of language and thought but also of communication. This way he distinguishes deliberate and non-deliberate metaphors, claiming that novel and some conventional metaphors are deliberate metaphors, which means they are deliberately created for communicative purposes, whereas dead and strongly conventionalized metaphors are considered as non-deliberate metaphors, which means they have no communicative value or, in other words, they lack rhetorical impact. However, Steen's ideas were challenged by a number of scholars. For example, Gibbs (2011), providing neurolinguistic and psycholinguistic evidence, asserts that both deliberate and non-deliberate metaphors, in Steen's terms, have similar communicative value due to our unconscious thinking.

In line with Gibbs' ideas, the paper attempts to analyze the possibility of rhetorical impact of dead and conventional metaphors in live basketball commentaries. To achieve this aim, a corpus of 46,559 words of transcribed Lithuanian live basketball commentaries was constructed and analyzed using the AntConc concordance program. The problem of conventionality versus novelty of metaphors was analyzed on the basis of the metaphor of CONFLICT following a three-step procedure suggested by Charteris-Black (2005). First, metaphorical expressions were identified in the analyzed discourse, then they were interpreted by relating them to conceptual metaphors, and finally they were explained in terms of rhetorical impact. The paper argues that though most metaphorical expressions used in the analyzed discourse are highly conventionalized and naturally comprehended as basketball terminology, we cannot downplay their influence on the spectators' subconscious minds. Extremely frequent use of the CONFLICT metaphor is aimed at arousing the spectators' emotions, or pathos (Cibulskienė, 2014).

References

Charteris-Black, J. (2005). *Politicians and Rhetoric. The Persuasive Power of Metaphor*. Palgrave Macmillan.

Gibbs, R. W. Jr. (2011). Are 'deliberate' metaphors really deliberate? A question of human consciousness and action. *Metaphor and the Social World 1(1)*: 26–51.

Cibulskienė, J. The Conflictual nature of metaphors in live basketball commentaries: a corpus-driven Approach. *Man and the Word 16(3)*: 36–60.

Steen, G. (2008). The paradox of metaphor: Why we need a three-dimensional model of metaphor. *Metaphor and Symbol 23*: 213–241.

Steen, G. (2013). The contemporary theory of metaphor – now new and improved! In Francisco Gonzálvez-García, María Sandra Peña Cervel and Lorena Pérez Hernández (eds) *Metaphor and Metonymy revisited beyond the Contemporary Theory of Metaphor: Recent developments and applications*. Amsterdam: John Benjamins. 27–65.

Jelena Čolovic-Marković (West Chester University)

To teach or not to teach explicitly: A qualitative examination of ESL writers' production of academic formulas in controlled and uncontrolled situations

Vocabulary research in corpus linguistics reveals that formulas are frequent and functional in academic prose (Biber & Conrad, 1999). Despite this, many ESL writers find their use challenging (Li & Schmitt, 2008). To help ESL writers learn to employ academic formulas in an expert-like manner, it has been suggested that they be explicitly taught (e.g., Simpson-Vlach Ellis, 2010). However, little attention has been paid to how direct teaching affects ESL writers' production of academic formulas.

This study investigates how students' abilities to produce academic formulas (i.e., a selected set from Simpson-Vlach and Ellis, 2010) in controlled (C-test) and uncontrolled (essay) situations differ between the students who receive explicit instruction and those who do not. The study involves a qualitative analysis of the target academic formulas produced in pre/post C-tests and two multi-draft essays written by 53 ESL students (23 in the control group and 30 in the experimental) enrolled in high-intermediate writing classes in an IEP in the US. The academic formulas produced in C-tests were analyzed manually. The target structures in essays were first extracted through the use of Text-Lex Compare v.2. 2 (Cobb, 2010) and then analyzed by hand.

The results reveal that when compared to the students who do not receive explicit instruction, those who do are able to produce, on a C-test, more target academic formulas and do so with greater accuracy. The results indicate that in essays, the treatment group, in contrast to the control, is able to produce a slightly greater variety of the target formulas and has a tendency to alter the original form of academic formulas. Although the findings suggest that explicit teaching of academic formulas may have a positive effect on ESL writers, more learning of academic formulas needs to take place.

References

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard and S. Oksefjell (eds) *Out of Corpora: Studies in Honor of Stig Johansson*. Amsterdam: Rodopi. 181–190.

Cobb, T. (2010). *Text Lex Compare* (Version 2.2). Retrieved from http://www.lextutor.ca/text_lex_compare/

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing 18:* 85–102.

Simpson-Vlach, R., & Ellis, C. N. (2010). An academic formulas list. New methods in phraseology research. *Applied Linguistics 31*: 487–512.

Parvin Gheitasi (Umeå University)

Formulaicity in the oral language production of young foreign language learners

Multiword units of language known as formulaic sequences are pervasive and essential in communication (Wray 2002). Although the identification of the main issues related to formulaicity has been under investigation in a number of research studies (Wood 2002; Conklin and Schmitt 2008), the present study is different in investigating the oral production of young learners in an EFL context. The results will shed further light on the process of early foreign language acquisition and also the role of formulaic sequences in young learners' language production.

This presentation reports on a study, which investigated the production and functions of formulaic sequences in the oral language production of young foreign language learners. A classroom with 10 students in the age range 9 to 10 years was

observed and video recorded for 14 sessions (90 minutes per session). The observations were accompanied by two elicitation tasks. The collected speech samples were transcribed chronologically. Formulaicity was defined and identified learner internally and based on pre-established criteria, which is adopted and adapted from the literature in the field (e.g. Wray 2002, Myles 2004). Learners' language production was analysed with regard to the type of input they received in the classroom and the context in which they were produced.

The results provided evidence of incidental learning of formulaic sequences from input; the language input provides instances for the learners to learn multi word units. In addition, formulaic sequences played different roles in the language production of the young learners. These sequences provide young language learners with an instrument to overcome their lack of knowledge, to improve their fluency, and to enjoy some language play. Formulaic sequences were used as a strategy to economize effort on processing and also to buy time for processing. Moreover, the data provided examples illustrating communicative functions of formulaic sequences where the use of formulaic sequences was affected by the relationship between the speaker and hearer. The data provided evidence regarding the segmentation stage; that is learners might be able to distinguish the component parts of the sequences. On the other hand, the language samples show learners' difficulties with some individual units within a sequence.

Cordula Glass (Universität Erlangen-Nürnberg)

Collocations, corpora, and cognition: Measuring dynamic processes in phraseological research

In many cases corpus research plays a central role in modern research on collocations (Evert 2005; Bartsch 2004). Furthermore, measures of collocational strength are frequently used to extract learner-relevant word combinations for the compilation of EFL material (Nesselhauff 2004, de Cock 2004). Cognitive research, however, suggests that patterns which emerge in the language acquisition process are rather to be regarded as dynamic and emergent structures (Bybee 2010; Larsen-Freeman/Cameron 2008; Ellis 2006; Tomasello 2005; Dąbrowska 2004). Thus, despite the fact that some measures seem to be an adequate way to

portray the judgments of adult native speakers (Michelbacher/Evert/Schütze 2011; Evert/Krenn 2005), corpus-based selection of collocations regarding association measures runs the danger to reduce collocations to a rather static phraseological phenomenon.

In order to test the cognitive value of corpus linguistic methods for more dynamic, usage-based models, this study contrasts corpus data with judgment data of over 200 native speakers from four different age groups. In a pseudo-longitudinal survey it then compares the general acceptance of collocational combinations as well as their age and context related acceptance scores with the most common association measures (t-score, z-score, MI) and also more cognitive methods of corpus analysis like collostructional analysis (Stefanowitsch/Gries 2003).

The results show that while association measures are able to extract established collocations they have difficulties accounting for the varying degrees of flexibility within a collocational combination. This then might give the impression that all collocations share a certain degree of idiomaticity and unpredictability which seems to obscure the fact that - like other linguistic patterns, such as morphological markers or grammatical constructions (Goldberg 2006, Bybee 2010) - knowledge of collocations as well develops in stages as, for example, predicted by Wray and Perkins (2000), constructionist approaches (Langacker 2009; Goldberg 2006; Stefanowitsch/Gries 2003) or theories based on complex adaptive systems (Ellis/Larsen-Freeman 2009; Larsen-Freeman/Cameron 2008). Thus, this study suggests that - especially in a language acquisition context – corpus data needs to be balanced by more speaker related methods like elicitation or judgment tasks.

References

Bartsch, Sabine (2004). *Structural and Functional Properties of Collocations in English.* Tübingen: Gunter Narr Verlag.

Bybee, Joan (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Cock, Sylvie de (2004). Preferred sequences of words in NS and NNS speech. *Belgian journal of English language and literature*. 225–246.

Dąbrowska, Ewa (2004). *Language, Mind and Brain - Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh: Edinburgh University Press.

Ellis, Nick (2006). Language Acquisition as Rational Contingency Learning. *Applied Linguistics 27(1)*: 1–24.

Ellis, Nick and Diane Larsen-Freeman (2009). *Language as a complex adaptive system*. Malden, MA/ Oxford/ Chichester: Wiley-Blackwell.

Evert, Stefan (2005). T*he Statistics of Word Co-Occurrences. Word Pairs and Collocations*. Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart, URN urn:nbn:de:bsz:93-opus-23714

Evert, Stefan and Krenn, Brigitte (2005). Using small random samples for the manual evaluation of statistical association measures. *Computer Speech & Language 19(4)*: 450–466.

Goldberg, Adele (2006). *Constructions at Work - The Nature of Generalization in Language*. Oxford/ New York: Oxford University Press.

Langacker, Ronald W. (2009). *Investigations in Cognitive Grammar*. Berlin/ New York: Mouton de Gruyter.

Larsen-Freeman, Diane and Lynne Cameron (2008). *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press.

Michelbacher, Lukas, Stefan Evert and Hinrich Schütze (2011). Asymmetry in corpus-derived and human word associations. *Corpus Linguistics and Linguistic Theory, 7(2)*: 245–276.

Nesselhauff, Nadja (2004). *Collocations in a Learner Corpus*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Stefanowitsch, Anatol and Stefan Gries (2003). Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics 8(2)*: 209–243.

Tomasello, Michael (2005). *Constructing a language - A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Wray, Alison and Michael Perkins (2000). The functions of formulaic language: an integrated model. *Language and Communication 20*: 1–28.

Beatriz González-Fernández (University of Nottingham)

The acquisition of collocations in relation to other components of vocabulary knowledge

Vocabulary knowledge is usually understood as the combination of various word knowledge components such as spelling, pronunciation, form-meaning link, derivatives and collocations (Nation, 2013). Nevertheless, although this multidimensional nature of vocabulary knowledge is widely acknowledged among

vocabulary researchers, few studies have examined the acquisition of multiple components concurrently (e.g., Schmitt, 1998; Webb, 2007; Li & Kirby, 2015). As a consequence, it is still unclear how these different dimensions of vocabulary knowledge are acquired and relate to each other, which leads to a lack of a generally accepted theory of vocabulary acquisition.

One of the central word knowledge dimensions is the collocational knowledge. Collocations are described as words that co-occur together more frequently than would be expected by chance (Biber, Johansson, Leech, Conrad, & Finegan, 1999). This study explores the acquisition of collocations within a framework of general vocabulary knowledge by asking the following question: How does collocational knowledge relate to the different components of vocabulary knowledge?

In order to answer this question, the present study investigated 144 Spanish learners of English with a battery of eight tests, which measured their receptive and productive knowledge of four key lexical components: the form-meaning link, derivatives, polysemy and collocations. This allowed for the comparison between collocational knowledge and other aspects of word knowledge.

The data was analysed using Structural Equation Modelling (SEM), a group of statistical techniques designed to test the validity of theoretical models that can be used to study the causal relationships between variables. The results of the SEM analysis indicated that collocational knowledge interrelates with the rest of the vocabulary knowledge components, although the strength of these connections vary depending on the dimension of word knowledge.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Li, M., & Kirby, J.R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics 3(5)*: 611–634.

Nation, I.S.P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge: Cambridge University Press.

Schmitt, N. (1998). Tracking the incidental acquisition of second language vocabulary: A longitudinal study. *Language Learning 48:* 281–317.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics 28(1)*: 46–65.

Lenore A. Grenoble (University of Chicago)

Formulaic language and contact-induced change: Evidence from Russian

Research has increasingly shown the importance of formulaic sequences in language change, acquisition and grammaticalization; usage and frequency are key factors in language change (Bybee 2003, Bybee & Cacoullos 2009; Diessel 2007; Wray 2012).

There has been sustained contact between Russian speakers and speakers of other languages for centuries. The last few decades have seen increasing language shift to Russian, resulting in profound changes in the phonology, lexicon and morphosyntax of Eurasian indigenous languages. This situation provides rich material for examining the mechanisms and results of language change in process.

Data from contact situations in Russian Eurasia show two effects of formulaic sequences: (1) pre-fabs in the source language provide models for contact-induced change, and (2) fixed formulae provide chunks of texts that 'rememberers' invoke when asked to speak the (forgotten) language (Dwyer 2012). In the first type the Russian pre-fabs provide models that are borrowed or adapted into the indigenous language, while in the second type speakers whose primary language is Russian resort to using remembered formulae in efforts to speak the indigenous language. The two different scenarios illustrate the different ends of the processes of acquisition and recall in spontaneous language use.

Adapting a formalism from Bao's (2010) work on substratum transfer in Singapore English, this paper focuses on the grammatical replication of two Russian constructions: (1) the modal *nado* 'necessary', used in Russian in an impersonal construction with a dative logical subject; and (2) the conditional formed with the particle by and the L-participle of the verb. Both the *nado*-construction and the by-construction are replicated in a variety of languages, and show similar borrowing pathways. Data are taken from a database of published and field recordings of typologically and genealogically different languages spoken in different regions: Aleut (Aleut-Inuit-Yupik); Evenki (Tungusic); Itelmen (Chukotko-Kamchatkan); and Kalmyk (Mongolic).

References

Bao, Zhiming (2010). A usage-based approach to substratum transfer: the case of four unproductive features in Singapore English. *Language 86(4)*:792–820

Bybee, Joan (2003). Mechanisms of change in grammaticization: the role of frequency. In Brian D. Joseph & Richard D. Janda (eds) *The Handbook of Historical Linguistics*. Oxford: Blackwell. 602–623.

Bybee, Joan & Rena Torres Cacoullos (2009). The role of prefabs in grammaticization: How the particular and the general interact in language change. In Roberta Corrigan, Edith Moravcsik, Hamid Ouali, & Kathleen Wheatley (eds) *Formulaic Language,* Volume I. Amsterdam: John Benjamins. 187–217.

Diessel, Holger (2007). Frequency effects in language acquisition, language use, and language change. *New Ideas in Psychology 25*:108–127.

Dwyer, Arienne M. (2012). Uprooted and replanted: recontextualizing a genre. In Filiz Kıral (ed.) *Turkic Historical Linguistics*. Istanbuler Texte und Studien. Würzburg: Ergon Türk Dilleri Araşırmaları 21.1: 47–82.

Wray, Allison (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics 32*: 231–254.

Ewa Guz (The John Paul II Catholic University of Lublin)

The relationship between fluency and the use of formulaic language in learner speech

While some evidence has been provided for the facilitative function of formulaic sequences in L1 production (Pawley and Syder 1983; Peters 1983; Wray and Perkins 2000; Wray 2002; Wray 2008; Pawley 2009), an analogical relationship between formulaicity and productive fluency in L2 has received little scholarly attention in recent years. Wood (2001, 2004, 2006, 2007, 2008, 2009) conducted a series of longitudinal studies in which gains in productive fluency over time were attributed to increases in learners' repertoires of automatized lexical phrases. Guz (2013, 2014) analyzed the relationship between breakdown and speed fluency and the use of formulaic sequences in native and non-native speech (L1 Polish, L2 English) using a cross-sectional paradigm and reported significant positive correlations between the number of formulaic sequences used and speed fluency.

This paper aims at investigating the relationship between formulaicity and fluency in learner speech basing on two distinct conceptualizations of formulaicity. Using samples of learner speech drawn from two data pools: an 12,679-word corpus of

monologic speeches delivered in English by fifty-three Polish academic students (C1-C2) and the spoken component of the PLEC learner corpus - PELCRA (Pęzik, 2012), we analyse the relationship between productive fluency and formulaic language use adopting two different methodologies for identifying formulaic sequences in learner data. First, we adopt an automated corpus-driven extraction procedure and identify the most frequent co-occurring sequences of 2 and more words using Compleat Lex Tutor's N-gram Phrase Extractor software (Cobb, 2015). Second, a more traditional, linguistic definition of formulaic sequences is utilized relying on Erman and Warren's (2000) restricted exchangeability test along with Wray's (2008) 11 diagnostic criteria for assigning intuitive judgements about formulaicity. Those candidate sequences are assigned the formulaic status if they appear in at least one of the twelve selected dictionaries of regular and idiomatic English. The resulting formulaic strings are then removed from the data. Breakdown and speed fluency of the samples are measured before and after the removal of the formulaic material using a set of objective temporal measurements recently proposed as valid indices of learner productive fluency by Bosker et al. (2013). The pre- and post-removal fluency scores are then compared. It is hypothesized that the fluency scores of formula-deprived speech will be lower than those of formula-rich samples suggesting formulaic sequences contribute to fluency. We also predict a high degree of variance between the scores depending on the methodology used and type of sequence identified as formulaic.

References

Bosker, H. R., Pinget, A. F., Quene, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing 30(2)*: 159–175.

Cobb, T. (2015). *N-Gram Phrase Extractor* [computer program]. Accessed 24 October at http://lextutor.ca/n_gram/.

Erman, B., and B. Warren, (2000). The idiom principle and the open choice principle. *Text 20(1)*: 29–62.

Granger S., & M. Paquot, (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (eds) *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins. 27–49.

Guz, E. (2013). Investigating the relationship between oral fluency and the use of formulaic sequences in the L2 speech of advanced learners of English. Paper presented at the 3rd International Conference on *Teaching and Learning Speaking in a Foreign*

Language: Speaking in a foreign language: Psycholinguistic and sociolinguistic perspectives, Konin, Poland.

Guz, E. (2014). Formulaic sequences as fluency devices in the oral production of native speakers of Polish. *Research in Language 12(2)*: 113–129.

Pawley, A. (2009). Grammarians' languages versus humanists' languages and the place of speech act formulas in models of linguistic competence. In R. Corrigan, E. A. Mo- ravcsik, H. Ouali, & K. M. Wheatley (eds) *Formulaic Language: Volume 1: Distribution and Historical Change*. Amsterdam: John Benjamins. 3–26.

Pawley, A. & F. H. Syder. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards, and R. W. Schmidt (eds), *Language and Communication*. London: Longman. 191–225.

Peters, A. M. (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.

Pęzik P. (2012). Towards the PELCRA Learner English Corpus. In P. Pęzik (ed.) *Corpus Data across Languages and Disciplines*. [Lodz Studies in Language. Vol. 28]. Frankfurt am Main: Peter Lang.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review 57*: 573–589.

Wood, D. (2004). An empirical investigation into the facilitating role of automatized lexical phrases in second language fluency development. *Journal of Language and Learning, 2(1)*: 27–50.

Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review 63*: 13–33.

Wood, D. (2007). Mastering the English formula: Fluency development of Japanese learners in a study abroad context.. *JALT Journal 29*: 209–230.

Wood, D. (2008). Mandarin Chinese speakers in a study abroad context: Does acquisition of formulaic sequences facilitate fluent speech in English? *The East Asian Learner (3)2*: 43–62.

Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics 12(1)*: 39–57.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Wray, A., & M. R. Perkins (2000). The functions of formulaic language: an integrated model. *Language and Communication 20(1)*: 1–8.

Glenn Hadikin (University of Portsmouth)

Building a community of practice? The role of language in citizen science

Citizen science is a relatively new phenomenon in which scientists outsource some of the processing of data to volunteers. The large data sets now assembled by electronic means are often too massive to be dealt with by a single scholar or even a team. One solution is to turn to the public, in a process akin to crowd sourcing. The crucial difference in citizen science is that it is the public's time and energy that is being solicited, not their financial support. Consequently, the success of a citizen science project depends on the project manager's ability to recruit and retain large numbers of volunteers; to communicate with them effectively so that they can fulfil their role; and to encourage interaction and discussion so that they are engaged and committed. There is, thus, a fundamental linguistic question: in what ways do language choices and behaviours facilitate or hinder the recruitment, retention, commitment and contribution of citizen scientists?

At FLaRN 2014 I presented an early version of my Lexical Selection model (Hadikin, 2015) and I will follow up by showing how the structure we are behaves in a corpus of citizen science forum discussions. I ask to what extent we can consider the world's largest citizen science website - the Zooniverse - a community of practice and what other strings of language are associated with the more successful projects when compared with the less successful projects. This paper is a first report on findings from 20 corpora gathered in this research and gives an overview of the specificity of citizen science discourse.

References

Hadikin, G. (2015). Lexical Selection and the Evolution of Language Units. *Open Linguistics,* vol. 1, 458–466, doi: 10.1515/opli-2015-0013

Ferdy Hubers (Radboud University Nijmegen)

Wendy van Ginkel (Radboud University Nijmegen)

Catia Cucchiarini (Radboud University Nijmegen)

Helmer Strik (Radboud University Nijmegen)

Ton Dijkstra (Radboud University Nijmegen)

How native speakers see the light: A study on the role of literal word meanings in non-transparent idioms

A highly debated issue in formulaic language research is the exact role literal word meanings play during idiom processing (Cacciari & Tabossi, 1988; Peterson et al., 2001; Rommers et al., 2013; Sprenger et al., 2006; Swinney & Cutler, 1979; Titone & Libben, 2014). A tacit assumption in many of these studies is that the native speakers involved in the experiments know the idiomatic expressions and their precise meanings for the simple reason that they are native speakers. In many studies subjects are asked to answer questions about their familiarity with the expressions under study, but they are not usually tested on their precise knowledge of the expressions and the exact meanings they assign to them. However, one can imagine that the extent to which subjects do indeed know the expressions and what they exactly mean can have a major impact on the results obtained.

Within the framework of our research on Dutch formulaic language perception, production and acquisition, we were interested to determine to what extent the idiomatic expressions we intend to use in our experiments are indeed known and interpreted correctly by Dutch native speakers, as these data are going to represent benchmarks in our research.

To gain more insight into the extent to which native speakers know idiomatic expressions and interpret them correctly, a study was conducted in which native speakers of Dutch answered open questions about the precise meaning of Dutch expressions and completed multiple choice items. We compare the results for the different tasks and analyse to what extent self-reported familiarity, transparency and image ability are related to performance. The results increase our

understanding of native knowledge of idiomatic expressions and provide new insights into their usability for research on idiom processing.

References

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research 22(2)*: 115–144. doi:10.1191/0267658306sr263oa

Hillert, D., & Swinney, D. (2001). The Processing of Fixed Expressions During SentenceComprehension. In A. Cienki, B. J. Luka, & M. B. Smith (eds) *Conceptual Structure, Discourse, and Language*. Stanford: CSLI Publications. 107–121

Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition 27(5)*: 1223–1237.

Raposo, A., Moss, H. E., Stamatakis, E. a., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia 47(2)*: 388–396. doi:10.1016/j.neuropsychologia.2008.09.017

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience 25(5)*: 762–76. doi:10.1162/jocn_a_00337

Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language 54(2)*: 161–184. doi:10.1016/j.jml.2005.11.001

Swinney, D., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior 18(5)*: 523–534. doi:10.1016/S00225371(79)90284-6

Titone, D., & Libben, M. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon 9(3)*: 473–496. doi:10.1075/ml.9.3.05tit

Wendy van Ginkel (Radboud University Nijmegen)

Ferdy Hubers (Radboud University Nijmegen)

Catia Cucchiarini (Radboud University Nijmegen)

Ton Dijkstra (Radboud University Nijmegen)

Helmer Strik (Radboud University Nijmegen)

Norming studies for idiom processing: native and non-native benchmarks

Idiomatic expressions are known to be particularly challenging for L2 learners (Cieślicka, 2015; Conklin & Schmitt, 2008; Steinel et al. 2007; Titone et al. 2015). Research has shown that L2 idiom processing is affected by idiom properties such as transparency, imageability, frequency and cross-language overlap (Steinel et al. 2007; Titone et al. 2015). Studies on idiom processing rely heavily on norming studies to gauge these properties of idioms and to establish benchmarks for performance. An interesting question with respect to L2 idiom processing is whether native performance should be taken as the reference or rather the performance of advanced L2 learners. So far little research has been devoted to comparing the performance of native speakers and L2 learners on idiom comprehension in relation to relevant features of idiomatic expressions such as frequency, transparency, imageability and cross-language similarity.

In this paper we report on a study in which both native speakers and advanced German learners of Dutch were tested on their receptive knowledge of 175 Dutch idiomatic expressions. All subjects were tested on both recognition and recall of the Dutch idiomatic expressions. We analyze the effect of the various idiom properties on receptive knowledge and make comparisons between the results of native speakers and advanced L2 learners. The outcomes provide insight into the usability of native and non-native benchmarks for receptive knowledge and their variability as function of important idiom properties.

References

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research 22(2)*: 115–144. doi:10.1191/0267658306sr263oa

Cieślicka, A. (2015). Idiom acquisition and processing by second/foreign language learners. In Heredia, R.R. & Cieślicka, A.B. (eds) *Bilingual Figurative Language Processing*. 208–244.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics 29(1)*: 72–89.

Steinel, M.P., Hulstijn, J.H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition 29*: 449–484.

Titone, D., Columbus, G., Whitford, V. Mercier, J. & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. In Heredia, R.R. & Cieślicka, A.B. (eds) *Bilingual Figurative Language Processing*. 171–207.

Rita Juknevičienė (Vilnius University)

Łukasz Grabowski (Opole University)

Comparing formulaicity of learner writing through phrase-frames: A corpus-driven study of Lithuanian and Polish EFL student writing

Acquisition of formulaicity in a foreign language is a difficult task, and research into learner corpora offers an abundance of evidence of how formulaic language is (mis)used by foreign learners of English. One of the ways to investigate formulaicity in learner language is the lexical bundles approach which has been used in many studies (e.g. Ädel & Erman 2012, Chen & Baker 2010, Juknevičienė 2013, Paquot 2013, 2014). This paper deals with similar multi-word units, termed phrase-frames, which are automatically generated by computer software. Phrase-frames are n-grams which are identical except for one word (Fletcher 2002-2007), e.g. it is * to, in the * of etc. where the gap is a variable 'wildword'. While phrase-frames have been investigated in the British National Corpus (Fletcher 2010, Gray & Biber 2013) and specialized English corpora (Forsyth & Grabowski 2015, Grabowski (in print), Fuster-Marquez 2014), research into learner corpora seems to be limited to Römer (2009) who investigated phrase-frames in expert and student, including non-native, academic writing.

Our study compares Lithuanian and Polish learner writing in terms of phrase-frames and aims to answer two questions: (1) Does the analysis of phase-frames highlight differences between advanced Lithuanian and Polish EFL learners?; (2) Are there any L1 specific tendencies in the range and distribution of phrase-

frames that could be accounted for by L1 transfer? Two corpora, namely, LICLE for Lithuanian learners and PICLE for Polish, were used to retrieve the data, both compiled as components of the ICLE project (Granger et al. 2009). For reference, we used data from the BNC and LOCNESS representing academic essays by British and American students (CECL). The kfNgram (Fletcher 2002-2007) software was used to generate phrase-frames while WordSmith Tools (v. 5, Scott 2008) was used for concordancing.

The preliminary findings reveal a number of differences between Polish and Lithuanian learners yet it remains debatable whether they arise due to interference from students' L1 or whether there are other factors (topic-related differences, teaching methods etc.) that account for the findings.

References

Ädel, A. & B. Erman (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: a lexical bundles approach. *English for Specific Purposes 31*: 81–92.

CECL (Centre for English Corpus Linguistics). LOCNESS. Louvain-la-Neuve: Universite catholique de Louvain. Available from https://www.uclouvain.be/en-cecl-locness.html

Chen, Y. & P. Baker (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology 14(2)*: 30–49.

Fletcher, W. (2002–2007). KfNgram. Annapolis: USNA. <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html> (20 November, 2011)

Fletcher, W. (2010). Phrases in English. <http://phrasesinenglish.org/> (20 September, 2014)

Forsyth, R. & Grabowski, Ł. (2015). Is there a formula for formulaic language? *Poznań Studies in Contemporary Linguistics 51(4)*: 511–549.

Fuster-Marquez, M. (2014). Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction 7(1)*: 84–121.

Grabowski, Ł. (in print). Phrase frames in English pharmaceutical discourse: a corpus-driven study of intra-disciplinary register variation. *Research in Language*.

Granger, S., E. Dagneaux, F. Meunier and M. Paquot (eds). (2009). *International Corpus of Learner English.* Version 2. Louvain-la-Neuve: Presses universitaires de Louvain.

Gray, B. & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics 18(1)*: 109–135.

Juknevičienė, R. (2013). Recurrent word sequences in written learner English. In I. Šeškauskienė and J. Grigaliūnienė (eds) *Anglistics in Lithuania. Cross-Linguistic and*

Cross-Cultural Aspects of Study. Newcastle upon Tyne: Cambridge Scholars Publishing. 178–197.

Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics 18(3)*: 391–417.

Paquot, M. (2014). Cross-linguistic influence and formulaic language: Recurrent word sequences in French learner writing. *EUROSLA Yearbook 14*: 240–261. Amsterdam: John Benjamins.

Römer, Ute. (2009). English in academia: Does nativeness matter? *Anglistik: International Journal of English Studies 20(2)*: 89–100.

Scott, M. (2008). *Wordsmith Tools*. Version 5. Oxford: Oxford University Press.

Mariusz Kamiński (College of Management "Edukacja" Wrocław)

A quantitative study of dictionary definitions

This paper compares dictionary definitions using two quantitative methods: hierarchical cluster analysis and correspondence analysis. The goal of the study is to show the practical utility of these methods in lexicography research. The research was conducted on definitions in LDOCE (2005) and Webster's Collegiate (2003), that is the dictionaries intended for different audiences. A random sample of definitions was drawn from each dictionary and converted into three-word clusters (trigrams). The data were processed in such a way as to generate frequency lists ordered according to the most frequent items in both dictionaries. Cluster analysis conducted on these data confirmed that the dictionaries differ in terms of frequency distribution of trigrams. Performing correspondence analysis, it was possible to visualise the data, and to see which word clusters are characteristic of each dictionary. As could be expected, there are substantial differences in lexical composition of the definitions between the dictionaries, which manifests itself in a different set of most frequent trigrams. This study shows that using the quantitative methods, it is possible to discover certain stylistic trends in the editors' policy towards definitions.

Jolanta Kovalevskaitė (Vytautas Magnus University)

Loïc Boizou (Vytautas Magnus University)

Erika Rimkutė (Vytautas Magnus University)

Morphological realisations of two-word syntactic relations in the *Dictionary of the Lithuanian nominal phrases*

This contribution to the analysis of Lithuanian multi-word expressions (MWEs) aims to describe the relation between the morphological and syntactic features of the morphologically variable two-word expressions extracted from the corpus-based *Dictionary of the Lithuanian Nominal Phrases* (Rimkutė et al. 2012). First insights on the quantitative description of Lithuanian MWEs which take into account the relations between morphological and syntactic features were discussed by Boizou et al. (2015). The following possible variability patterns of two-word expressions have been distinguished: fixed + fixed, fixed + variable, variable + fixed, variable + variable. Analysed MWEs have shown that models variable + variable and fixed + variable are the most frequent (respectively 46% and 44% of MWEs).

In relation to the previously mentioned research on variability patterns, our study aims to establish the most typical links between the morphological features and syntactic functions. Apart from the different verbal categories (tense, mood and person, various gerund and participle forms), the main emphasis will be put on the case category, in order to define the syntactic realisations of each case. For example, the genitive is one of the most frequent cases in the following syntactic relations: a) in subordinate attributive phrases with the pattern fixed + variable, where the genitive noun (as the fixed constituent) is the typical non-agreeing attribute (universiteto-GEN.SG *darbuotojas* "university worker"); b) in subordinate object phrases fixed + variable, where the genitive case (of the variable constituent) is related with verbal government (*laukti vasaros*-GEN.SG "to wait for the summer"). This research will provide quantitative results that will be compared with the data from contemporary Lithuanian.

References

Boizou L., Kovalevskaitė J., Rimkutė E. (2015): Lietuvių kalbos dvižodžių junginių morfologinių ir sintaksinių ypatybių sąsajos. *Darbai ir dienos* 63 (in print).

Rimkutė E., Bielinskienė A., Kovalevskaitė J. (eds) (2012): *Lietuvių kalbos daiktavardinių frazių žodynas*. Kaunas: VDU leidykla, http://donelaitis.vdu.lt/lkk/pdf/daikt_fr.pdf.

Yuriy Kovalyuk (Yuriy Fedkovych Chernivtsi National University)

Finding the common ground: phraseological theory and terminology revisited

Against all odds and claims that "phraseology as a distinct field has never really taken hold in Anglo-American linguistics" (Norrick 2007: 615), it has now become a "major field of pure and applied research" (Cowie 1998: 1) in Western linguistics. This process is, however, challenged by two factors: confusing and ample terminology (Cowie 1998: 1, Moon 1998: 1, Granger & Paquot 2008: 27, Naciscione 2010: 17) and different approaches regarding its scope. To avoid terminological puzzles, scholars have opted for inclusive terms, such as phraseme (Mel'čuk 1988), phraseological unit (Cowie 1998), fixed expression (Moon 1998), prefabricated sequence (Wray 2008), etc. and have adopted either a theory driven (Eastern tradition) or a data driven (Western tradition) approach. Yet, as Granger & Paquot maintain, both frameworks will best work in combination, i.e. "corpus-based methods should be incorporated into mainstream approaches" (2008: 45). Therefore, the present paper will first focus on each of the approaches in isolation. It will then examine the links between phraseology and its foregrounding disciplines: grammar, semantics, pragmatics, discourse and enthnolinguistics to account for the apparent conflicting doctrines in the field. Finally, a survey into phraseological terminology will be conducted and a classification of phraseological units based on the reconciliation of the purely linguistic and the data driven approaches will be suggested.

References

Cowie, A. P. (1998). Introduction. In A. P. Cowie (ed.). *Phraseology: Theory, Analysis, and Applications*. New York: OUP. 1–23.

Granger, S., & Paquot, M. (2008). In S. Granger and F Meunier (eds) *Phraseology: An Interdisciplinary Perspective*). Amsterdam & Philadelphia: John Benjamins. 27–51.

Mel'čuk, I. A. (1988), Semantic description of lexical units in an explanatory combinatorial dictionary: basic principles and heuristic criteria. *International Journal of Lexicography 1(3)*: 165–88.

Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Clarendon Press.

Naciscione, A. (2010). *Stylistic Use of Phraseological Units in Discourse*. Amsterdam & Philadelphia: John Benjamins.

Norrick, N. (2007). English phraseology. In H. Burger, D. Dobrovol'skij, P. Kühn & N. Norrick (eds) *Phraseology: An International Handbook of Contemporary Research* (Vol. 2). Berlin & New York: Walter de Gruyter. 615–619.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. New York: OUP.

Benjamin Kremmel (University of Nottingham, University of Innsbruck)

Tineke Brunfaut (Lancaster University)

Charles Alderson (Lancaster University)

How important is knowledge of multi-word expressions for reading comprehension?

Several studies have contrasted the contribution of FL vocabulary and syntactic knowledge to reading performance (Shiotsu & Weir, 2007; van Gelderen et al., 2003, 2004). Despite the increasingly recognized formulaic nature of language, the majority of these studies have followed a dichotomous view of these components, focusing on individual words as constituting vocabulary and disregarding formulaic sequences completely. As a result, the contribution made by phraseological knowledge to reading ability has not been investigated systematically (Martinez & Murphy, 2011).

This paper reports on a study that examined the impact of including a phraseological component in explaining variance in reading performances. Test scores of 418 learners of English as a foreign language (EFL) were modeled in a structural equation model, showing that a measure of multi-word expression knowledge outperformed traditional syntactic and vocabulary measures in predicting reading comprehension variance. Additional insights into the role of knowledge of multi-word expressions were gained through verbal protocol analysis of 15 EFL learners answering reading comprehension items that targeted

the understanding of such phrasal expressions within written context. The findings hint at an underestimated, but critical, role of phraseological knowledge in FL reading, and are relevant to both the assessment and the teaching of EFL reading ability.

References

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly 45(2)*: 267–290.

Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing 24(1)*: 99–128.

van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). Roles of linguistic knowledge, metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension: a structural equation modelling approach. *International Journal of Bilingualism 7(1)*: 7–25.

van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed and metacognitive knowledge in first and second language reading comprehension: a componential analysis. *Journal of Educational Psychology 96(1)*: 19–30.

Batia Laufer (University of Haifa)

Liubov Baladzhaeva (University of Haifa)

Attrition of collocations in the native language of bilingual and monolingual immigrants.

Most research on collocation problems has focused on use and learning, particularly in L2 (e.g. Alterberg & Granger, 2001; Hasselgren, 1994; Howarth, 1996; Kaszubski, 2000; Nesselhauf, 2005; Peters, 2015; Webb & Kagimoto, 2011). Very few studies investigated the attrition of L1 collocations in the context of L2 learning (Isurin, 2007; Laufer, 2003).

We report on two studies in which we investigate whether Russian immigrants in Israel, with and without knowledge of Hebrew (L2), experience attrition of collocations in Russian (L1). In both studies we asked the participants to judge correctness of collocations, and in the second study the participants were also

required to indicate how confident they were about their answers. We compared three groups of adult participants: immigrants with knowledge of Hebrew (n=37 in study 1 and 44 in study 2), immigrants without knowledge of Hebrew (n= 16 and 30) and monolingual controls, residents of Russia and Kazakhstan (n= 21 and 21). In the second study the participants were matched on age and education. Sociolinguistics variables of the participants were collected as well.

We used ANOVAs and post-hoc tests for the comparison of the three groups in each study, and performed correlations between the correction judgement scores and each one of the sociolinguistic variables. Participants with no knowledge of Hebrew were found to perform no differently than participants with knowledge of Hebrew, and both performed significantly worse than the controls. There was no correlation between attrition and factors known to affect attrition, like age of arrival in Israel, length of residence and level of education. In the second study, test results correlated positively with the amount of usage of Russian. We explain the results in terms of direct and indirect L2 influence and discuss the notion of 'second hand attrition'.

Agnieszka Leńko-Szymańska (University of Warsaw)

A corpus-based analysis of the development of phraseological competence in EFL learners using the CollGram profile

One way to study the development of phraseological competence in L2 learners is to explore learner corpora. The methodology so far applied by corpus-based studies of L2 phraseology have consisted in comparing frequencies of n-grams retrieved from learner data at two or more proficiency levels. These studies do not analyse the use of phraseology in individual learner texts, but examine recurrent expressions holistically in a corpus. Such an approach ignores individual variation in the use of multi-word units. In addition, it does not consider the strength of co-occurrence of the retrieved items.

A new method of studying phraseology was recently proposed by Bestgen & Granger (2014). CollGram is "a technique that assigns to each pair of contiguous words (bigrams) in a learner text two association scores (…) computed on the

basis of a large reference corpus". It produces three measures – mutual information, t-score and the number of idiosyncratic units, which together form a CollGram profile, and which, according to the authors, "quantify the collocation strength of each text" (p. 31). So far CollGram has been applied to trace the development of phraseological competence in intermediate and advanced learners (Granger & Bestgen 2014; Bestgen & Granger 2014).

The purpose of this study is to analyse learner data from a wider range of proficiency levels (pre-A1 – C1). The EFL texts analysed in the study were drawn from two corpora: the International Corpus of Crosslinguistic Interlanguage (Tono et al. 2012) and the International Corpus of Learner English (Granger et al. 2002). 120 essays written in English by L1 German, Polish and Spanish were rated on the CEFR scale (Council of Europe 2001) by three raters. Next, a CollGram profile was computed for each essay using COCA as a reference corpus. Changes in the values of the CollGram measures for six CEFR proficiency levels were analysed using ANOVA.

The presentations will discuss the results of the analyses, which reveal a complex pattern of growth in the use of native phraseology by EFL learners at different proficiency levels. They also demonstrate that the CollGram profile is a valuable instrument in tracing the development of phraseological competence in EFL learners.

References

Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing 26*: 28–41.

Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching 52(3)*: 229–252.

Granger, S., Dagneaux, E. Maunier, F. & Paquot, M. (2009). *International Corpus of Learner English* (Version 2). Presses universitaires de Louvain.

Tono, Y., Kawaguchi, Y. & Minegishi, M. (eds). (2012). *Developmental and Cross-linguistic Perspectives in Learner Corpus Research*. Amsterdam and Philadelphia, PA: John Benjamins.

Saskia E. Lensink (Leiden University)

Niels O. Schiller (Leiden University)

Arie Verhagen (Leiden University)

Parts and wholes - on how to empirically test for the cognitive reality of multi-word units

According to usage-based theories of language, the way we use language shapes the way we store it in our mental lexicons (Bybee, 2006). This view would imply that completely regular, frequent combinations of words could be stored as wholes. Indeed, there is a growing body of psycho- and neurolinguistic studies that finds evidence for the existence of multi-word units (Arnon & Snider, 2010; Tremblay & Baayen, 2010; Tremblay & Tucker, 2011). Moreover, Siyanova-Chanturia, Conklin & Van Heuven (2011) found phrasal frequency effects in both L1 and L2 speakers for a specific phrasal pattern, which could suggest that speakers also make use of multi-word units in their L2. Recently we replicated and extended these findings for a wide range of multi-word units in Dutch and have found further evidence that L1 speakers and advanced L2 speakers make use of multi-word units in their language production.

In a series of experiments we had Dutch participants read out loud frequent trigrams and quadgrams in their L1 or L2, English. Using the newest mixed-effects modelling techniques (Wood, 2006), it is shown that when speaking in their native language, participants are sensitive to both the frequencies of the single words and the frequency of the whole multi-word unit itself. These findings on production data are reminiscent of predictions from dual route models where multiple processing routes are hypothesized to exist and can be employed at the same time in comprehension (Baayen, Dijkstra & Schreuder, 1997). However, when producing L2 quadgrams, it seems like only the quadgram frequencies contribute significantly to the speed of talking.

The talk will focus on the different techniques one can employ to study the cognitive reality of multi-word units, discussing the findings from our lab on both L1 and L2 speakers and the implications of these findings for linguistic theory.

References

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language 62*: 67–82.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language 37(1)*: 94–117.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language,* 711–733.

Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. Jou*rnal of Experimental Psychology: Learning, Memory, and Cognition 37(3)*: 776.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication*, 151–173.

Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon 6(2)*: 302–324.

Wood, S. N. (2006). *Generalized additive models: an introduction* with R. CRC press.

Hiroyuki Matsumoto (Hokkai Gakuen University)

Neil Heffernan (Kureme University)

Can utilization of lexical units facilitate EFL reading proficiency?

Lexical units, or formulaic language, in language chunks have been studied primarily in memory and the production of language. In most research, they are helpful in developing native-like fluency and word selection (Hunston & Francis, 2000; Lewis, 1993; Nattinger & DeCarrico, 1992). In terms of language comprehension, however, little has been found except that the mental act of separating a sentence into chunks can improve reading comprehension (Eskey & Grabe, 1988; Lewis, 1993). In this study, an experiment was conducted to examine whether utilization of lexical units actually has a facilitating and positive effect on English-as-a-foreign-language (EFL) reading comprehension, on the basis of strategy instruction: (1) a meaning strategy with a focus on the connection

of key words, (2) a logic strategy with a focus on discourse markers, (3) a schema strategy that helps activate prior knowledge, and (4) a lexical-unit strategy that utilizes both lexical cohesion and pragmatic functions. The participants were 150 tenth- and eleventh-grade students with limited English proficiency in Japan. The treatments were arranged for three experimental groups and one control group: X1 (all the strategies), X2 (except for the lexical-unit strategy), X3 (the lexical-unit strategy), and X4 (no strategy). Based on data from three reading proficiency tests with multiple-choice questions, a chunking test by separating sentences into meaningful chunks, and a strategy questionnaire on a five-point Likert scale, structural equation modelling (SEM) was constructed to evaluate the intensity of causality between the utilization of lexical units and EFL reading proficiency. The results indicate the possibility that utilization of lexical units had both facilitating and synergetic effects on the participants' EFL reading proficiency. The synergetic effect of utilizing lexical units was recognized because the X1 experimental group demonstrated the highest causality, which implies the importance of clarifying semantic and logical connection in successful EFL reading.

References

Eskey, D. E. & Grabe, W. (1988). Interactive models for second language reading: Perspectives on instruction. In P. L. Carrell, J. Devine, & D. Eskey (eds) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press. 223–238.

Hunston, S., & Francis, G. (2000). *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins Publishing Company.

Lewis, M. (1993). *The Lexical Approach: The state of ELT and a way forward*. Hove: Language Teaching Publications.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford University Press.

Natsumi Okuwaki (Tsuru University)

The processing of spoken and written formulaic sequences by L2 speakers

Formulaic sequences (FS) have been found to present a processing advantage compared to non-formulaic equivalents by L1 speakers (e.g. Gibbs et al., 1997) and sometimes by L2 speakers (e.g. Conklin & Schmitt, 2008). This is either because FS is processed as unanalysed holistic units (Wray, 2002) or because they are highly automatised (Cordier, 2013). It may also be due to the phrasal frequency (Siyanova-Chanturia, 2015). Whatever implications it may have, faster processing is considered advantageous for L2 speakers.

It is not clear, however, whether faster processing is observed for different types of FS. Qian (2015) compared processing on three types of FS and showed that frequent written sequences were processed faster than infrequent idioms. Based on Qian (2015), the present study further investigates the processing of spoken and written formulae to see if 1) FS are processed more quickly than matched novel phrases, 2) processing speed is different among types, and 3) phrasal frequency has any effect on processing.

30 Japanese L1-speakers learning English participated in the study. 45 target items, listed into 3000- and 4000-word bands, were taken from the PHRASE List (Martinez & Schmitt, 2012). The items were consisted of three types of English FS (15 for each type); spoken FS, written academic FS, and FS used in both genres.

For each FS, a non-formulaic equivalent is created (45 matched phrases), and 45 ungrammatical sequences were used as distractors. In addition, 15 higher-frequency FS, listed into 1000- and 2000-word bands, were chosen to compare the processing speed with the lower-frequency FS.

The participants were required to judge the grammaticality of the items by pressing a key. Their reaction times and error rates were collected.

Findings of this study will be discussed in terms of processing advantages and the presence/absence of frequency effect on the processing of FS by L2 speakers.

Carlos Prado-Alonso (University of Oviedo)

On the register variation of idiomatic *Do so*

Do so constructions, as in '*I bought a car yesterday, and Peter has done so today'*, are verbal anaphors that have been considered fairly formulaic syntactic constructions, as they contain structurally unique features which always take the same form, and acquire a meaning that is not strictly predictable from the individual parts.

Do so constructions have received extensive attention from a theoretical perspective but very little research has dealt with the analysis of these constructions in naturally occurring discourse. Research has focused mainly on the categorical factors —i.e. semantic and syntactic— that determine the use of the construction. It has been argued, for instance, that the extent of application of *do so* anaphora depends principally on factors such as: (a) non-stativity of the antecedent (Guimier 1981); (b) antecedent not headed by be (Levin 1986); (c) coreferentiality of subjects in the antecedent and *do so* clauses (Souesme 1987); (d) adjunct status of any "orphan" in the *do so* clause (Culicover and Jackendoff 2005); (e) non-contrastive status of any adjunct in the *do so* clause (Stirling and Huddleston 2002); (f) antecedent embedded rather than matrix predicate (Levin 1986); (g) voice or category differences between antecedent and the *do so* clause (Stirling and Huddleston 2002); and (h) adverse connotations of the antecedent (Bolinger 1970).

By contrast, this study provides an analysis of *do so* anaphora in naturally occurring discourse providing an in-depth investigation of the textual factors affecting the distribution and pragmatic use of these types of idiomatic and formulaic expressions in Present-day English written texts. The data for the study have been taken from six computerised corpora of British and American Present-day English, namely the LOB, FLOB, FLOB, FROWN, BE06, and AmE06 corpora, comprising texts from the 1960s, 1990s and 2000s (for details see Hofland et al. 1999 and Baker 2009).

In sum, the analysis sheds light on the linguistic and textual factors that drive the pragmatic use and the distribution of *do so* verbal anaphora and shows that, in addition to syntactic and semantic factors, the linguistic features of the texts in

which they occur also play an important role in the use of these types of formulaic expressions.

References

Baker, P. (2009). The BE06 Corpus of British English and Recent Language Change. *International Journal of Corpus Linguistics 14(3)*: 312–337.

Bolinger, Dwight. (1970). The meaning of do so. *Linguistic Inquiry 1*: 140–144.

Culicover, P.W., Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Hofland, K., A. Lindebjerg and J. Thunestvedt. (1999). *ICAME Collection of English Language Corpora*. 2nd edition, CD-ROM version. Bergen: The HIT Centre.

Guimier, C. (1981). Sur la Substitution Verbale en Anglais. *Modèles Linguistiques 3(1)*: 135–161.

Levin, L. (1986). *Operations on Lexical Forms: Unaccusative Rules in Germanic Languages*. PhD dissertation. Cambridge, MA, MIT.

Souesme, J. (1987). Valeurs et Emplois Respectifs de DO et DO SO. *Modèles Linguistiques 9*: 65–92.

Stirling, L & R. Huddleston (2002). Deixis and Anaphora. In R. Huddleston & G. Pullum (eds) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University. 1449–1564.

Geraint Paul Rees (Universitat Pompeu Fabra)

The selection of EAP vocabulary: A formulaic solution

There are a number of flaws in current approaches to the selection of vocabulary for English for Academic Purposes (EAP) courses and materials. Firstly, current single-item wordlist approaches are inadequate since they do not account for the influence of colligational context in conditioning word meaning (Hyland & Tse, 2009). Secondly, those approaches which do take into account multi-word strings rely on frequency of occurrence alone, to the detriment of semantic concerns (Biber et al., 2004; Simpson-Vlach & Ellis, 2010). Finally, on both these approaches academic vocabulary is regarded as monolithic when, in fact, there is a great deal of variation in the lexis used between academic disciplines.

In an attempt to address these problems I present a corpus-based study of verb collocation in academic journal articles which employs an innovative two-stage methodology for the selection of vocabulary for EAP. First, the frequency of

occurrence of a sample of verbs is compared across three sub-corpora, each consisting of journal articles from a different academic discipline. Then, the semantic behavior of verbs in each discipline is compared employing techniques from Corpus Pattern Analysis (Hanks, 2004). The study provides evidence for discipline-specific formulaic sequences, and the need to take into account the semantics of verbs and verb collocates when selecting vocabulary for EAP.

This raises the question of how formulaic sequences should be addressed in EAP materials and lexicography in general. I present some preliminary proposals in this regard. Given that EAP plays an increasingly important role in governing students' economic and academic opportunities, this research has significant implications not only for classroom practice but also for materials and assessment design, and the publishing industry.

References

Biber, D., Conrad, S. M., Reppen, Byrd, R. P., Helt, M., Clark, V., Urzua, A. (2004). *Representing Language Use in the University: Analysis of the TOEFL® 2000 Spoken and Written Academic Language Corpus. TOEFL Monographs*. Princeton: Educational Testing Service.

Hanks, P. (2004). Corpus pattern analysis. In G. Williams & S. Vessier (eds) *Proceedings of the Eleventh EURALEX International Congress*. Universite de Bretagne-Sud, Lorient: EURALEX.

Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: corpus evidence for specificity. *International Journal of English Studies 9*: 111–130. Retrieved from http://dialnet.unirioja.es/servlet/articulo?codigo=3104270

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics 31*: 487–512.

Teresė Ringailienė (Vytautas Magnus University)

Collocations in popular scientific and academic discourse in English and Lithuanian

Popular scientific discourse still takes a sufficiently ambiguous position: some popular scientific magazines, both printed and online, have a wide readership; however, the research on this sub-genre has still been fragmentary and

inexhaustive. For instance, da Silva (2010) focuses on verbal and mental processes in science popularization news; Mussolf (2007) explores the use of metaphors; Calsamiglia and van Dijk (2004) discuss popularization discourse in the Spanish press; Petrėnienė (2013, 2010, 2003) analyses such features as the use of comparisons, metaphors, and the explanation of terms in science populiarization.

The aim of the present research is twofold: first, it explores the differences in the collocations used in popular scientific and academic discourse; second, it compares the collocations used in English and Lithuanian popular scientific discourse. For the analysis, the Corpus of Popular Scientific Articles (COPSA) is used, which is composed of an English and a Lithuanian sub-corpus, each containing 500,000 words. For the comparison with academic discourse, the Corpus of Contemporary American English (COCA) and Dabartinės lietuvių kalbos tekstynas were used. The research is limited to the analysis of the collocations related to the academia.

The present research has revealed that there are a number of differences between the collocations used in popular scientific and academic discourse as well as between English and Lithuanian. First, research methods and even researchers themselves tend not to be focused on in popular scientific discourse, contrary to academic texts where the description of data and references to scholars play a major role. Second, the collocations with the collocate scholar (*mokslininkas/ė*) and university (*universitetas*) emphasise the tendency towards international research in English and local research in Lithuanian. Third, different collocational patterns suggest not only differences between popular scientific and academic discourse, but also a different conceptualization of certain phenomena in English and Lithuanian.

Jūratė Ruzaitė (Vytautas Magnus University)

Formulaic vagueness markers in spoken Lithuanian

This study is a corpus-based analysis of one type of vagueness markers, i.e. general extenders, in spoken Lithuanian. This analysis is concerned with vague multiword expressions encoding shared knowledge and most commonly referred to as

general extenders (GEs) (Overstreet 1999), e.g. *ir panašiai* (='and similar') and *ir viską* (='and everything'). Other terms for these formulaic expressions include 'set marking tags', 'utterance final tags', 'clause terminal tags', 'extension particles', 'generalized list completers', 'generalizers', and 'final coordination tags'.

The main body of research on GEs focuses on English L1 thus producing English-dominated models of analysis and interpretation. Cross-linguistic comparisons and studies of lesser-researched languages are still scarce (e.g. Polish (Wierzbicka 1991), Montreal French (Dubois 1992), Swedish (Winter and Norby 2000), German (Overstreet 2005), Persian (Parvaresh & Tayebi 2014), and Spanish (Fernández 2015)).

This paper presents findings from the initial stages of an ongoing corpus-based study into the forms and functions of vague language in face-to-face conversations in Lithuanian. The corpus used for the present investigation is a subcorpus of The Corpus of the Contemporary Lithuanian Language (CCLL), which consists of 557,822 words of naturally occurring spoken discourse among speakers of Lithuanian.

The findings indicate, firstly, that the structure and length of GEs as well as the length of the exemplars they modify allow for a great degree of variation. Similarly to some other languages, in spoken Lithuanian adjunctive GEs are more frequent than disjunctive GEs. The analysis of co-text has revealed that in spoken interaction GEs tend to cluster with non-standard language use, colloquialisms, different hedging devices, and hesitation markers. With regard to social functions, GEs are predominantly used to soften utterances and to mark approximation, to save face, to build intersubjectivity, to indicate the extent of commitment, and to mark in-group membership; some types of GEs are employed for emphasis.

References

Channell, J. (1994) *Vague Language*. Oxford: Oxford University Press.

Dubois, S. (1992) Extension particles, etc. *Language Variation and Change 4*: 179–203.

Fernández, J. (2015) General extender use in spoken Peninsular Spanish: metapragmatic awareness and pedagogical implications. *Journal of Spanish Language Teaching 2(1)*: 1–17.

Overstreet, M. (2005) And stuff und so: Investigating pragmatic expressions in English and German. *Journal of Pragmatics 37(11)*: 1845–1864.

Overstreet, M. (1999) *Whales, Candlelight and Stuff Like That: General Extenders in English Discourse*. New York/Oxford: Oxford University Press.

Parvaresh, V. & T. Tayebi (2014) Vaguely speaking in Persian. *Discourse Processes 51(7)*: 565–600.

Wierzbicka, A. (1991) *Cross-cultural Pragmatics: The Semantics of Human Interaction*. Berlin: de Gruyter.

Winter, J. & Norrby, C. (2000) Set marking tags 'and stuff'. *Proceedings of the 1999 Conference of the Australian Linguistic Society*. http://www.als.asn.au/proceedings/als1999/winter&norrby.pdf

Birutė Ryvitytė (Vilnius University)

Evaluative meaning and disciplinary values: A corpus-based study of the adjective *svarbus* important in the Corpus Academicum Lithuanicum

The interface between academic values and the evaluative function of discourse has attracted considerable attention over the last two decades (Giannoni 2010, Hunston 2011, etc.). The purpose of this presentation is to investigate disciplinary values in academic writing, focusing on the use of adjective + noun patterns, in particular on the use of adjective *svarbus* 'important'. The analysis proceeds through a detailed quantitative and qualitative analysis of these patterns in the Corpus of Academic Lithuanian (CorALit: http://coralit.lt/) comprising authentic texts representing the main fields of study and research developed in Lithuania. The analysis focuses not only on the adjective + noun pattern, but also on collocations within and around the pattern. The empirical focus of the study will be on the 'hard' and 'soft' disciplines as represented by the corpus of 9 million words. Comparison of the results for the corpus shows that there are differences in form and meaning between the disciplines. The features revealed by the analysis are indicative of the epistemological characteristics of these disciplinary discourses. It is argued that these differences reflect the different academic writing cultures and norms of 'hard' and 'soft' sciences.

References

Giannoni, D.S. (2010). *Mapping Academic Values in the Disciplines: A Corpus-based Approach*. Berlin: Peter Lang.

Hunston, S. (2011). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London: Taylor & Francis.

Svetlana Schukina (Belarusian State Academy of Music)

Formulaicity in English musical vocabulary

This work originates from the author's experience as a second language teacher at the Belarusian State Academy of Music. It deals with the English language for music academic purposes and international relations.

The primary aim of the paper is to increase the awareness of the problem of formulaic speech, translating and writing, in order to improve the conversational fluency and written abilities in musicians' international professional communication. As it is, our aim goes well with the importance of the three main functions of formulaic language, or "formulaic sequence" (A. Wray), as the most important ones, i.e. – communication, production and learning strategies.

The musical vocabulary incorporates musical terminology; terminological semantic adaptations; expression markings – instructions added to the music (by a composer or an editor) to indicate the manner of performance – dynamic, tempo and mood markings; words with no direct equivalents in other languages; musical lexis of no terminological profile; words of the common language.

The paper considers the identification of formulaic sequences in each of the above groups, and gives evidential examples of the availability of clichés, collocations, hackneyed phrases and idioms in all the groups, though in different degrees. Being stipulated by specific qualities of the musical term, they are traced even in terminology, despite the well-known assertion about the standardization, which is a key issue therein and involves a statement about "a central difference between terminology and formulaic language…", i.e. "…the former refers to a standard whereas the latter refers to the mental lexicon"(P. ten Hacken & F. Parra).

The article also provides the coverage of such topics of the musical vocabulary as second-language acquisition, teaching, finding prefabricated phrases of high frequency to be used at master-classes of internationally renowned musicians;

when compiling applications to organizing committees of international musical competitions, etc., and outlines perspectives in the domain.

References

Anichkov, I. (1997). *Works on Linguistics*. St Petersburg: Nauka.

Hacken, P. & Fernandez-Parra, M. (2008). Terminology and Formulaic Language in Computer-Assisted Translation. *SKASE Journal of Translation and Interpretation 3(1)*: 1–16.

Fernandez-Parra, M. (2008). Translating Formulaic Expressions in Instruction Manuals: A Corpus Study. *Newcastle Working Papers in Linguistics 14:* 51–60.

Maslyko, E.A., (1990). *Communicating while teaching/learning English*. Minsk: Vysheishaya shkola.

Wood, D. (ed.) (2010). *Perspectives on Formulaic Language: Acquisition and Communication*, London/New York: Continuum.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: CUP.

Wray, A. & Perkins M.R. (2000). The functions of formulaic language: an integrated model. *Language & Communication 20*: 1–28.

Inesa Šeškauskienė (Vilnius University)

Justina Urbonaitė (Vilnius University)

From freedom to liberty to custodial sentence in the Criminal Code of the Republic of Lithuania, or some strategies of translating Lithuanian collocations with *laisvė* 'freedom' into English

Notably, legal language abounds in formulaic expressions, which include field-specific terms and longer units. Formulaicity here refers to understanding words and word strings as wholes (Wray 2002). However, the degree of formulaicity may vary. Phrases like on the contrary are fully fixed in form. Some other combinations, especially those involving nouns and verbs, are semi-preconstructed (for the term, see Wray 2002: 7), i.e. allowing more collocational freedom. For example, in legal English we could use to pass a law or to adopt a law, but *to accept a law is unacceptable. These collocations are often language-specific, sometimes also register- and genre-specific.

The present investigation focuses on collocations with the Lithuanian word *laisvė* 'freedom, liberty', one of the key words in legal language, and their rendering into English in the Criminal Code of the Republic of Lithuania. The research aims at 1) interpreting those collocations from the semantic point of view, relying on the Conceptual Metaphor Theory and its later developments (Lakoff and Johnson 1980/2003; Winter 2001; Deignan 2005), 2) identifying their equivalents in English and 3) identifying a strategy of translation: a) keeping the metaphor in both the source text (ST) and the target text (TT) or b) keeping it in the ST and discarding it in the TT (Abdulah and Shuttleworth 2013: 613–614). The methodology of research involved the AntConc programme (2014), employed to search for the word in question and its immediate co-text in both languages, the Metaphor Identification Procedure (Steen et al. 2010) and metaphorical pattern analysis (Stefanovitsch 2004, 2006).

The results suggest that collocational patterns with the word *laisvė* are very stable. They are all interpretable as metaphors but not always rendered as such in English. Those cases are analysed in more detail. Translation strategies are determined by culture-specific conceptualization, language resources and the legislator's intention.

References

Abdulah, S. and M. Shuttleworth (2013). Metaphors in the translation of English technical texts into Malay: a preliminary study. *Journal of Asian Scientific Research 3(6)*: 608–629.

Anthony, L. (2014). *AntConc* (Version 3.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/. Accessed March 2014.

Deignan, A. (2005). *Metaphor and Corpus Linguistics*. Amsterdam: John Benjamins Publishing.

Lakoff, G. and M. Johnson. (1980/2003). *Metaphors We Live By*. Chicago: University of Chicago Press.

Steen, G.J., A.G. Dorst, J.B. Herrmann, A.A. Kaal, and T. Krennmayr (2010). Metaphor in usage. *Cognitive Linguistics 21(4)*: 765–796.

Stefanowitsch, A. (2004). Happiness in English and German: a metaphorical pattern analysis. In: Achard, M. and S. Kemmer (eds) *Language, Culture, and Mind*. Stanford, CA: CSLI Publications. 137–149.

Stefanowitsch, A. (2006). Words and their metaphors. In: Stefanowitsch, A. and S.Th. Gries (eds) *Corpus-Based Approaches to Metaphor and Metonymy*. 63–105.

Winter, S. L. (2001). *A Clearing in the Forest. Law, Life and Mind*. Chicago: the University of Chicago Press.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: CUP.

John J. Sidtis (Nathan Kline Institute/New York University)

Diana Van Lancker Sidtis (New York University/Nathan Kline Institute)

Formulaic expressions may have a special home in the brain: Evidence from functional imaging

Clinical observations in the 19th and early 20th centuries on the effects of brain damage on language significantly influenced modern concepts of language organization. In the past several decades, brain and language studies have become the domain of functional imaging. In spite of advanced technology and signal processing, imaging results have typically been discordant with clinical observations. In contrast with the practice of decomposing brain images based on speculation about the structure of a complex behavior, we developed an approach that identifies brain patterns that predict an individual's behavior during the scan, a performance-based analysis. A reproducible relationship that predicts speech rate has been identified: as speech rate increases, blood flow increases in Broca's area (the left inferior frontal region) and decreases in the right caudate nucleus (part of the basal ganglia).

For this study, a group of 16 normal, right-handed, native speakers of American English were scanned during spontaneous productions of 60 sec monologues. Monologues were transcribed and the number of words in pause fillers, sentence stems, proverbs and idioms, and speech formulas were determined and converted to the percentages of the total number of words in formulaic expressions. Surprisingly, performance-based analyses revealed that the percentage of pause fillers in formulaic expressions was predicted by the same brain regions that predict syllable rates during repetition. In contrast, the percentage of words in speech formulas was predicted by the opposite pattern: an increase in the right inferior frontal region coupled with a decrease in the left caudate. These results are consistent with clinical observations that implicate the right hemisphere and basal ganglia in formulaic expressions and further suggest that not all formulaic

expressions have the same brain organization. When properly analyzed incorporating clinical knowledge and sophisticated linguistic approaches, functional imaging may contribute to a better conceptualization of language organization.

Narcisse Torshizi (McMaster University)

Magda Stroińska (McMaster University)

Mechanism of transition from patriotism to nationalism: Some insights from the views of homeland in second generation Persians in Canada

In Maslov's (1943) hierarchy, the need to be loved and accepted (to belong) is preceded only by physiological requirements and by the need for safety (Baumeister & Leary, 1995, DeWall et al, 2011 and Kune, 2011). This desire to belong concerns not only family and friends, but also a greater community, a tribe or a nation. One usually realizes the importance of this type of belonging when one chooses or is forced into exile, i.e. when one leaves the space inhabited by one's community. While the need to be part of a homeland community is both basic and natural, it can be transformed into the need to assert the superiority of one's community and the need to protect it from alleged foreign influences. Modern day politicians regularly manipulate the attachment to one's homeland and feelings of "true patriot love" to transform them into nationalism and xenophobia.

According to Wierzbicka (1997: 156), words related to homeland "often differ from one another in particularly telling ways, offering valuable insight into different national traditions and historical experiences." Homeland is "a physical place with cultural associations" (Liebler, 2010: 596) and its emotional effects on individuals could be an important factor in the integration of immigrants in multicultural countries like Canada. This paper analyses terms that are used to describe one's native land in Farsi. Based on interviews with ten 2nd generation Persian immigrants to Canada and using Wierzbicka's semantic primitives approach, we look at four terms used (*vatan, mihan, sarzamine madari* and *sarzamine niyakan*) to see how the relationship to the physical space considered one's homeland, reflected by language choices, impacts the construction of individual and group identity. This homeland is perceived as a product of the common

language, shared land and culture; "it implies a community of people tied with emotional as well as geographical bond" (Vamvakidou et al., 2010: 545-46). In particular, we analyse which terms refer to the physical bond with the country of one's ancestry and which point to a concept of mythical national space with no physical dimensions. It is the latter that is often evoked in nationalistic discourse, in particular in terms of defending the safety and purity of one's national culture and national space (Farhamy 2007). We analyse linguistic means of establishing the distinction between patriotism and nationalism in the discourse of 2nd generation Persian immigrants but we believe that the results may be applicable to other national and linguistic groups. This topic seems particularly relevant in the world where the notion of nation-states is at odds with the needs of refugees seeking a new homeland but still carrying with them the feeling of belonging to a different national space?

References

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin 117(3)*: 497–529.

DeWall, C., Deckman, T., Pond, R. S., & Bonser, I. (2011). Belongingness as a core personality trait: How social exclusion influences social functioning and personality expression. *Journal of Personality 79(6)*: 979–1012.

Farhamy, Lisa M. (2007). Iranian Nationalism. Public Purpose – *The Journal of American University's School of Public Affairs*, Vol. V: 19–30.

Kune, N. (2011). *The need to belong: rediscovering Maslow's hierarchy of needs* Paul H. Brookes Publishers.

Liebler, C. A. (2010). Homelands and indigenous identities in a multiracial era. *Social Science Research 39(4)*: 596–609.

Sapir, E., & Mandelbaum, D. G. (1985). *Selected Writings of Edward Sapir in Language, Culture and Personality*. Vol. 342. Univ of California Press.

Vamvakidou, I., Kyridis, A., Troullou, M., & Keramyda, M. (2010). Researching active citizenship and life long learning in CICE conference publications (1998 – 2008). In: Peter Cunningham and Nathan Fretwell (eds) *Lifelong Learning and Active Citizenship. Proceedings of the twelfth Conference of the Children's Identity and Citizenship in Europe Academic Network*. London: CiCe 2010. 548–560.

Wierzbicka, A. (1997). *Understanding Cultures Through Their Key Words: English, Russian, Polish, German, and Japanese*. Oxford: Oxford University Press.

Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of….* (Edited by John B. Carroll.).

Diana Van Lancker Sidtis (New York University/Nathan Kline Institute)

John J. Sidtis (Nathan Kline Institute/New York University)

Characterization and classification of formulaic expressions: A view from language disorders

Formulaic expressions constitute a large, heterogeneous array, having in common that they are not newly created and are known to a language community. Linguistic examination has elucidated their characteristics, such as unitary structure, nuanced, nonstandard semantics, and idiosyncratic links to social context. Information about the effects of neurological damage can cast light on these characterizations and motivate classification. This report describes the effects of brain dysfunction on production of formulaic expressions in spontaneous speech, lending external support to the linguistic characterizations. Finally, a recent analysis suggests a basis for a motivated classification of selected subsets of formulaic expressions.

We have identified an overabundance of formulaic expressions in the speech of persons with left hemisphere damage and language disorder, coupled with impoverishment in persons with right hemisphere damage, implicating the right hemisphere. This portrayal conforms well to known properties of right hemisphere function: representation of complexly configured material, emotional experiencing, and management of social contingencies. Later studies revealed that persons with Alzheimer's disease (who have intact subcortical nuclei) show significantly copious production, while those with Parkinson's disease (who have dysfunctional subcortical nuclei) present with diminished proportions. These observations implicate subcortical nuclei in the production of formulaic expressions, a proposal that fits well with known properties of this cerebral site: unitary motor gestures and procedural memory. These observations converge to provide a foundation for the dual-process model. Recently, we compared incidence of formulaic expressions in monologues produced by healthy and Parkinson speakers during functional imaging. A significant interaction was found for these two groups with respect to sentence stems (I guess, so) and pause fillers (uh, um). Further analysis of formulaic behaviors in association with well defined cerebral dysfunction can further elucidate the characteristics and classifications of formulaic expressions and contribute to production models accommodating both the dual modes of language.

Laura Vilkaitė (University of Nottingham)

Reading collocations in a second language: an eye-tracking study

Collocations have been defined in corpus linguistics as words co-occurring together more frequently than expected by chance (Biber et al, 1999). So far a number of studies have investigated the psychological reality of collocations and showed that they have certain processing advantages when compared to novel language, both for native speakers and for language learners (e.g. Durrant & Doherty, 2010; Siyanova & Schmitt, 2008; Wolter & Gyllstad, 2011, 2013). However, these studies mostly looked at processing of adjacent collocations only (e.g. *spend time*), while collocations are very frequently used non-adjacently (e.g. *spend a lot of time*). Research that looked at processing of non-adjacent collocations by native speakers (Vilkaitė, in press) seems to suggest that non-adjacent collocations also facilitate processing, but the facilitative effect is smaller. The question remains if the same effect holds when processing non-adjacent collocations in one's L2.

The present study tried to address this question using an eye-tracking technique. 40 advanced non-native speakers of English read a list of sentences (N = 40) for comprehension and their eye-movements were recorded. The sentences contained phrases in one of the four conditions: adjacent collocations (*receive treatment*), controls (*arrange treatment*), non-adjacent collocations (*receive any form of treatment)* and non-adjacent controls (*arrange any form of treatment*). A number of eye-tracking measures were analysed using mixed effects modelling to investigate the effect of collocational status and adjacency as well as speaker's proficiency on collocation processing speed.

The results of the study suggest that for non-native speakers there is a clear processing advantage for adjacent collocations, even if it is moderated by language proficiency. However, there seems to be no processing advantage for non-adjacent collocations, suggesting that even advanced non-native speakers might process non-adjacent collocations qualitatively differently than native speakers.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* Harlow: Longman.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory 6(2)*: 125–155.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes 64(3)*: 429–458.

Vilkaitė, L. (2016). Are non-adjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics 32(4)*: 430–449.

Alexander Wahl (Radboud University)

MERGE: An algorithm for the corpus extraction of continuous and discontinuous formulaic sequences

Computational techniques for the automatic extraction of formulaic sequences from corpora are an important area of current research (Gries 2013). While different methods have been developed, there is no accepted best approach. I have developed and tested an extraction algorithm I call MERGE (Multi-word Expressions from the Recursive Grouping of Elements), which performs well and addresses limitations of some earlier techniques.

MERGE first extracts all adjacent and discontinuous bigrams in a corpus (with the maximum possible size in words of a discontinuity specified by the user). These bigrams are then scored and ranked according to the log likelihood association measure (Dunning 1993). Next, the winner is merged into a single lexical unit, and all corpus instances of the winning bigram are replaced by this unit. This process then repeats until reaching some minimum association threshold. The output is a rank-ordered list of sequences, with the more highly-ranked items representing the algorithm's best hypotheses for formulaic sequences.

Crucially, at later iterations, one (or both) of the elements of the winning "bigram" may themselves be earlier, multi-word merged output. This, coupled with the fact that the bigrams may be discontinuous, means MERGE learns sequences of various lengths, and with gaps. Traditional extraction approaches

have often been ill-equipped to handle such sequences, as they may extract only continuous n-grams of certain sizes (Evert 2005; but see e.g. Da Silva 1999).

As validation of MERGE, I report an empirical investigation wherein human raters assigned scores to extracted sequences based on how well they represent true formulaic sequences. Mixed-effects modeling revealed that raters' scores correlate with the rank order of extraction (marginal R^2=0.64; conditional R^2=0.84), indicating MERGE does find true formulaic sequences, with better sequences found early, and quality decreasing with increasing iterations. Thus, MERGE represents a powerful new approach for formulaic sequence extraction.

References

Da Silva, Joaquim Ferreira, Gaël Dias, Sylvie Guilloré and José Gabriel Pereira Lopes (1999). Using LocalMax algorithm for the extraction of contiguous and non-contiguous multiword lexical units. *Lecture Notes in Artificial Intelligence: Progress in Artificial Intelligence*, 1996: 113–132.

Dunning, Ted (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics 19(1)*: 61–74.

Evert, Stefan (2005). *The Statistics of Word Co-occurrences: Word Pairs and Collocations*. Dissertation. Universität Stuttgart.

Gries, Stefan Th. (2013). 50-something years of work on collocations: What is or should be next ... *International Journal of Corpus Linguistics 18(1)*: 137–165.

Stuart Webb (University of Western Ontario)

Nguyen Thi My Hang (University of Danang)

Examining second language receptive knowledge of collocation and the factors that affect learning

This study aimed to measure receptive knowledge of collocation in a way similar to how receptive knowledge of single-word items is often measured, and examined the relationship between collocation knowledge and knowledge of single-word items at different frequency levels. A 180-item collocation test was used to measure receptive knowledge of verb-noun and adjective-noun collocations that were made up of words taken from the 1,000, 2,000 and 3,000 word frequency levels. The focus of the study was limited to two types of

collocations. Verb-noun and adjective-noun collocations were chosen since these two types of collocations have been found to be frequent but problematic for learners (Nesselhauf, 2003). The results of the collocation test were then contrasted with the results of a new version of the Vocabulary Levels Test (Nation, 1983; Schmitt, Schmitt, & Clapham, 2001) to explore the relationship between knowledge of collocation and single-word items. To gain further insight into the results, the extent to which five factors (node word frequency, collocation frequency, mutual information scores, congruency, part of speech) predict knowledge of multi-word combinations was examined.

The results indicated that the participants were not close to a level of mastery of collocation knowledge at any word frequency level; knew less than 50% of each type of collocation overall, and that their knowledge of collocation significantly decreased at each level. The analysis also revealed that there were significant large positive correlations between knowledge of collocations and single-word items, and that node word frequency was the strongest predictor of receptive knowledge of collocation.

References

Nation, P. (1983). Testing and teaching vocabulary. *Guidelines 5*: 12–25.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics 24*: 223–242.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing 18(1)*: 55–88.

Seung-yun Yang (Touro College/Nathan Kline Institute)

Diana Van Lancker Sidtis (New York University/Nathan Kline Institute)

Listeners' identification and evaluation of Korean idiomatic utterances produced by persons with left- or right-hemisphere damage

Purpose: Previous research has suggested that formulaic and novel language are processed by disparate cerebral systems, lending support to the notion that these two language modes differ essentially in their processing characteristics. This study explores this notion further by investigating the effects of left- or right-

hemisphere damage on the production of matched idiomatic or literal expressions. Healthy listeners' abilities to identify, evaluate, and perceptually characterize the utterances were measured.

Method: Twenty one native speakers of Korean with left- (LHD-with aphasia) or right-hemisphere damage (RHD-without language deficit) and healthy controls (HC) produced six ditropically ambiguous (idiomatic or literal) sentences. Following instruction, utterances were elicited from the study participants without textual support in order to simulate natural speech production. Healthy listeners identified the sentence types and provided goodness ratings for each sentence and performed perceptual ratings of voice quality

Results: Healthy native Korean listeners successfully discriminated the intended meanings of ditropically ambiguous sentences produced by HC speakers. Listeners showed significantly decreased performance in discriminating utterances produced by RHD individuals, whose utterances also generated poor goodness ratings compared to those obtained from LHD and HC. Goodness ratings correlated with listeners' performance. Nasal voice quality was associated with idiomatic sentences.

Conclusion: Successful identification by healthy listeners of the intended meanings of ditropic utterances produced by healthy speakers affirms native competence for distinguishing the meanings from the auditory stimulus alone, implying knowledge of formulaic expressions. The negative effect of RHD on listeners' identification and goodness ratings lends support to the notion that cerebral processes are allocated differently for these two modes of language. Sentence type was significantly associated with selected voice qualities. These findings support previous reports of prosodic information serving to signal idiomatic versus literal meanings as well as a right-hemisphere involvement in formulaic language and the dual-processing model of language.

PARTICIPANTS

Name	Affiliation	Email address
Hana Al-Mutairi	University of Nottingham	aexhma@nottingham.ac.uk
Gundula Baehre		gundulabaehre@gmail.com
Liubov Baladzhaeva	University of Haifa	baladjaeva@gmail.com
Keith Bateson	McMaster University	batesokg@mcmaster.ca
Donata Berūkštienė	Vytautas Magnus University	donaber@gmail.com
Jennifer Boutz	University of Maryland	jboutz@umd.edu
Nihal Çalışkan	Yıldırım Beyazıt University	caliskanihal@hotmail.com
Gareth Carrol	University of Birmingham	g.carrol@bham.ac.uk
Jiaoyue Chen	University of Southampton	jc6e10@soton.ac.uk
Svitlana Chugu	Vite Knute, Vinnytsia	svetachugu@gmail.com
Jurga Cibulskienė	Lithuanian University of Educational Sciences	jurga.cibulskiene@leu.lt
Jelena Čolovic-Marković	West Chester University	jmarkovic@wcupa.edu
Natalia Maria Durus	Institut National des Langues et Civilisations Orientales	natalia.durus@gmail.com
Maria Fernandez-Parra	Swansea University	m.a.fernandezparra@swansea.ac.uk
Parvin Gheitasi	Umea University	parvin.gheitasi@umu.se
Cordula M. Glass	Friedrich-Alexander-Universität Erlangen-Nürnberg	cordula.glass@fau.de
Beatriz González-Fernández	University of Nottingham	bea.gonzf@gmail.com
Lukasz Grabowski	Opole University	lukasz@uni.opole.pl
Lenore Grenoble	University of Chicago	grenoble@uchicago.edu
Jonė Grigaliūnienė	Vilnius University	jone.grigaliuniene@ gmail.com
Ewa Guz	The John Paul II Catholic University of Lublin	ewasik@o2.pl
Glenn Hadikin	University of Portsmouth	glenn.hadikin@port.ac.uk
Ferdy Hubers	CLS/CLST, Radboud University	f.hubers@let.ru.nl
Anne Sofie Jakobsen	University of Copenhagen	asj@hum.ku.dk
Rita Juknevičienė	Vilnius University	rita.juknevicienе@takas.lt
Danguolė Kalinauskaitė	Vytautas Magnus University	danguole.kalinauskaite@vdu.lt
Mariusz Kamiński	University of Applied Sciences, Nysa	mariusz20ski@gmail.com

Name	Affiliation	Email address
Jurga Kasteckienė	Vilnius University	jurgast@yahoo.com
Seval Kömürcü	Freiburg University	seval.koemuercue@anglistik.uni-freiburg.de
Jolanta Kovalevskaitė	Vytautas Magnus University	j.kovalevskaite@hmf.vdu.lt
Yuriy Kovalyuk	Yuriy Fedkovych Chernivtsi National University	yuriy_kovalyuk@yahoo.de
Myroslava Kovalyuk	Yuriy Fedkovych Chernivtsi National University	jarich.mv@gmail.com
Benjamin Kremmel	University of Nottingham, University of Innsbruck	benjamin.kremmel@nottingham.ac.uk
Batia Laufer	University of Haifa	batialau@research.haifa.ac.il
Agnieszka Lenko-Szymanska	University of Warsaw	a.lenko@uw.edu.pl
Saskia Lensink	LUCL, Leiden University	s.e.lensink@hum.leidenuniv.nl
Justina Mandravickaitė	Vilnius University	mandravickaite@gmail.com
Hiroyuki Matsumoto	Hokkai Gakuen University	matsumoto@econ.hokkai-s-u.ac.jp
Alexander Nakhimovsky	Colgate University	adnakhimovsky@colgate.edu
Natsumi Okuwaki	Tsuru University	okuwaki@tsuru.ac.jp
Magali Paquot	Université catholique de Louvain	magali.paquot@uclouvain.be
Rūta Petrauskaitė	Vytautas Magnus University	r.petrauskaite@hmf.vdu.lt
Carlos Prado-Alonso	University of Oviedo	pradocarlos@uniovi.es
Geraint Paul Rees	Universitat Pompeu Fabra	geraintpaul.rees@upf.edu
Teresė Ringailienė	Vytautas Magnus University	terese.ringailiene@vdu.lt
Michael Rodgers	University of Nottingham	Michael.Rodgers@nottingham.ac.uk
Jūratė Ruzaitė	Vytautas Magnus University	j.ruzaite@hmf.vdu.lt
Birutė Ryvitytė	Vilnius University	birute.ryvityte@flf.vu.lt
Svetlana Schukina	Belarusian State Academy of Music	schukina.minsk@gmail.com
Inesa Šeškauskienė	Vilnius University	inesa.seskauskiene@flf.vu.lt
John Sidtis	Nathan Kline Institute/New York University	john.sidtis@nyu.edu
Magda Stroinska	McMaster University	stroinsk@mcmaster.ca
Narcisse Torshizi	McMaster University	torshizin@mcmaster.ca
Justina Urbonaitė	Vilnius University	justina.urbonaite@flf.vu.lt

Name	Affiliation	Email address
Diana Van Lancker Sidtis	New York University/Nathan Kline Institute	diana.sidtis@nyu.edu
Irina Vaynshteyn	Touro College /Nathan Kline Institute	iv288@nyu.edu
Laura Vilkaitė	University of Nottingham	laura.vilkaite@nottingham.ac.uk
Alexander Wahl	Radboud University, Donders Institute	a.wahl@psych.ru.nl
Stuart Webb	University of Western Ontario	swebb27@uwo.ca
Alison Wray	Cardiff University	wraya@cardiff.ac.uk
Seung-yun Yang	Touro College / Nathan Kline Institute	seung-yun.yang3@touro.edu
Monique Yoder	LCC International University	myoder@lcc.lt
Rūta Zukienė	Vilnius University	ruta.zukiene@gmail.com

Notes

THE OLD CAMPUS OF VILNIUS UNIVERSITY

Location of rooms

1. Writers' Hall
2. Krėvės' Auditorium
3. German Room
4. Canada Room
5. University Café