The 9th Inter-Varietal Applied Corpus Studies (IVACS) Conference

Language, Communities & Mobility

University of Malta

June 13 – 15, 2018

IVACS 2018 Wednesday 13th June 2018

The conference is to be held at the University of Malta Valletta Campus, St. Paul's Street, Valletta.

12:00 – 12:45	Conference Registration			
12:45 – 13:00	**Conferencing opening welcome and address** University of Malta, Dean of the Faculty of Arts: Prof. Dominic Fenech Room: Auditorium			
13:00 – 14:00	**Plenary**: Dr Robbie Love, University of Leeds Overcoming challenges in corpus linguistics: Reflections on the Spoken BNC2014 Room: Auditorium			
Rooms	**Auditorium (Level 2)**	**Meeting Room 1 (Level 0)**	**Meeting Room 2 (Level 0)**	**Meeting Room 3 (Level 0)**
14:00 – 14:30	Adverb use in spoken interaction: insights and implications for the EFL classroom - Pascual Pérez-Paredes & Geraldine Mark	'Yeah, no, everyone seems to be saying that': 'New' pragmatic markers as represented in fictionalized Irish English - Ana Mª Terrazas-Calero & Carolina Amador-Moreno	On discourse markers in Lithuanian argumentative newspaper discourse: a corpus-based study - Anna Ruskan	Lexical Bundles in the Description of Drug-Drug Interactions: A Corpus-Driven Study - Lukasz Grabowski

	Auditorium (Level 2)	Meeting Room 1 (Level 0)	Meeting Room 2 (Level 0)	Meeting Room 3 (Level 0)
14:30 – 15:00	"Um so yeah we're <laughs> you all know why we're here": hesitation in student presentations - Kristin Horan	So you have to follow different methods and clearly there are so many methods: analyzing assertiveness in Brazilian university students spoken interlanguage - Bárbara Malveira Orfanò	The Study of Ideological Bias through Corpus Linguistics in Syrian Conflict News from CNN and Russia Today News Outlets - Pimkarn Wasuwong	Using learner corpus data to inform the development of a diagnostic language tool and individualised learning programme - Caitlin Neachtain
15:00 – 15:30	**Coffee Break**			
Rooms	**Auditorium (Level 2)**	**Meeting Room 1 (Level 0)**	**Meeting Room 2 (Level 0)**	**Meeting Room 3 (Level 0)**
15:30 – 16:00	Factors impacting the usage of UK-based international foundation - Dana Therova	Pronouns in CHET and CECheT: revealing authorial presence - Begoña Crespo & Isabel Moskowich	Where are they from? Geographical representation of migrants before and after Brexit - Dario Del Fante	Between an /r/ and a schwa: Illustrating some of the joys and woes of big data for speech analysis - Alexandra Vella & Sarah Grech
16:00 – 16:30	A Corpus-based Examination of the Vocabulary Used in Medical Multiple Choice Questions - Emily Harms	Evidence of nesting in metaphoric language - Katie J. Patterson	Unveiling the journalist's stance: Focus on The Economist's coverage of the Brexit referendum - Roberta Facchinetti, Sara Corrizzato & Valeria Franceschi	Varieties of English in English Language Teaching in Germany - Lisa Scheiwe

16:30 – 17:00	Corpus-driven Approach to Lexical Semantic Change: A Case Study - Gang Yao	NOT THE 99 PER CENT. (Under-) representation of the less wealthy in British fiction corpora - Michael Pace-Sigge	Some methodological challenges in the identification of discriminatory attitudes and counter-speech in online newspaper comments - Stavros Assimakopoulos & Rebecca Vella Muskat	Something borrowed and something new: a dual approach for studying inter-varietal differences multilingually - Ilmari Ivaska, Silvia Bernardini & Adriano Ferraresi
17:00 – 17:30	An analysis of a corpus of Teacher Talk in the secondary-level EFL classroom: highlighting differences in modality and in the use of phrasal verbs - Eric Nicaise		The Trojan Horse: the Construction of Iraqi Shiites in Times of Wars in the US Press - Dhiaa Kareem	
17:30	**Welcome Reception**			

Thursday 14th June 2018

08:30 – 09:00	Registration			
Room	**Auditorium (Level 2)**	**Meeting Room 1 (Level 0)**	**Meeting Room 2 (Level 0)**	**Meeting Room 3 (Level 0)**
09:00 – 09:30	A Corpus-Based Analysis of Social Interaction in LTE Contexts - Fiona Farr, Angela Farrell & Elaine Riordan	This is England, speak English! A corpus-assisted critical study of language ideologies in the British press - Gavin Brookes & David Wright	'I hear whispers and I don't like it': Conflict in a Corpus of Spoken Interactions from the Construction Industry - Almut Koester & Michael Handford	Where is the I? Cultural Diversity and the Self in L2 Student Writing - Jelena Runić
09:30 – 10:00	The design and annotation of the TEG learner corpus of Irish - Aoife Ní Ghloinn, Elaine Uí Dhonnchadha & Anne O' Keeffe	Using corpus linguistic methods to uncover verbal cues to deception - Mathew Gillings	Stoefpears Run the World': A Corpus Study into English Code-Mixing in Dutch Youths' Computer-Mediated Communication - Lieke Verheijen, Roeland van Hout & Laura de Weger	Exploring the Use of Collocation in the Writing of Saudi University Students Using Corpus Methodology - Huda Yahya Khoja
10:00 – 10:30	'That's the dilemma you know': Patterns and functions of *you know* as used by L1- and L2-speakers of English - Giovani Santos	Peeling back the layers: using CL and CA to examine Focus Group talk - Róisín Ní Mhocháin	"Straight talking money" A corpus assisted critical discourse analysis of payday loans websites - Hazel Price & Dan McIntyre	The Relationship between Phraseology and Writing Quality: Implications for the Assessment of Student Writing in an English as a Lingua Franca Context - Lee McCallum
10:30 – 11:00	**Coffee**			

Room	Auditorium (Level 2)	Meeting Room 1 (Level 0)	Meeting Room 2 (Level 0)	Meeting Room 3 (Level 0)
11:00 – 11:30	A Corpus-Aided Comparative Analysis of Native and Non-Native Speaking English Language Teachers' Approaches to In-Class Spoken Feedback - Jane Seely	Corpus design and construction: the challenges faced by minoritized languages - Dawn Knight	Predatory discourses and the incitement of violence against women in an online discussion forum - David Wright & Ikechukwu E. Onyenwe	The readability of Flemish legal documents - Michael Bauwens & José Tummers
11:30 – 12:00	"I went to the cinema" and "I came to the conclusion": Zoning in on acquisitional patterns of learner grammar development - Geraldine Mark, Pascual Pérez-Paredes & Anne O'Keeffe	Linguistic Resources for facilitating MOOC discussion forum: A Corpus Driven Approach - Shi Min Chua	On formulaicity of Russian patient information leaflets: a corpus-driven study - Lukasz Grabowski	Linked noun groups: teaching awareness of genre-specific uses - Michael Pace-Sigge
12:00 – 13:00	**Lunch**			
13:00 – 14:00	**Plenary**: Prof. Ute Römer, Georgia State University Verb constructions in second language acquisition: From learner corpus analysis to pedagogical recommendations Room: Auditorium			
Rooms	**Auditorium (Level 2)**	**Meeting Room 1 (Level 0)**	**Meeting Room 2 (Level 0)**	**Meeting Room 3 (Level 0)**
14:00 – 14:30	Lexical development in English school children's writing from six to sixteen - Philip Durrant & Mark Brenchley	The Corpus of Irish Narratives - Chris Fitzgerald	The diagnostic potential of Contextual Prosodic Theory in establishing forensically the features of manipulative	Conceptualising context in corpus linguistics: moving beyond metadata - Michael Handford & Dawn Knight

		newspaper discourse in Russian - Marija Milojkovic	
14:30 – 15:00	Let's talk about maths: using classroom data for collaborative professional development - Joanna Baumgart	Difficult Words: An Analysis of Terminology in a Corpus of Irish Education Materials - Mícheál J. Ó Meachair & Elaine Uí Dhonnchadha	Exploring markers of emergent professional identity and status in an international mobile community - Margaret Healy	On corpus linguistic approaches to study language: a critical reflection - Milena Hebal-Jezierska & Łukasz Grabowski
15:00 – 15:30	**Coffee**			
Rooms	**Auditorium (Level 2)**	**Meeting Room 1 (Level 0)**	**Meeting Room 2 (Level 0)**	**Meeting Room 3 (Level 0)**
15:30 – 16:00	Using a Multimodal Corpus in EFL Classrooms; an Action Research on Learner Attitude - Sara Aljohani	You can't achieve a shrimp: a corpus-driven approach to increased precision in noun-verb collocations - Jane Templeton	Populism in Italian politicians' social messages. A corpus-based study of three politicians' social media discourse - Dario Del Fante	Stories and 'story-likes': How lecturers use real and hypothetical events to scaffold knowledge in different EMI contexts - Siân Alsop & Hilary Nesi
16:00 – 16:30	Using a learner corpus to develop English language teaching (ELT) materials targeted at Spanish learners - Julie Moore	Understanding the Interests of Communities of TripAdvisor Travellers through a Corpus-Based Discourse Analytic Approach - Phoenix Lam	Science in the UK media: An analysis of two newspapers - Geri Popova	Investigating advanced translation learners' problems of lexical cohesion: A corpus-based study of lexical repetition in English to Chinese translation - Jun Pan & Honghua Wang
16:30 – 17:00	A corpus-based comparison of registers in EFL school	A learner corpus research on the diagnosis of accuracy in	Corpus Analysis of Modal Verb unda in Georgian	The Effect of Indirect Coded Corrective Feedback and

	textbooks for secondary schools in France, Germany and Spain - Elen Le Foll	grammatical morphemes by interlanguage users: A case of prospective English teachers in Turkey - Erdem Akbas & Zeynep Olcu-Dincer	- Sophiko Daraselia & Nino Sharashenidze	Error-Tagged Learner Corpora on the Reduction of Errors in L2 Writing Tasks - Chiachieh Tang
17:00 – 17:30	The ELT Archive Textbook Corpus: How much has language teaching changed? - Rachel Allan	Multimodal Corpus in Teaching Russian as a Foreign Language - Michal Kozdra	Refugees in the Spanish written media: A corpus-based study from a semantic preference perspective. Towards a categorization of a negative portrayal - Gema Alcaraz-Maĕrmol & Jorge Soto Almela	The Use of Inferential Lexical Bundles in Academic Writing of Expert Writers: A Contrastive Study - Fatih Güngör
20:00	**Conference Dinner** Venue: Cargo Bar & Dine 19:00 Transport to Birgu			

Friday 15th June 2018

08:30 – 09:00	Registration			
Room	**Auditorium (Level 2)**	**Meeting Room 1 (Level 0)**	**Meeting Room 2 (Level 0)**	**Lecture Room 3 (Level 0)**
09:00 – 09:30	Rhetoric and Reader Pronouns in Economics Academic Writing: A Corpus-Based Contrastive Analysis of English, French and Spanish Research Articles - Niall Curry	A corpus-informed exploration of the creation of intimacy in 'First Dates' (Ireland) - Brian Clancy	CorCenCC: applying the sociolinguistics of new speakers within a contemporary corpus of Welsh - Steve Morris & Dawn Knight	The 'Research Article' event: Towards a more unified dress-code - Maria N. Melissourgou & Laura Maruster
09:30 – 10:00	Language Assessment in Ireland at Senior Cycle and the CEFR: A comparative analysis of assessment instruments across languages examined at Senior Cycle in relation to the CEFR - Sorcha Ryan & Ann Devitt	The pragmatics of poverty in nineteenth century English pauper letters - Ivor Timmis	"Box, Box Now": A Corpus-Based Analysis Of Formula One Radio Messages - Jukka Tyrkkö & Hanna Limatius,	Exploring disciplinary differences in academic writing using semantic tags - Michael Barlow & Phuong le
10:00 – 10:30	**Coffee**			

Room	Auditorium (Level 2)	Meeting Room 1 (Level 0)	Meeting Room 2 (Level 0)	Lecture Room 3 (Level 0)
10:30 – 11:00	Developing a CEFR-based and corpus-based curriculum for English for Academic Purposes - Stergiani Kostopoulou	'Sure where else is there? 'Pragmatic markers in radio advertising in Ireland - Joan O'Sullivan	Language, mobility and climate change at the crossroad: A diachronic corpus-assisted discourse analysis of the climate-migration nexus in the press - Cinzia Bevitori & Jane Helen Johnson	Teaching the "unlearnable" - A methodology for the use of corpora for Hungarian as a foreign language - Szilvia Szita
11:00 – 11:30	What "should" happen in a Modern foreign language class? A corpus-based analysis of secondary level Inspectorate reports in Ireland - Stéphanie O'Riordan	"Hey, can I borrow your…: Is sharing your corpus linguistic tools with others a good idea? - Jane Evison & Elaine Vaughan	Corpora for Remembering and Learning: The Second Amendment, the 2008 Supreme Court Ruling, and the Absolute Phrase - James Vanden Bosch	
11:30 – 12:30	**Plenary**: Dr Rachelle Vessey, Birkbeck University of London From cross-cultural to superdiverse corpus linguistics: language ideologies and communities Room: Auditorium			
12:30	**Conference Closing**			

Table of Contents

Overcoming challenges in corpus linguistics: Reflections on the Spoken BNC2014

Robbie Love, University of Leeds

The *Spoken BNC2014* (Love et al. 2017, Love forth.) is an important component of the new *British National Corpus 2014*; a large dataset representing current British English usage across different situations, which is being compiled by Lancaster University in collaboration with Cambridge University Press. It is the successor to the spoken component of the original *British National Corpus* (Crowdy 1995) and was released publicly via Lancaster University's CQPweb server (Hardie 2012) in September 2017.

In this talk, I want to (a) reflect on the compilation of and initial reception to the Spoken BNC2014, and (b) discuss some wider and emergent challenges in corpus linguistics which I have noticed through my work on this corpus.

In terms of corpus construction, I pay attention to other contemporary spoken corpus projects such as the spoken component of *CorCenCC* (Knight et al. 2016) and *FOLK* (Schmidt 2016), and consider the role of representativeness in corpus design. I argue that representativeness is an ideal but that it is inevitable – due to practical constraints – that there will be some differences between the original design of a large 'national' corpus and the finished product, and that it is important to be honest, critical and realistic about representativeness. I demonstrate this by discussing precisely what the Spoken BNC2014 does and does not represent, showing that the corpus makes a range of valuable contributions to the field.

My work on the Spoken BNC2014 has raised a few challenges – and funny stories – which I will share in this talk. These address issues such as:

- the benefits and risks of researchers engaging with the media;
- the role of a 'national' British corpus in the context of an increasingly multilingual, multicultural and mobile UK society;
- the growth of corpus linguistics and 'blind faith' in its methods and results;
- the responsibility for good corpus research;
- Who and what corpus linguists are.

I conclude by noting that, as corpus linguistics grows and interacts increasingly with other disciplines, public corpora such as the Spoken BNC2014 are being used more and more, for a wider range of purposes and by more diverse groups of people. The growth and spread of corpus linguistics is undoubtedly a good thing, though it brings with it the increasing responsibility of us all to protect its integrity.

References:

Crowdy, S. (1995). The BNC spoken corpus. In G. Leech, G. Myers, & J. Thomas (Eds.), *Spoken English on Computer: Transcription, Mark-Up and Annotation* (pp. 224-234). Harlow: Longman.

Hardie, A. (2012). CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics 17*(3), 380-409.

Knight, D., Neale, S., Watkins, G., Spasić, I., Morris, S., & Fitzpatrick, T. (2016, June). *Crowdsourcing corpus construction: contextualizing plans for CorCenCC (Corpws Cenedlaethol Cymraeg Cyfoes – The National Corpus of Contemporary Welsh).* Paper presented at the IVACS 2016 conference, Bath Spa University, UK.

Love, R. (forth). *Overcoming Challenges in Corpus Construction: The Spoken British National Corpus 2014.* New York: Routledge.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics, 22*(3), 319-344.

Schmidt, T. (2016). Good practices in the compilation of FOLK, the Research and Teaching Corpus of Spoken German. *International Journal of Corpus Linguistics, 21*(3), 396-418.

Adverb use in spoken interaction: insights and implications for the EFL classroom

Pascual Pérez-Paredes, Faculty of Education, University of Cambridge
Geraldine Mark, Mary Immaculate College, University of Limerick

Research has shown that adverb use is a powerful predictor of communicative competence among EFL learners (Pérez-Paredes & Diez-Bedmar 2012). Most research in learner adverb use has focused on academic written registers and in general these studies have shown evidence of overuse of adverbs along with a lack of register awareness. Studies that examine adverb use in spoken language and those that consider pragmatic aspects are not abundant (excepting Pérez-Paredes (2010), Aijmer (2011), Gablasova et al. (2017). Equally, studies comparing the use of adverbs in the same task or research condition are rare. This paper describes the findings of a study comparing high frequency adverb use in spoken communication across four L1 datasets. The study draws on the Chinese, Spanish and German components of the LINDSEI database (Gilquin, DeCock and Granger 2010) and the first language English data in the LOCNEC extended corpus. Participants across all four datasets complete the same three tasks in English. Taking a combined quantitative and qualitative approach, the study examines the contexts of use and underlying pragmatic functions of adverbs across each different task. This paper describes some of the insights from this study and the implications for the language classroom in relation to task effect, first language and language patterning.

References:

Aijmer, K. (2011). Well I'm not sure I think… The use of well by non-native speakers. *International Journal of Corpus Linguistics*, 16(2), 231-254.

Díez-Bedmar, B., & Pérez-Paredes, P. (2012). A cross-sectional analysis of articles in learner writing. In Y. Tono, Y. Kawaguchi, & M. Minegishi (Eds.), *Developmental and Crosslinguistic Perspectives in Learner Corpus Research* (pp. 139-158). John Benjamins.

Gablasova, D., Brezina, V., McEnery, T., & Boyd, E. (2017). Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics*, 38(5), 613-637.

Gilquin, G., De Cock, S., & Granger, S. (2010). The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM.

Pérez-Paredes, P. (2010). The death of the adverb revisited: Attested uses of adverbs in native and non-native comparable corpora of spoken English. *Exploring New Paths in Language Pedagogy Lexis and Corpus-based Language Teaching*. Equinox, London, 157-172.

'Yeah, no, everyone seems to be saying that': 'New' pragmatic markers as represented in fictionalized Irish English

Ana Mª Terrazas-Calero, Mary Immaculate College, University of Limerick
Carolina Amador-Moreno, University of Extremadura

The unpredictability of linguistic trends is best exemplified by the rise and fall in popularity of pragmatic markers (PMs). That would certainly be the case of upsurging 'yeah, no', which can be noticed in many different varieties of English around the world nowadays. Language observers, commentators and non-specialists have condemned its use, and although relatively unexplored by academics, this PM has been analyzed in the speech of young, male speakers in AmE and AuE (Liberman 2008; Burridge and Florey 2002). Using a corpus of spoken, informal interviews and conversations in Australia, Burridge and Florey (2002) found that, formally, 'yeah no' is a multifunctional marker which can create cohesion, indicate assent/dissent, and act as a hedge and mitigating, face-saving device (150). Although the only studies conducted on this PM to date have been carried out on AuE and AmE, 'yeah, no' has also been 'spotted' in other varieties like NZE or SAE, but nothing has been said about its use in Irish English (IrE). The present paper, therefore, investigates the use of 'yeah, no' in contemporary IrE as portrayed in Paul Howard's incredibly successful Ross O'Carroll-Kelly (RO'CK) series. By means of quantitative and qualitative analyses of the use of this marker in a corpus of eight RO'CK novels and plays spanning a period of five years, we will explore the use and the pragmatic functions 'yeah, no' serves in contemporary Dublin English according to Howard. Our results will, then, be contrasted with the spoken data in the Limerick Corpus of Irish English so as to corroborate the validity of Howard's representation. Finally, we will also argue that Howard's keen decision to present this marker in his books is a clear attempt at mirroring naturalistic and spontaneous Dublin English in the written mode.

References:

Burridge, K., & Florey, M. (2002). 'Yeah-no He's a Good Kid': A Discourse Analysis of Yeah-No in Australian English. *Australian Journal of Linguistics*, 22: 149-171.

Liberman, M. (2008, April 3). Yeah no [Blog port]. Retrieved from http://itre.cis.upenn.edu/~myl/languagelog/archives/005523.html

On **discourse markers in Lithuanian argumentative newspaper discourse: a corpus-based study**

Anna Ruskan, Vilnius University

Functional distribution and variation of discourse markers have been investigated widely in argumentative newspaper discourse in Germanic and Romance languages (Dafouz-Milne 2008; Speyer, Fetzer 2014; Breeze 2016). Attention has been paid to textual and interpersonal functions of discourse markers, their co-occurrences in sequences and contribution to argumentative patterns. The studies reveal a number of universal as well as language and culture specific features of discourse markers while developing authorial argumentation. The aim of the current study is to explore the textual and interpersonal functions of discourse markers in Lithuanian argumentative newspaper discourse by comparing their formal and functional distribution in editorials and opinion articles. Although both argumentative genres represent a subjective authorial opinion and appeal to the reader, they may differ in the choice of discourse markers, characterised also in terms of (inter)subjectivity (Traugott 2010). The functional profile of discourse markers is identified by analysing their position, scope, co-occurrences and (inter)subjective dimension. The study combines qualitative and quantitative methods of analysis. The data have been drawn from the self-compiled corpus of editorials and opinion articles obtained from the Lithuanian newspaper Verslo žinios 'Financial news' in the years 2016-2017. The argumentative dimension of discourse markers is also explored through the analysis of their English translation correspondences found in the Europarl Corpus. The combination of monolingual and parallel corpora allows for identifying the formal and functional versatility of discourse markers. References Breeze, R. 2016. Negotiating alignment in newspaper editorials: The role of concur-counter patterns. Pragmatics, 26 (1).

References:

Dafouz-Milne, E. 2008. The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics* 40, 95-113.

Speyer, A. & A. Fetzer. 2014. The coding of discourse relations in English and German argumentative discourse. *The Pragmatics of Discourse Coherence*, eds. Gruber, H. & G. Redeker. Amsterdam/Philadelphia: John Benjamins. 87-120.

Traugott, E. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. *Subjectification, Intersubjectification and Grammaticalization*, eds. Davidse, K., L. Vandelanotte & H. Cuyckens. Berlin/New York: De Gruyter Mouton. 29-74.

Lexical Bundles in the Description of Drug-Drug Interactions: A Corpus-Driven Study

Lukasz Grabowski, University of Opole

Recently conducted studies of lexical bundles showed that these prefabricated multi-word sequences constitute important building blocks of specialist discourses (e.g. Biber et al. 2004; Biber 2006; Goźdź-Roszkowski 2011), including a pharmaceutical one (Grabowski 2015). The aim of this presentation is to extend the latter research by exploring lexical bundles in English pharmaceutical texts describing drug-drug interactions. The study uses 300 text samples extracted from the DrugDDI Corpus originally collected in the Drugbank database (Segura-Bedmar et al. 2010). In short, drug-drug interactions are situations whereby drug affects the action of another one (Rowland 2008 :1), they account for 6-30% of all adverse reactions and that is why they constitute a particularly significant problem in clinical practice (Ionescu & Caira 2005: 296). Consequently, it is essential that professionals in the medical and pharmaceutical field, notably non-native speakers of English, be familiar with recurrent phraseologies used to describe drug-drug interactions. Hence, an attempt is made in this study to describe the use and discourse functions of 4-word lexical bundles used in pharmaceutical texts to describe drug-drug interactions. However, apart from presenting new descriptive data on recurrent phraseologies used in the pharmaceutical domain, another aim of this study is to reflect on the ways lexical bundles have been typically explored so far across different text types and genres. Although the lexical bundles approach is largely an empirical, data-driven method of identification and exploration of frequent phraseologies in texts, it is not devoid of limitations and methodological problems. The specific problem, among others, discussed in this presentation concerns filtering out overlapping bundles (e.g. by employing Formulex method proposed by Forsyth (2015) as compared with the traditional lexical bundles approach). Hence, this research is also intended as an opportunity to reflect on the methodologies used to explore recurrent phraseologies in specialist registers.

References:

Biber, Douglas, Conrad, Susan & Cortes, Viviana. 2004. "If you look at…: Lexical bundles in university teaching and textbooks". *Applied Linguistics* 25(3): 371–405.

Goźdź-Roszkowski, Stanisław. 2011. *Patterns of Linguistic Variation in American Legal English. A Corpus-Based Study*. Frankfurt am Main: Peter Lang.

Forsyth, Richard. 2015. Formulib: Formulaic Language Software Library. <http://www.richardsandesforsyth.net/zips/formulib.zip>

Rowland, Malcolm. 2008. "Introducing Pharmacokinetic and Pharmacodynamic Concepts". *Drug-Drug Interactions*. Second Edition, A. David Rodrigues (ed.), 1–29. New York/London: Informa Healthcare.

"Um so yeah we're <laughs> you all know why we're here": hesitation in student presentations

Kristin Horan, National University of Ireland

The Cambridge, Limerick and Shannon (CLAS) Corpus is a one million word corpus of hospitality and business language recorded in the Shannon College of Hotel Management in 2008-2009. The corpus is a collaborative project between Mary Immaculate College, Shannon College of Hotel Management and Cambridge University Press, with the data feeding into the English Profile project.

Using a 150,000 word sub-corpus of the CLAS corpus, this study explores the use of hesitation by students during the delivery of academic presentations. The sub-corpus is comprised of data from undergraduate students spanning three years of a degree programme and derived from a variety of business and hotel-specific modules. The sub-corpus consists of roughly half native and half non-native speaker data, which allows the researcher to compare hesitation across both cohorts.

This paper identifies the types of hesitation markers used most frequently by the students in the presentations, comparing the commonalities and differences between native and non-native speaker usage. The paper further examines the functions of these markers, by investigating the placement of the hesitation markers within clause structure and their proximity to other hesitation devices such as laughter and pauses. Methodologically, the paper will also reflect on the choices and challenges of annotating the types of hesitation in the dataset.

So you have to follow different methods and clearly there are so many methods: analyzing assertiveness in Brazilian university students spoken interlanguage

Bárbara Malveira Orfanò

Most of the literature on boosters originates from the study of amplifiers, a syntactic rather than a pragmatic entity. Quirk et al. (1985: 590) distinguish between two types of amplifiers: maximizers, which denote an absolute degree of intensity and therefore occupy the extreme upper end of the scale and boosters, which denote a high degree but without reaching the extreme end of the scale. Differently, Carter and McCarthy (2006: 279) take a pragmatic approach to the study of boosters and explain that speakers sometimes need to assert a claim or viewpoint quite directly and more confidently. Following their approach, this study investigates boosting markers in two corpora. The first a learner oral corpus being compiled at the Federal University of Minas Gerais/Brazil (BRASE) and the second a sub-corpus from the British Academic Spoken English (BASE). The Brazilian learner corpus comprises oral presentations recorded in an English for Academic Purpose class and has, at the present moment, 50,000 tokens. The BASE corpus comprises lectures and seminars from different disciplines and has 1.644,942 tokens. Analytical tools from the area of Corpus Linguistics are employed and items functioning as boosters are investigated in both corpora indicating similarities and differences between the two datasets. Employing both quantitative and qualitative analysis preliminary results demonstrate that Brazilian students overuse booster items when the results are compared with the BASE corpus. This is an indication that this group of learners expresses themselves in a more assertive way differing from natural oral academic discourse. Thus, the study shall provide important insights into learners oral interlanguage and consequently contribute to the body of research on English for Academic Purposes considering the current status of the area in Brazil.

References:

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: a comprehensive guide; spoken and written English grammar and usage*. Ernst Klett Sprachen.

Quirk, R., Greenbaum, S., & Leech, G. and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London and New York: Longman.

A Study of Ideological Bias Through Corpus Linguistics in Syrian Conflict News from CNN and Russia Today News Outlets

Pimkarn Wasuwong, Thammasat University

Bias in news writing has become an important issue since it distorts readers' perception towards the actual situation. That is, writers from different news outlets tend to write news stories in favour of their readers who share the same background knowledge, culture and values. So, there is a tendency that the writers introduce bias in the story to draw readers' agreement. Consequently, readers are suggested to engage in critical reading to detect the bias. However, it is not an easy skill since sometimes bias is not introduced to the story explicitly. The present research applied a multi-disciplinary approach: Ideological Square Positive self- presentation and negative other- presentation (van Dijk, 1993) and corpus linguistics to detect bias in Syrian conflict news from CNN and Russia Today (RT) news outlets. This hybrid approach benefits in reducing the gap between qualitative and quantitative research. That is, using only socio-political knowledge does not provide valid results in analysing ideological bias. It needs statistical device from corpus linguistics, such as frequency to help in generalization. In present study, three collocation patterns which are noun + noun, adjective + noun and adverb+ verb are examined through the ideological square principle based on the frequency of occurrence in both news corpora. Noun are used to reflect country name indicating group relations, whereas adjective and adverb are the supplement providing extra information on intensity and operation in the Syrian war. Additionally, the study investigated how ideologies are carried through syntax, semantics and pragmatics which explain the ways writers use language to introduce bias to a news story. The results reveal that both news outlets introduced bias by reporting imbalanced viewpoints according to ideological square. Moreover, the present research found that news writers applied various linguistic tools such as pragmatic presuppositions, active and passive voice and lexical choice to carry the ideologies. In future research, researchers may apply this multidisciplinary approach as a prototype to detect bias in other text genres.

Reference:

Van Dijk, T. A. (1993). Principles of critical discourse analysis. Discourse & society, 4(2), 249-283.

Using learner corpus data to inform the development of a diagnostic language tool and individualised learning programme

Caitlin Neachtain, Mary Immaculate College

The application of learner corpus methods to second language acquisition research is less common in the study of lesser-used languages, yet the use of corpus methods can provide valuable data with which to understand language proficiency levels in general and also to assess individual proficiency, in particular with a view to creating effective learner-oriented instructional materials. CAES (Corpas Anótáilte de réir Earráidí Scríofa) is a growing collection of texts written in Irish by student teachers during their third level study. This custom-built corpus comprises 300 individual texts (circa 70,000 words). In this project, error analysis is used to identify and analyse the grammatical errors that are most prevalent in students' written production in order to build a diagnostic tool with an in-built feedback function aimed at supporting and improving students' language awareness and promoting efficient, independent language learning. Based on the compiled data, the corpus informs the creation of meaningful tasks for language learning, with learning objectives prioritised for each learner individually in order to maximise the impact of the diagnostic intervention. The corpus also provides a resource bank of phrases and sentences for use in ongoing Irish language diagnostic testing and explicit grammar instructional activities within the online learning programme. Some key issues addressed by this study include identifying common patterns in learners' interlanguage, variability in individual output based on the usage context, and the reliability of learners' beliefs in relation to their language proficiency level across all language skills based on the analysis of the corpus. While development and analysis of the CAES learner corpus is ongoing, this current study provides baseline results that support the incorporation of diagnostic language tools to improve learner outcomes, with particular regard to customised programmes for lesser-used languages.

Factors impacting the usage of UK-based international foundation students' academic vocabulary in assessed writing

Dana Therova, The Open University

To date little research has been conducted into the writing of international foundation students at UK-based universities. This is surprising considering that academic writing is one of the core academic skills that foundation-level students are expected to develop during the foundation year as well as the fact that written assignments continue to remain the main mode of assessment in higher education institutions in the UK. This study therefore focuses on assessed writing of international foundation students at a UK-based university and investigates the main contributing factors impacting the deployment of academic vocabulary in their written assignments. The aim of this longitudinal study, conducted over a period of one academic year, was addressed by the identification of individual academic expressions based on the Academic Word List (Coxhead 2010) and the more recent New Academic Vocabulary List (Gardner and Davis 2013) with the assistance of AntWordProfiler (Antony 2013). The identified academic items were subsequently analysed in their co-textual environments for grammatical accuracy using AntConc (Antony 2014). The textual analysis was complemented by individual semi-structured interviews identifying the students' perceptions of the main factors that impacted the usage of academic vocabulary in their assessed writing. The textual data point to the effect that the task and topic have on selection and usage of academic vocabulary. The interview data highlight the importance of feedback and exposure to academic expressions. These findings have potentially important pedagogical implications not only for international foundation practitioners, but also for broader EAP and other higher education contexts characterised by a diverse student population as they also highlight the difficulties encountered with certain aspects of grammar relating to academic vocabulary shared by a multilingual group of students.

References:

AntCon (Version 3.4.4) [Computer Software]. Developed by Anthony, L. (2014), Tokyo, Japan: Waseda University. Available
from http://www.laurenceanthony.net/software/antconc/ AntWordProfiler (Version 1.4) [Computer Software]. Developed by Anthony, L. (2013), Tokyo, Japan: Waseda University. Available from: http://www.laurenceanthony.net/software/antwordprofiler/

Coxhead, A. (2000). 'A New Academic Word List', *TESOL Quarterly*, 34(2), 213 – 238.

Gardner, D. and Davies, M. (2013). 'A New Academic Vocabulary List', *Applied Linguistics*, 31(4), 1 – 24.

Pronouns in CHET and CECheT: revealing authorial presence

Begoña Crespo, University of A Coruña
Isabel Moskowich, University of A Coruña

One of the many linguistic features in which authorial presence is assumed to be detected (Biber, 1988; Hyland, 1996; Atkinson, 1999) is the use of pronouns. First person pronouns are particularly important as they, as central pronouns (Quirk, 1985; Chamonikolasová, 1991) are especially revealing of the author`s presence and his/her involvement in the message conveyed. Stemming from the idea that scientific English is detached and object-centred (Atkinson, 1999) our aim is twofold: on the one hand, we want here to ascertain that this is so and whether this characteristic is maintained across time in general terms. On the other hand, being more specific, we will compare the results obtained in both the History and Chemistry texts compiled for the Coruña Corpus of English Scientific Writing. The reason for this is that we are conscious that they represent different writing traditions, the soft and the hard sciences in Hyland's words (1996), and want to show the differences. We will also consider other variables that may exert some kind of influence on the use of pronouns such as the sex of the author. The material we will be using in this study forms part, as already mentioned, of the so-called Coruña Corpus. Two subcorpora have been selected on this occasion, namely, The Corpus of History English Texts (CHET) and the Corpus of English Chemistry Texts (CECheT). They contain text extracts belonging to works published in the eighteenth and nineteenth centuries. Both of them are still beta versions at the moment of writing this abstract. Some of the characteristics of these subcorpora will be also described (Crespo and Moskowich, 2010).

References:

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Chamonikolasová, J. (1991). "Pronouns, functional sentence perspective and imntonation" in Brno Studies in English, 19, 55-64.

Crespo, B. and Moskowich, I. (2010). "CETA in the Context of the Coruña Corpus". *Literary and Linguistic Computing* 25/2: 153-164.

Hyland, K. (1996). "Talking to the Academy: Forms of Hedging in Science Research Articles" in *Written Communication*, 13/2, 251–281.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985). *A Comprehensive English Grammar*. London: Longman.

Where are they from? Geographical representation of migrants before and after Brexit

Dario Del Fante, University of Padova

The unforeseeable Brexit result would arguably constitute a watershed for EU history. However, Brexit is not only an isolated case, but the evidence of a bigger phenomenon, namely the upsurge of populist political tendencies among European countries. Their political campaigns have transformed peoples' concerns about the increasing number of terroristic attacks and of immigrants arriving in Europe into real fear, often depicting migrants as related to terrorism or as actual terrorists (Wodak and Krzyżanowski 2017). Within this context, the press plays a decisive role in reporting and commenting social and political reality (Fairclough 2001). Hence, this study analyses how Italian and British press reported migration discourse before and after Brexit, focusing on referential and predicational strategies, which are generally used for categorising social actors, attributing negative and positive characteristics in implicit or explicit predicates. (Reisigl and Wodak 2001: 45-46). These strategies aim at achieving emotive effects and so contribute to emotive coercion and cognitive coercion (Hart 2010). This study, in line with SFL (Halliday and Matthiessen 2004), falls in the tradition of Corpus Assisted Discourse Study (Partington et al. 2013), which combines the quantitative approach of Corpus Linguistics to the qualitative approach of Discourse analysis. The data set consists of two different corpora, one for the UK and one for Italy, composed of articles from different newspapers in order to be politically balanced, in the period January-July 2016 and January-July 2017. Preliminary results show a prevalent negative representation of nationalities investigated both in UK and in Italy, as for 2016 and for 2017. There is a decrease of attention given to migrants in 2017 than 2016.

References:

Fairclough, N. (2001). *Language and Power*. London: Longman.

Halliday, M. A. K., and Matthiessen, C. M. (2004). *An Introduction to Functional Grammar*. Third Edition. London: Hodder Arnold.

Hart, C. (2010). *Critical discourse analysis and cognitive science: New perspectives on immigration discourse*. Springer.

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies* (CADS) (Vol. 55). John Benjamins Publishing.

Reisigl, M., & Wodak, R. (2001). *Discourse and discrimination: Rhetorics of racism and antisemitism*. Routledge.

Taylor, C. (2014). Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International journal of corpus linguistics*, 19(3), 368-400

Wodak, R., & Krzyżanowski, M. (2017). Right-wing populism in Europe & USA. *Journal of Language and Politics*, *16*(4), 471-484.

Between an /r/ and a schwa: Illustrating some of the joys and woes of big data for speech analysis

Alexandra Vella, University of Malta
Sarah Grech, University of Malta

Research is beginning to show that examining phonetic variables such as "r" and schwa is a useful way forward, in attempts at establishing the degree of identifiability of Maltese English (and possibly other varieties of English) as a variety in its own right. Research on variation at the phonetic/phonological levels often needs to go beyond identifying and analysing features of variation in a given data set. Related variables such as the provenance of a feature or a set of features, the prevalence of particular variants across speakers, or, as in the case of post-vocalic "r" and schwa in Maltese English, the degree of interaction of different features within a given speech community are all key factors in understanding how such variation carries meaning in a society. In research on this type of variation, being caught between a rock and a hard place is not uncommon, as researchers need big data involving (ideally) natural speech to achieve statistically robust conclusions, but also manageable data that must often be heavily pre-processed before analysis can even begin. These restrictions can result in some features being more attractive as a research prospect simply because their distribution is widespread, while other features equally relevant as variables to be studied, are less readily researched because their occurrence is narrower. This paper presents an analysis of post-vocalic "r" and schwa in a relatively small Maltese English corpus of one hour of speech. Post-vocalic "r" is expected to be more difficult to analyse, as its distribution must be carefully balanced in order to obtain useful results. By contrast, schwa can be predicted to be slightly easier to analyse, even though this feature may also respond differently under particular conditions. For each feature, a total of 2 minutes of speech is transcribed and segmented. The analysis of both features is discussed both in relation to their occurrence within the data, and in relation to other sociolinguistic variables pertinent to Maltese English. The discussion is also framed by a consideration of some of the issues relevant to working with corpora of natural speech data.

A Corpus-based Examination of the Vocabulary Used in Medical Multiple Choice Questions

Emily Harms, St. George's University

A Corpus-based Examination of the Vocabulary Used in Medical MCQs Research has shown that the language used in multiple choice questions (MCQs) in nursing assessments negatively affects the performance of English language learners (ELLs) and negatively impacts pass rates for ELLs (Lujan, 2008). This demonstrates that high-stakes exams may have construct-irrelevant variance and inaccurately assess students' knowledge (Bosher & Bowles, 2008). While some research has been done in the field of nursing, limited research has been done in the medical field. Little is known about the language used in medical MCQs or how it affects ELLs' performance. With a growing number of ELLs in medical school (Hamilton & Woodward-Kron, 2010) and the reliance on MCQs in high-stakes licensure examinations, this issue is becoming increasingly relevant. This paper will present the literature review and methodology for a proposed dissertation using corpus linguistics to examine the vocabulary used in medical MCQs and how the vocabulary affects ELLs' performance. The researcher plans to use three established word lists to determine the frequency of vocabulary in a medical MCQ textbook (the Academic Vocabulary List, the Medical Academic Vocabulary List, and the New General Service List) and then examine the remaining vocabulary. The researcher will also focus on possible construct-irrelevant variance of MCQs based on the types and frequency of the language used. Through awareness of the frequency and types of vocabulary in MCQs, English support programs for ELLs in medical school can make more informed curricula choices to better support ELLs' success on high-stakes medical examinations.

References:

Bosher, S., & Bowles, M. (2008). Nursing education research: The effects of linguistic modification on ESL students' comprehension of nursing course test items. *Nursing Education Perspectives*, 29(3), 165-172.

Hamilton, J., & Woodward-Kron, R. (2010). Developing cultural awareness and intercultural communication through multimedia: A case study from medicine and the health sciences. *System*, 38(4), 560-568. doi:10.1016/j.system.2010.09.015

Lujan, J. (2008). Linguistic and cultural adaptation needs of Mexican American nursing students related to multiple-choice tests. *Journal of Nursing Education*, 47(7), 327-330. doi:10.3928/01484834-20080701-07

Evidence of nesting in metaphoric language

Katie J. Patterson, Universidad Austral de Chile

Metaphoricity is often regarded as a distinctive linguistic phenomenon, in opposition to literal, or non-figurative language. Recent research from a corpus-linguistic perspective has begun to show, however, that such a dichotomist stance to metaphor does not bear scrutiny (Deignan 2005; Partington 2006; Philip 2011). Our ability to manipulate or bend the limits of linguistic conventions (semantically, lexically, grammatically), in order to cope with communicative demands is one area where this dichotomy does not hold up. The focus of this chapter is to explore a nesting (cf. Hoey: 2005) pattern of grew that is specific to its use in metaphoric contexts, and compare this to its absence in non-metaphoric contexts. The data are taken from a 49m-word corpus of nineteenth century writings. The findings go some way to suggesting that as a metaphor, grew is qualitatively a different lexical item, when compared to its non-metaphoric use(s). It is proposed that Hoey's (2005) Drinking Problem Hypothesis can account for these lexical differences, providing a psychological explanation for what drives us as language users to identify metaphor. Crucially, adopting lexical priming as a means to exploring metaphor shifts the perspective of metaphoricity to the individual language user: the findings show that a metaphoric sense of an item appears to be dependent on the primings activated in a reader. It can thus be argued, based upon the lexical priming approach, that metaphoricity is inherent in the language user rather than the language itself, and that its manifestation is often dependent on the individual's interpretation of the language.

References:

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

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language.* London: Routledge.

Partington, A. 2006. 'Metaphors, motifs and similes across discourse types: Corpus-assisted discourse studies (CADS) at work'. In: A. Stefanowitsch & S. Gries (eds.) *Corpus-Based Approaches to Metaphor and Metonymy.* Berlin: Mouton de Gruyter, pp. 267-304.

Philip, G. 2011. *Colouring Meaning: Collocation and Connotation in Figurative Language.* Amsterdam: John Benjamins.

Unveiling the journalist's stance: Focus on The Economist's coverage of the Brexit referendum

Roberta Facchinetti, University of Verona
Sara Corrizzato, University of Verona
Valeria Franceschi, University of Verona

Despite being viewed as among the most objective textual types within the journalistic genre, news reports and news features may (and often do) include relevant information about the journalist's stance (van Krieken & Sanders 2016, Orna-Montesinos 2015, Perrin 2012, among others), to the point that Richardson (2006) suggests approaching news discourse as an argumentative genre. Bearing this in mind, the present paper will investigate the reports and features dealing with the Brexit referendum that was held in the United Kingdom on June 23rd 2016. To carry out the analysis, we have compiled a corpus of the entire coverage of the Brexit issue as reported by the online edition of The Economist between January and December 2016, thus covering the six months before and after the referendum. Working within the framework of both Corpus Linguistics and Critical Discourse Analysis, the study intends to identify if and to what extent the journalist's stance transpires from the texts. Stance-taking strategies embedded in the selected data will be analysed, with special attention to the following: lexical choices (including foregrounding and backgrounding), modality, textual structure and, from a wider perspective, the rhetorical construction of the whole piece. The data confirm the presence of an 'angle' transpiring from the majority of texts analysed, which convey a multiplicity of obvious or hidden communicative purposes that go far beyond the sheer reporting, thus and testifying to the strong role the journalist's slant may play in shaping the news narrative.

References:

Orna-Montesinos, C. (2015) "The discourses of sustainability in news magazines The rhetorical construction of journalistic stance." *Revista Espanola De Linguistica Aplicada* 28/2: 442-464.

Perrin, D. (2012) "Stancing: Strategies of entextualizing stance in newswriting." *Discourse, Context & Media* 1/2-3: 135-147.

Richardson, J. E. (2006) *Analysing Newspapers. An Approach from Critical Discourse Analysis*. Oxford: Palgrave Macmillan.

van Krieken, K. & Sanders, J. (2016) "Diachronic changes in forms and functions of reported discourse in news narratives." *Journal of Pragmatics* 91: 45-59.

Varieties of English in English Language Teaching in Germany

Lisa Scheiwe, Paderborn University

A sound knowledge of different varieties of English is vital for English as a foreign language students to improve their ability to communicate successfully in intercultural contexts. Thus, this paper addresses in how far German school books of English use representative recordings of different varieties of English. In a pilot study, I compiled a corpus of audio material of two school books for German grammar schools aimed at students aged 12 to 14. With a focus on two local varieties, Welsh English (WE) and New York City English (NYCE), I analysed selected phonological features (e.g. the FACE vowel for WE and the PRICE vowel for NYCE) auditorily, as well as acoustically. I compared these results to linguistic descriptions of these varieties (e.g. Wells, 1982; Gordon 2004) as well as the reference varieties RP and General American (GA) to determine to which extent the local varieties WE and NYCE were represented in the school book recordings. The analysis showed that there is a stronger tendency towards the pronunciation of the reference varieties RP and GA rather than the local varieties in both school books. I will discuss how representative these school book recordings of local varieties are.

References:

Gordon, Matthew J. (2004). "New York, Philadelphia, and Other Northern Cities: Phonology". In: *A Handbook of Varieties of English Phonology*. Ed. by Edgar W. Schneider et al. Vol. 1. A Handbook of Varieties of English 1. Berlin: Mouton de Gruyter, pp. 282–299.

Wells, John C. (1982c). *The British Isles. Vol. 2. Accents of English*. Cambridge: Cambridge University Press.

Corpus-driven Approach to Lexical Semantic Change: A Case Study

Gang Yao, University of Murcia &
Shanghai International Studies University

Research on the change of different linguistic aspects has a long-standing tradition, whereas studies of lexical semantic change have moved inside our field of vision only after the last century because of lexical irregularities and technical limitations. Seeking to understand and explain why and how semantic change happens, we chose our words of interest—English *migrant* and its Spanish counterpart *migrante*—to investigate the issue, which was creatively approached from a corpus-driven perspective with the assistance of dictionary resources. Analysis of the frequency information, collocational profile, and statistic tests of Eng. *migrant* and Sp. *migrante* in historical corpora has revealed that both words have undergone a meaning change in their respective histories and illustrated a close connection behind these changes. Based on a qualitative interpretation of the corpus result, socio-historical and linguistic factors, accompanied by mechanisms of semantic change such as metaphor and language borrowing, are involved in accounting for the respective cases. The findings and interpretations presented in the present study can not only shed new light on the causes, mechanisms, and consequences of lexical semantic change, but also test the possibility of implementing an innovative approach to this subject.

Keywords: lexical semantic change, corpus-driven approach, causes, mechanisms, *migrant*, *migrante*

NOT THE 99 PER CENT. (Under-)representation of the less wealthy in British fiction corpora

Michael Pace-Sigge, University of Eastern Finland

The issue of power and the use of language has been widely researched (for example: Bernstein, 1968; Fairclough, 1989), as has the issue of power-relations and control (cf. Bordieu, 1991; Partington, Taylor & Duguid, 2013). A corpus-based piece of research, which looks at the frequency, collocations and semantic associations of words that express the presence or absence of power and economic descriptors, fits therefore into a tradition of exploring how the world is represented through the written word. Indeed, "[c]orpus stylistics can contribute to the exploration and development of descriptive tools that aim to characterize meanings in texts" (Mahlberg, 2007: 240).

This paper seeks to determine the extent to which the lexical choices in British literature over the last 200 years support the view that *homo economicus* is reflected in fictional writing – and how the world of work finds representation in novels. For this, a corpus of 500 full-text Victorian novels, representing the 19th century; and the material in the *British National Corpus: Written Fiction* (BNC_W-F) representing the 20th century, will be investigated to see how key words express the writer's relationship to wealth and earning an income.

This research has been inspired by Karl Marx's *Das Kapital* and Thomas Piketty's book *Capital in the Twenty-First Century* where the authors, unusual for economists, make frequent reference to works of literature.

References:

Bernstein, B. 1973. *Class Codes and Control.* St. Albans: Paladin.

Bordieu, P. 1991. *Language and Symbolic Power*. Cambridge: Polity Press / Blackwell

Fairclough, N. 1989. *Language and Power*. Harlow: Longman.

Mahlberg, M. 2007. Corpus stylistics: Bridging the gap between linguistic and literary studies. In *Text, Discourse and Corpora*, M. Hoey, M. Mahlberg, M Stubbs & W. Teubert (eds),pp. 219-246. London: Continuum.

Marx, Karl. [1859] Das Kapital: Kritik der politischen Oekonomie. See: http://www.mlwerke.de/me/me13/me13_003.htm (last accessed 18/09/17).

Partington, A., Alison Duguid, and Charlotte Taylor, 2013. *Patterns and Meanings in Discourse*. Amsterdam: John Benjamins.

Piketty, T. 2014. *Capital in the twenty-first century*. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Some methodological challenges in the identification of discriminatory attitudes and counter-speech in online newspaper comments

Stavros Assimakopoulos, University of Malta
Rebecca Vella Muskat, University of Malta

While a lot of research in the area of critical discourse analysis has focused on the representation of minorities in the mass media, a surprisingly limited amount of work has dealt with the very same topic in the social media context. As Törnberg and Törnberg acknowledge, "a central reason for this striking lack of studies is the methodological difficulties relating to handling and analysing the large amounts of unstructured textual data that often characterise social media" (2016: 132). In this talk, we will focus on the challenges we faced while classifying such data as positive or negative in terms of their evaluative content in relation to migrants and LGBTQI individuals. The corpus at hand, which was compiled for the purposes of the EU-funded CONTACT project (http://reportinghate.eu), comprises comments posted in reaction to news items focusing on the two communities in Maltese online news venues (cf. Assimakopoulos and Vella Muskat 2017). While annotating our data in a way that would allow for eventual analysis using corpus linguistics techniques, it soon became obvious that the seemingly simple binary classification that we had started off with would not do justice to the complex nature of our collected comments, and that a more nuanced approach would be needed in order to identify both discriminatory attitudes and counter-speech in this context. Against this backdrop, and using examples from our corpus and its analysis, we will discuss in our talk the challenges posed by: i) the need to control for comment size and organisation while sampling data for in-depth annotation, ii) the fluidity of evaluative attitudes using both direct and indirect strategies even within single comments, iii) the existence of a substantial number of ambiguous comments that cannot be straightforwardly classified, and iv) the presence of code-switching in the particular bilingual context of Malta.

References:

Assimakopoulos, Stavros and Rebecca Vella Muskat. 2017. Xenophobic and homophobic attitudes: National report for Malta. Msida: University of Malta.

Törnberg, Anton and Petter Törnberg. 2016. Muslims in social media discourse: Combining topic modeling and critical discourse analysis. Discourse, Context & Media 13: 132-142.

Something borrowed and something new: a dual approach for studying inter-varietal differences multilingually

Ilmari Ivaska, University of Bologna
Silvia Bernardini, University of Bologna
Adriano Ferraresi, University of Bologna

Today's global mobility has contributed to the switch to new forms of superdiverse multilingual realities. Consequently, in order to understand what is typical for a certain variety in a certain language and on the other hand across languages, one has to tease apart language-specific tendencies from more general ones. Corpora are precious resources when doing such interpretations, but they tend to be limited in comparability across languages. On the other hand, compiling equally representative corpora in multiple languages is hardly feasible. We suggest a dual approach to address the problem: available corpora and quantitative corpus-driven methods can be used as a point-of-departure for the inquiry, so as to detect potential signals for varietal differences parallelly in multiple languages. These findings can then be tested by compiling a smaller collection of new, rigorously documented and highly comparable data in all the studied varieties and languages. In our exemplary study, we contrast learner language and translated language with non-translated first language in English, Finnish, and Italian. We use multiple existing corpora, a cross-linguistically consistent annotation scheme of the Universal Dependencies Initiative, and a corpus-driven method, i.e. Key Structure Analysis, to find linguistic features that distinguish these varieties from each other, and compare the findings between languages. We also collect highly comparable texts in all these varieties and languages, and complement them with experimental surveys and interview data from the same participants, structured based on the findings in the large-scale corpora. Our preliminary findings support the applicability of the dual approach: we were able to identify both language-specific and general inter-varietal differences. Interestingly, we also found evidence that the studied varieties – learner language translated language – often diverge very similarly from the reference variety – non-translated first language. In a first stage, these findings are interpreted through exploratory collostructional searches in the corpus. Subsequently, our well-curated data will shed more light on them and also on the way the constructions involved are used, perceived, and thought of, by the respective language user groups. The dual approach could be generalized to research designs of any language varieties where sufficiently comparable existing corpora are not available.

An analysis of a corpus of Teacher Talk in the secondary-level EFL classroom: highlighting differences in modality and in the use of phrasal verbs

Eric Nicaise, Université catholique Louvain/ University College Louvain-en-Hainaut

Teacher Talk (TT) is a key aspect of English as a foreign language (EFL) lessons. Setting aside the wide range of listening materials available, the EFL teacher is in many instances the only model of English which students are exposed to throughout their time at secondary school. According to research, 'teacher talking time' takes up an impressive 70% of classroom time in general (Cook, 2008; Meunier, 2012). The talk will focus on two linguistic features of Teacher Talk used at secondary-level education in EFL classes: phrasal verbs and modal forms. In particular, it will investigate whether the use of these two linguistic features differs according to the two following contexts: English language classes where the teachers teach their native language, and EFL classes in the French-speaking part of Belgium, where francophone EFL teachers teach English. This investigation has been made possible through the exploration of the CONNEcT Corpus (an acronym for A Corpus of Native and Non-native EFL Classroom Teacher Talk). While still in the process of being expanded, the Corpus now amounts to over 200, 000 words. After discussing the motives and rationale behind this study, the corpus linguistics and CADS approaches which were adopted to examine the use of phrasal verbs and modal forms in the teacher's classroom speech will be presented. Following a quantitative analysis, both language areas will then be investigated within the context of two major classroom-based functions: language explanation and instruction giving. The findings reveal considerable variation between the two sub-corpora of teacher talk: while the francophone English teachers tend to use their teacher talk to comment on these two language areas, the teachers whose mother tongue is English tend to incorporate them into their classroom speech as part of the target language that the students are learning. Drawing on these findings, some implications for EFL teaching will be considered.

References:

Cook, V. (2008) *Second Language Learning and Language Teaching*. (2nd edition). Beijing: Foreign Language Teaching and Research Press.

Meunier, M. (2012) 'Formulaic language and second /foreign language teaching', in *Annual Review of Applied Linguistics* ,Vol. 32, pp. 1-34.

The Trojan Horse: the Construction of Iraqi Shiites in Times of Wars in the US Press

Dhiaa Kareem Al Janaby, Newcastle University

The Iraqi Shiites are consistently viewed as a disadvantaged and repressed group in spite of their large population, which makes up 60-65% of the total population when compared to their counterpart Sunni Muslims, who make up much of Iraq's elite in the Iraqi government. 'As a result of their repression and suffering, Iraqi Shiites were always seen by the opponents Iraq faced in its wars as a Trojan horse that could revolt and overthrow Saddam's regime from the inside. Thus, the Shiites were targeted by Iran in the Iraq-Iran war of 1980- 1988, the Iraqi invasion of Kuwait in 1990 and in the 2003 US-led invasion in a series of discourses aimed at mobilizing them against Saddam. The aim of this study was to investigate the construction of the Iraqi Shiites in the major US press in times of wars and conflicts, namely: the Iraq-Iran war and the US-led invasion, to see how they are represented and whether there is a change in the reporting of this group. To achieve this aim, a synergy of corpus linguistics using a critical discourse analysis approach represented by the Discourse-Historical Approach (DHA) was employed. The results showed that in both wars (Iraq-Iran and the US-led Invasion) the Iraqi Shiites were constructed as being misrepresented in their political representation in the Iraqi government. However, this construction was weighted more heavily in 2003. Also, the Shiites were constructed as being unmoved by Iranian calls to revolt owing to the differences between Iraqi Shiites and those of Iranian Furthermore, the accumulated history of repression of the Shiites in earlier wars against them was seen as a case of human rights violation in 2003, whereas this aspect was not highlighted in the press during the Iraq-Iran war.

A Corpus-Based Analysis of Social Interaction in LTE Contexts

Fiona Farr, University of Limerick
Angela Farrell, University of Limerick
Elaine Riordan, University of Limerick

This paper examines social interaction in language teacher education contexts, namely teaching practice feedback sessions, classroom discussions, reflective practice discussions, and online teaching portfolio entries. We believe there is a need to start developing evidence-based accounts of teacher education interactions to assist those working in the profession (as teachers, researchers and educators) as it is still the case that 'very little substantial research has been undertaken in teacher education, both in education generally and in the field of language teaching' (Freeman 2001:74). In response to this research gap, we draw on a range of spoken and written data collected from these contexts differentiated in the following ways: experientially, using teachers along the cline from novice to experienced; educationally, examining pre- and in-service education programmes; by mode of interaction, face-to-face (tutorials, TP feedback, lectures) and online modes of computer-mediated communication (blogs, discussion forums, chat, and e-portfolios); and by interactional relationship paradigms (student teachers with each other, with peer tutors, with practice tutors and with lecturers).

Through the use of quantitative and qualitative corpus techniques to investigate the data, we examine the ways in which various types of social interaction involving different participant paradigms and distinct modes can either facilitate processes of teacher development or pose problems for teachers. This paper closes by considering how the findings presented can feed into a better understanding of professional practices and develop new practice-oriented and research agendas.

References:

Freeman, D. (2001) 'Second language teacher education', in R. Carter and D. Nunan (eds), *The Cambridge Guide to Teaching English to Speakers of Other Languages*, Cambridge: Cambridge University Press, 72-79.

Freeman, D. and Johnson, K. E. (1998) 'Reconceptualizing the knowledge-base of language teacher education', *TESOL Quarterly*, 32(3): 397-417.

This is England, speak English! A corpus-assisted critical study of language ideologies in the British press

Gavin Brookes, University of Nottingham
David Wright, Nottingham Trent University

This talk will explore British press representations of people living in the UK who can't speak English following the 2011 Census, which was the first to ask questions about respondents' linguistic background and competence in English. This study takes a corpus-assisted approach to critical discourse analysis (Baker et al., 2008) and is based on a corpus of right-leaning national newspaper articles published between 2011 and 2016. The analysis evinces the tendency for the press to focus on immigrants (particularly in school and healthcare contexts) and construct these groups with recourse to four recurring representational themes, which we relate to argumentational strategies or topoi (Reisigl and Wodak, 2001). It is argued that these representational strategies (and their associated topoi) are problematic as they present a series paradoxes, overlook the difficulties of language learning and cultural assimilation, and generally contribute to a broader UK anti-immigrant media narrative which serves to legitimise exclusionary and discriminatory practices against people from minority linguistic and ethnic backgrounds.

References:

Baker, P.,Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery, T. and Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306.

Reisigl, M. and Wodak, R. (2001). *Discourse and Discrimination: Rhetorics of Racism and Antisemitism*. London and New York: Routledge.

'I hear whispers and I don't like it': Conflict in a Corpus of Spoken Interactions from the Construction Industry

Almut Koester, Vienna University of Economics and Business
Michael Handford, Cardiff University

Communication in the construction industry has been characterized as 'adversarial' (Emmett and Gorse, 2003), but few studies have examined actual instances of interaction in the industry. Some recent studies exploring naturally-occurring construction interactions using discourse analysis and corpus methods (Handford, 2014; Handford and Matous 2011, 2015) have found mixed results, identifying some conflictual but also many consensual interactions. The first part of the paper will introduce CONIC (Construction Industry Corpus, JSPS Grant number 00466781), a 300,000-word multimodal, contextually-rich corpus of construction industry interactions. The corpus comprises audio and video recordings of meetings, on-site interactions and problem-solving encounters from the UK and several Asian countries (including Vietnam, Singapore and Bangladesh), as well as interview data and written documents. Most of the interactions involve speakers using English as a Lingua Franca. Using a combination of corpus methods and discourse analysis, the paper then examines the corpus to identify the characteristics of conflictual versus collaborative discourse within this community of practice. We will also address the theoretical question of how conflictual discourse can be characterised: whether there are any universals, or if it is the context alone which determines what is understood as collaborative or conflictual.

Research areas addressed by the paper:
• CL and real word contexts (workplace discourse)
• CL and English as a Lingua Franca (ELF)
• Spoken Corpora

References:

Handford, M. 2014. Communication in the construction industry. In: *The Routledge handbook of language and professional communication*. London: Routledge, pp. 363-381.

Handford, M. and Matous, P. 2011. Lexicogrammar in the international construction industry: a corpus-based case study of Japanese-Hong-Kongese on-site interactions in English. *English for Specific Purposes* 30(2), pp. 87-100.

Handford, M. and Matous, P. 2015. Problem-solving discourse on an international construction site: Patterns and practices. *English for Specific Purposes* 38, pp. 85-98.

Where is the I? Cultural Diversity and the Self in L2 Student Writing

Jelena Runić, Johns Hopkins University

This presentation discusses problems raised by cultural differences with respect to self-narration and use of the first person in academic writing, by exploring a large sample of student essays. Western culture most commonly views events from one's own perspective (independent self), while Asian cultures typically position the self as part of the event (interdependent self) (cf. Markus and Kitayama 1991). The question that should be asked then is how multilingual writers from Asian cultures approach western academic prose and self-narration, especially in the context of increased student mobility.

That the concept of *individuality* varies across cultures has been supported even by anecdotal evidence. Thus, Markus and Kitayama (1991:224) report that parents persuading children to eat do so differently in the U.S. and in Japan. To illustrate, an American parent would say: "Think of the starving kids in Ethiopia, and appreciate how lucky you are to be different from them." A Japanese parent would react differently by saying: "Think about the farmer who worked so hard to produce this rice for you; if you don't eat it, he will feel bad, for his efforts will have been in vain." Furthermore, research on differences in self-construction from the field of cultural psychology demonstrates that Asians, Asian-born immigrants, and Asian Americans are inclined to tell different self-narratives than European Americans (Wang 2008, 2009, *i.a.*)

This presentation explores original student essays and their strategies for self-representation of western prose. In the first part of the presentation, the audience will be shown portions of autobiographical student texts. Strategies for the inclusion of the interdependent self in western writing style will be revealed. Thus, samples of low-stake assignments emphasizing the writer's voice and analytical writing will be demonstrated. Finally, the implications of this type of self-narration will be addressed, key among which being textual parsing, analytical writing, writer's voice, writer's contribution, and intellectual property.

References:

Markus, Hazel Rose, and Shinobu Kitayama (1991). Culture and the Self: Implications for Cognition, Emotion, and Motivation. Psychological Review 98, 2: 224-253.

Wang, Qi (2008). Being American, Being Asian: The Bicultural Self and Autobiographical Memory in Asian Americans. *Cognition* 107, 2: 743-751.

Wang, Qi (2009). Are Asians Forgetful? Perception, Retention, and Recall in Episodic Remembering. *Cognition* 3, 1: 123-131.

The design and annotation of the TEG learner corpus of Irish

Aoife Ní Ghloinn, Maynooth University
Elaine Uí Dhonnchadha, Trinity College Dublin
Anne O' Keeffe, Mary Immaculate College

This paper will report on the compilation and annotation of the TEG learner corpus of Irish – 150,000 words of spoken and written learner data, produced as part of the TEG (Teastas Eorpach na Gaeilge/ European Certificate of Irish) proficiency tests in Irish, at B1-C1 level of the Common European Framework of Reference for Languages (CEFR). The aim of the current phase of the TEG corpus project is to analyse learner grammar in Irish at B1-C1 levels, in order to compile profiles of grammatical competence at those levels. The paper will focus particularly on the development and piloting of a semi-automatic error annotation system for this corpus, in order to facilitate retrieval and analysis of grammatical structures used correctly and incorrectly by learners at each proficiency level. While there is a growing body of literature on the annotation of learner corpora (van Rooy, 2015; Lüdeling and Hirschmann, 2015) and significant developments in the linguistic annotation of Irish (Uí Dhonnchadha and Van Genabith, 2006), the TEG corpus will be the first to bring together these two strands of research in order to provide comprehensive grammatical annotation of learner Irish. This paper will provide an overview of the various methods used in the grammatical annotation of learner corpora, including approaches to issues such as corpus architecture, the provision of an explicit target hypothesis, the categorisation of errors, and the tools used to facilitate the annotation process and analysis. It will then describe the methods and tools developed for this project, in order to provide an annotation system for learner Irish that is accurate, consistent and efficient, while enabling meaningful analysis.

References:

Van Rooy, B. 2015. Annotating learner corpora. In: Granger, S., Gilquin, G., & Meunier, F. (eds.). *The Cambridge Handbook of Learner Corpus Research*. Cambridge University Press.

Lüdeling, A. & Hirschmann, H. 2015. Error annotation systems. In: Granger, S., Gilquin, G., & Meunier, F. (eds.). *The Cambridge Handbook of Learner Corpus Research*. Cambridge University Press.

Uí Dhonnchadha, E & Van Genabith, J., .2006. A Part-of-Speech Tagger for Irish using Finite State Morphology and Constraint Grammar Disambiguation, LREC 2006, available at http://www.lrec-conf.org/proceedings/lrec2006/.

Using corpus linguistic methods to uncover verbal cues to deception

Mathew Gillings, Lancaster University, UK

Research from the field of psychology has shown that there are significant differences between the behaviours of a truth teller and a liar; investigating these differences in more detail may point towards cues to deception. These cues can be divided into three categories: physiological cues (such as heart rate and sweat); behavioural cues (such as body language and eye movement); and verbal cues (such as speech rhythm or grammatical differences).

This work-in-progress talk will focus purely on verbal cues, outlining how I am using corpus linguistic methods to investigate the language of deception. Drawing upon previous research and methods from psychology, I will outline my experiment which was designed to extract truthful and deceptive language in a range of scenarios, and from a range of participants from different social backgrounds (i.e. differences in region, gender, age, and socioeconomic status). The focus of this talk will be on how corpus methods can be used to analyse this language.

The bulk of this talk will comment on the relatively uncharted territory which lies between corpus linguistics and forensic psychology. Research from the field of psychology in automated deception detection has so far only been carried out using LIWC (Pennebaker et al, 2001). However, in more recent years, Archer and Lansley (2015) and McQuaid et al (2015) have applied corpus linguistic methods to the field, using Wmatrix to investigate part-of-speech and semantic differences between truthful and deceptive corpora. It is clear to corpus and forensic linguists that we have a lot more to offer psychology and the field of deception detection than we have done already, and these methods have the potential to uncover deceptive cues that have so far remained unfound.

References:

Archer, D. and C. Lansley. (2015). Public appeals, news interviews and crocodile tears: an argument for multi-channel analysis. *Corpora*. Vol. 10(2): 231-258.

McQuaid, S., M. Woodworth, E. Hutton, S. Porter, and L. ten Brinke. (2015). 'Automated insights: verbal cues to deception in real-life high-stakes lies'. *Psychology, Crime and Law*. Pp. 1-31.

Pennebaker, James & E. Francis, M & J. Booth, R. (2001). Linguistic Inquiry and Word Count (LIWC): LIWC2001. Mahwah: Lawrence Erlbaum Associates.

'Stoefpears Run the World': A Corpus Study into English Code-Mixing in Dutch Youths' Computer-Mediated Communication

Lieke Verheijen, Tilburg University
Roeland van Hout, Radboud University
Laura de Weger, Radboud University

English loanwords are quite common in Dutch, especially in advertising, commercials, and business communication, but also in everyday speech (Zenner, Speelman, & Geeraerts 2015), and English elements have become a salient aspect of communicative practices in Dutch youth language. Linguistic borrowing is often preceded by code-mixing (or code-switching), i.e. alternating between two or more languages within a single conversation or context. The present study aimed to explore Dutch youths' code-mixing in their written computer-mediated communication (CMC), specifically to what extent and how they code-mix with English in their online messages. We quantitatively and qualitatively analysed a large-scale corpus of Dutch written CMC. The corpus consisted of messages by youths of different genders and age groups (adolescents, young adults), and composed via different social media (SMS, Twitter, MSN chat, WhatsApp). It contained 8619 switches to English. Previous research (De Decker & Vandekerckhove 2012, Verheijen 2016) inspired us to examine the switches for various language-internal factors, namely their length, number, lexical category, integration, semantic field, intentionality, and frequency, as well as for the language-external factors of gender, age, and CMC mode. Though the youths revealed a certain level of creativity in their code-mixing via memes, puns, and graphemic and morphological integration, the English elements they used were mostly conversational words and fixed phrases such as greetings, affective language, swear words, and expressions. The results imply that Dutch youths mainly use English as part of their 'teenage talk': to boost their expressivity and distinguish themselves from older speakers.

References:

De Decker, B., & Vandekerckhove, R. (2012). English in Flemish adolescents' computer-mediated discourse: A corpus-based study. *English World-Wide*, 33(3), 321-351.

Verheijen, L. (2016). De macht van nieuwe media: hoe Nederlandse jongeren communiceren in sms'jes, chats en tweets. In: D. van de Mieroop, L. Buysse, R. Coesemans, & P. Gillaerts (Eds.), *De macht van de taal: Taalbeheersingsonderzoek in Nederland en Vlaanderen* (pp. 275-293). Leuven / Den Haag: Acco.

Zenner, E., Speelman, D., & Geeraerts, D. (2015). A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal of Bilingualism*, 19(3), 333-346.

Exploring the Use of Collocation in Saudi University Students' Writing Using Corpus Methodology

Huda Yahya Khoja, University of Leeds

English Foreign Language (EFL) learners, Saudi learners in this context, face many challenges in the process of learning and using the language in a native-like way. One of these difficulties is the correct use of collocation, where they know two words, but fail to connect them accurately. Even though there has been a continuously growing interest in investigating EFL learners' written production of collocation, there has been little research carried out in the context of Arab EFL learners. This was especially the case in studying the three types of collocations investigated in this study using corpus approaches. EFL studies involving Saudi Arabian learners have thus far focused on controlled environments through elicitation tests or instructional interventions. Motivated by the written production data, this research uses 16 written samples selected randomly from two levels of Saudi foundation-year students (pre-intermediate and intermediate) at King Abdulaziz University (KAU). The methodology follows three steps in analysing learners' written texts: the manual extraction of collocation, then referring to the native corpus, i.e. the BNC, and finally contrasting the two levels' use of collocation. The research findings show that Saudi learners are able to produce Established Collocations in a high number, without much difference between pre-intermediate and intermediate level learners. The findings have also reflected on the types of produced collocations and to which of the three types of collocations investigated were found to be the most and least problematic. This research further indicates the possible similarities and differences in the uses of the two levels of Saudi learners, despite learners' language proficiency not being consistent to relate completely for a better or lower use of collocations. Thus, the research was able to draw upon some problematic issues in those learners' use of collocation, which can contribute to pedagogy, whereas any in depth exploration of what may cause the production of Non-Established Collocations is still limited.

'That's the dilemma you know': Patterns and functions of *you know* as used by L1- and L2-speakers of English

Giovani Santos, Mary Immaculate College, University of Limerick

Studies on second language acquisition have shown that the immersive study-abroad experience has a very positive effect on spoken fluency, while offering opportunities to develop linguistic features found in the pragmatic system of the target language (Kinginger 2013). Within such a system, we find pragmatic markers (PMs), which are a core feature of spoken language and part of the toolkit for fluency in speaking (O'Keeffe et al. 2007). PMs have been widely studied within L1 contexts – including contrastive studies between different L1s – and have provided us with invaluable insights into the multi-functionality of such linguistic features. However, despite their pivotal role in natural and fluid conversation, there is still much ground to be covered regarding the form and function of PMs in L2 discourse. This paper, part of a broader and innovative PhD study, investigates the use of the PM you know in the spoken discourse of Brazilian university students in Ireland. The results and reflections of this study are based on a sample corpus of spoken language, which comprises four 30-minute informal interactions between 6 participants and the researcher. Drawing on the Limerick Corpus of Irish English (LCIE), and a sample of the LCIE 2.0 (an updated version of LCIE which is being currently built in a parallel project), it was possible to compare and contrast the patterns and functions of the PM you know in the participants' L2 against those of the Irish English data. Buysse (2017) notes that despite being one the most common PMs, you know, by comparison to other PMs, has not been given as much attention in L2 studies. This paper, thus, aims to address this gap and contributes to the growing body of studies on L2 pragmatics as well as on Irish English and its impact on L2-speakers of English in Ireland.

References:

Buysse, L. (2017) 'The pragmatic marker you know in learner Englishes', in *Journal of Pragmatics*, 121, 40-57.

Kinginger, C. (2013) 'Language socialization in study abroad', in Chapelle, C. A. (ed.) *The Encyclopedia of Applied Linguistics*, Oxford: Blackwell Publishing Ltd, 3163-3168.

O'Keeffe, A., McCarthy, M. and Carter, R. (2007) *From Corpus to Classroom: Language Use and Language Teaching*, Cambridge: Cambridge University Press

Peeling back the layers: using CL and CA to examine Focus Group talk

Róisín Ní Mhocháin, Bath Spa University

This paper looks at how Corpus Linguistics (CL) and Conversation Analysis (CA) can be combined to support an analysis of Focus Group discussion with English as a second language speakers. The discussion topic revolves around their experiences both learning English in school and using it as a necessary conversational resource as part of their daily lives in Ireland. The wider study examines the data in terms of three layers; linguistic (CL), structural (CA), and thematic. This paper looks at how Corpus Linguistics is used to identify some potential areas of interest for CA analysis of the data. CA is particularly relevant in this study because, grounded in ethnomethodology, it also aims to make sense of the ordinary, of which we can say talk is primary (Punch and Oancea 2014). Baxter (2010: 120) notes that analysing talk-in-interaction using CA 'provides extraordinarily rich evidence of the underlying rules of how language works'. The interest in this focus group data source in particular is to evidence how participants, with varying proficiency in English, individually contribute to and manage the discussion, and ensure their voices are clearly heard. The intention is to examine the focus group corpus data in light of the thematic findings to see if the difficulties participants say they have with speaking English is evident in the structural and linguistic data of the focus group interaction itself.

References:

Baxter, J. (2010) 'Discourse-analytic approaches to text and talk'. In Litosseliti, L. (ed.) Research Methods in Linguistics. London: Bloomsbury pp.117-137.

Punch, K. F. and Oancea, A. (2014) Introduction to research methods in education. 2nd edn. London: SAGE.

"Straight talking money" A corpus assisted critical discourse analysis of payday loans websites

Hazel Price, University of Huddersfield
Dan McIntyre, University of Huddersfield

As a result of the global economic crisis in 2008, payday loans rose in popularity in the UK. In recent years, however, the Financial Conduct Authority (FCA) has imposed a price cap on such short-term, high interest loans and introduced regulatory changes to protect customers against inappropriate advertising. In this paper we use Wmatrix (Rayson 2009) to analyse the linguistic strategies employed by payday lenders on their websites in the light of these changes. We compare the language of one specific payday lender, Wonga, across two periods (2014 and 2017) to track linguistic developments after the changes in legislation. We then compare payday loans websites (e.g. Wonga, QuickQuid) with the loan webpages of UK banks (e.g. NatWest, HSBC) to assess how the two types of loan providers interact with their customers. Our initial results show register differences between Wonga and major banks, through Wonga's significant overuse of contractions (e.g. -n't, -'ll) and first person plural pronouns. Semantic analysis shows that banks significantly overuse lexis pertaining to the domains of LAW AND ORDER, SUITABILITY, and HELPING. Compared to banks, semantic categories overused in the Wonga corpora include MONEY, TIME: LATE, and STRONG OBLIGATION OR NECESSITY. The language used by Wonga indicates a focus on paying loans back instead of explaining available loans and their suitability for customers. Research has shown that payday loan customers are typically young, single and on very low incomes. In contrast to how Wonga advertise their loans (as being for one-off, unexpected items), customers report using their borrowed money for everyday essentials such as groceries (Citizens Advice Bureau, 2016). Consequently, we argue that Wonga's linguistic practices are inappropriate for their customer base (even in the light of changes to their advertising) since they have the potential to negatively affect social mobility by trapping borrowers in a spiral of debt.

References:

Citizens Advice Bureau (2016) 'Payday loans: An improved market?'
<http://bit.ly/1ThCDOZ>

Rayson P (2009) Wmatrix: A Web-based Corpus Processing Environment. Lancaster University. <http://ucrel.lancs.ac.uk/wmatrix/>

The Relationship between Phraseology and Writing Quality: Implications for the Assessment of Student Writing in an English as a Lingua Franca Context

Lee McCallum, University of Exeter

The relationship between phraseology and second language proficiency judgements has been increasing in popularity with observations from Howarth (1998) and more recent computational studies from Crossley, Cai and McNamara (2012).

Under the dimensions of complexity, accuracy and fluency, these studies have unquestionably shown phraseology is a key consideration for examiners when making proficiency judgements (Kyle & Crossley, 2015). Yet, many of these studies have focused on 'large'- grained measures that are simply extracted by automatic tools such as the number of bigrams and their presence in native reference corpora. In doing so, these studies emphasise language considered accurate under native speaker production. However, pluralistic views of proficiency that operate in English as a Lingua Franca contexts, involve the ability of non-native speakers to communicate in situations where deviations from 'native speaker norms' are commonly present. These deviations often involve atypical phraseological units that are considered erroneous under traditional SLA theories including: '*make a research*' and 'cause *profits*' which are not naturally produced in native texts.

This study uses a corpus of graded placement test essays from learners in the UAE to investigate the relationship between typical native and atypical non-native phraseological combinations that are operationalised under CAF dimensions. The study differs from much existing research by investigating phraseology across a wide grade range (CEFR A2-B2) and featuring measures that tap into learner and reference-corpus based measures. The study's research design uses MI scores to explore and isolate atypical features and shed light on their relationship with writing grades.

References:

Crossley, S.A., Cai, Z., & McNamara, D.S. (2012). Syntagmatic, paradigmatic and automatic n-gram approaches to assessing essay quality. Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference (pp.214-219).

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, *19* (1), 24-44.

Kyle, K., & Crossley, S.A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings and application. *TESOL Quarterly*, *49*(4), 757-786.

A Corpus-Aided Comparative Analysis of Native and Non-Native Speaking English Language Teachers' Approaches to In-Class Spoken Feedback

Jane Seely, Mary Immaculate College, University of Limerick

Classroom Discourse (CD) and Teacher Talk (TT) have received much attention over the years across a range of research perspectives. Much of recent CD work draws on Conversation Analysis (CA) and, more recently, there has been growing synergy between CA and Corpus Linguistics (CL). In the area of Language Teacher Education (e.g. CELTA), trainee teachers are introduced to the concept of TT as something to be minimised and very little mention is made of the types of TT used in the classroom. In other words, the focus is on reducing the quantity of TT rather than on the nature of the talk itself. This paper will report on a quasi-longitudinal mixed-methods study of TT which uses data from classroom interactions over 9 months and also draws on interviews with the teachers. This 150,000 word corpus comprises 15 Native (NS) and Non-Native English (NNS) speaking teachers, at varying career stages, ranging in age from 23 to 36. Specifically, this paper will compare and discuss findings on teachers' spoken feedback in the classroom in relation to the Native English-Speaking teachers and their Non-Native English Speaking counterparts. The results indicate that, regardless of career stage, the NNS teachers are more direct in their approach to corrective feedback, with more examples of direct repair and form-focused feedback than the NS teachers. Additionally, while the majority of teachers interviewed stated that they were robust in their positive feedback, the corpus data showed a disparity between NS and NNS feedback whereby the NNS teachers seemed to show more positive reinforcement than their NS counterparts. This brings to light the usefulness of using corpus data in teacher training, in a mixed methods format, not least of all to highlight where there is disparity between teacher beliefs about TT and how they actually use it in the classroom context.

Corpus design and construction: the challenges faced by minoritized languages

Dawn Knight, Cardiff University

This paper provides a detailed overview of the ESRC/AHRC -funded CorCenCC (Corpws Cenedlaethol Cymraeg Cyfoes - The National Corpus of Contemporary Welsh) project: A community driven approach to linguistic corpus construction. CorCenCC will be the very first large-scale general corpus of Welsh language, comprising 10 million words of spoken, written and electronic language. It will be the first corpus to represent modern Welsh and will be revolutionary in that it is community-driven, using mobile and digital technologies to enable public collaboration. CorCenCC will be open-source and freely available for use by professional communities and anyone with an interest in language and will help to construct a picture of how modern Welsh is evolving and will contribute to the debate on the future direction of the language. CorCenCC will enable, for example, community users to investigate dialect variation or idiosyncrasies of their own language use; professional users to profile texts for readability or develop digital language tools; to learn from real life models of Welsh; and researchers to investigate patterns of language use and change. Corpus design and construction in a minority language context such as that of Welsh poses interesting challenges, but also presents opportunities perhaps not open to developers of corpora for larger languages. During this presentation, I will discuss and examine these challenges and opportunities in more detail. The presentation will then look at some of the potential applications of a corpus of contemporary language in the context of a minoritised language such as Welsh.

"I went to the cinema" and "I came to the conclusion": Zoning in on acquisitional patterns of learner grammar development

Geraldine Mark, Mary Immaculate College, University of Limerick
Pascual Pérez-Paredes, University of Cambridge
Anne O'Keeffe, Mary Immaculate College, University of Limerick

Corpus-based Second Language Acquisition and usage-based accounts of language learning are increasingly focusing on the developmental nature of the acquisition of patterns in learner language as opposed to solely focusing on learner errors. By using corpus linguistic methods, it is possible to look in detail at patterns of usage as learners' proficiency develops through notional levels. In doing so, it is possible to explore whether growth in lexico-grammatical repertoire converges with L1 usage. Existing work on morphosyntactic patterning by Ellis and his associates in relation to a usage-based model of acquisition (Ellis et al. 2016) becomes relevant to this exploration as they have shown evidence of Zipfian patterns of usage with regard to the frequency of use of English verbs and, in particular, verb argument constructions (VACs). In this study, we will use the Cambridge Learner Corpus (CLC), a 55-million-word corpus of learner exam data, from over 200,000 exam scripts, across 200 countries, from candidates of over 140 first language backgrounds. The CLC is calibrated across the six levels of the CEFR and this allows for tracking of the evolving acquisition of syntactic forms, uses and distributional patterns, across different levels, learners, contexts, and L1 backgrounds. It also allows us to explore whether register is also an important variable in the acquisition patterns. Specifically, this study will look at the most frequent patterns involving the use of a verb in the past simple and their arguments at different levels of proficiency. It is hoped that the findings will offer an interesting complement to studies on morphosyntactic patterning and show how syntactic complexity evolves.

References:

Ellis, N. C., Römer, U. & O'Donnell, M. B. (2016). *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar.* Language Learning Monograph Series. Wiley-Blackwell.

Linguistic Resources for facilitating MOOC discussion forum: A Corpus Driven Approach

Shi Min Chua, The Open University

Massive Open Online Courses (MOOCs) have enabled people from around the world to access university-level courses freely online. One signature of MOOCs is its discussion forums where learners with different experiences can interact with each other, thus engaging in dialogic learning (Ferguson & Sharples, 2014). To enhance this learning experience, educators and mentors also participate in the forums to facilitate the discussions. To examine how educators and mentors use different linguistic resources to engage learners, open up and move forward discussions, a corpus of 96078 discussion postings (5,444,003 tokens) from 12 MOOCs was constructed. A corpus-driven approach was used in this study to explore the linguistic resources in this new form of computer-mediated communication, which is situated in a public learning environment. A keyword analysis was conducted by comparing the facilitators' postings (9460 postings, 460,499 tokens) against the learners' postings (86618 postings, 4,983,504 tokens). Using log-likelihood ratio test, 430 keywords in the facilitators' postings were found significant at $p<.000001$. Preliminary categorization of the top 50 keywords by reading 200 randomly selected concordance lines of each keywords resulted in ten functional categories, including namely encouragement for learners (e.g., point, glad, questions, great), course-specific terminology (e.g., data, corpus), social responses (e.g., hi, thanks), hedging (e.g., might), moving forward discussion (e.g., if, questions), agreement and answer (e.g., yes), directives for platform course issues (e.g., please, feedback), pre-empting course content (e.g., week), giving examples (e.g., example), directing attention (e.g., see, this). This variety of functional categories suggests that educators and mentors engaged with learners with different strategies. Further analysis of these 50 keywords on collocations and the whole discussion threads where the keywords occur showed that some strategies (pre-empting content and directing attention) appeared to close down the conversations whereas encouragements triggers positive responses from learners.

Reference:

Ferguson, R., Sharples, M. (2014) Innovative pedagogy at massive scale: Teaching and learning in MOOCs. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 98–111.

On formulaicity of Russian patient information leaflets: a corpus-driven study

Lukasz Grabowski, University of Opole

Conducted using a corpus-driven approach to study texts, this study focuses on the exploration of key vocabulary and phraseology in a specialist text type, namely patient information leaflets originally written in Russian and collected in a custom-designed corpus. More specifically, the aim of this corpus-driven register analysis is to, first, identify keywords and recurrent sequences of words that contribute the most to the leaflets' formulaicity and, second, to describe their use and discourse functions. The keywords were identified using three different methods (G2, Hedges' g and Neozeta) and the overlapping ones were selected for further qualitative analyses. As for the recurrent multi-word items with the largest coverage in the corpus, they were identified using a recently proposed method called Formulex (Forsyth 2015), which helps one avoid the problem of overlapping sequences of n-words or of those sequences of words that constitute fragments of longer n-grams. Apart from descriptive data and methodological novelties (comparison of the output of three keyword extraction methods, the application of the Formulex method), the results revealed, among others, that frequent use and discourse functions of the identified patterns result from situational contexts of use as well as communicative purposes of patient information leaflets. Since corpus linguistic phraseological research conducted on texts written in Russian is still scarce, it is hoped that the results of this descriptive and exploratory study will provide an opportunity for a methodological reflection and inspire further research in the future.

References:

Dunning, T. (1993). "Accurate Methods for the Statistics of Surprise and Coincidence". *Computational Linguistics*, 19 (1): 61-74. Forsyth, R. (2015). Formulib: Formulaic Language Software Library. User notes http://www.richardsandesforsyth.net/docs/formulib.pdf

Hedges, L. (1981). Distribution Theory for Glass"s Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics*, 6 (2): 107–128.

Craig, H. & Kinney, A. F. (Eds) (2009). *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press.

Linked noun groups: teaching awareness of genre-specific uses

Michael Pace-Sigge, University of Eastern Finland

The English language makes use of a large number of multiword-units (MWU) and these are often highly idiomatic. Amongst word pairs and trigrams displaying idiomaticity, a form occurring less frequently is that of a pair of nouns which is linked by a conjunction. Some of these are widely used and easily recognised, for example *law and order* or *boy and girl*. However, there are a number of less obvious trigrams of that sort, for example *nose and cheekbone; increase or decrease*. Relatively rare are nominal pairs with the conjunction *but*. The Collins Cobuild English Grammar (which is based on a Pattern Grammar – led classification) refers to these as *linked noun groups*.

Significantly, however, the occurrence pattern of such pairs is dependent on genre. Consequently, this paper will look at 19th and 20th century fiction texts, written and spoken academic English (BAWE and BASE) as well as casual spoken British English to highlight the most frequently occurring pairs in order to give an indication of the most natural usage patterns. Domain specificity also relies on order: thus, for example, *gold and silver* is more commonly used in reference to the prices of these metals, whereas *silver and gold* is connected, firstly, with song lyrics. Lastly, the example of one fiction (Dickens) and one non-fiction (Marx) writer will be given to demonstrate idiosyncratic (or creative) uses.

For any learner of the English language, awareness of idiomatic patterns like these serves to assist their register-specific application and it will be shown how this can be employed at various levels of text composition.

References:

Biber, D., and Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Collins Cobuild English Grammar. (1990). London and Glasgow: Collins.

Sardinha, Tony Berber (2017). Lexical priming and register variation. In Pace-Sigge, M. and Patterson, K.J. *Lexical Priming. Applications and Advances.* Amsterdam: John Benjamins, pp. 189-230.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Verb constructions in second language acquisition: From learner corpus analysis to pedagogical recommendations

Ute Römer, Georgia State University

This talk adopts a usage-based perspective on language acquisition to investigate how knowledge of verb-argument constructions (VACs) develops in English as a second language learners across proficiency levels. I will first present findings from an analysis of L1 German and L1 Spanish learner use of English VACs, such as the 'V n n' or ditransitive construction (e.g., *he brought her chocolate*) and then discuss what the findings mean for language pedagogy.

Findings presented in this talk come from a large-scale study on L2 English construction development which uses methods from Corpus Linguistics and Natural Language Processing to systematically extract VACs from a pseudo-longitudinal learner corpus. The study addresses the following research questions:

1. How does the VAC repertoire of learners develop across proficiency levels?
2. How does the distribution of verbs in VACs in learner production develop across proficiency levels?
3. What role do formulaic sequences play in the L2 acquisition of VACs?

To address these questions, data on verbs and the constructions they occur in was exhaustively extracted from a dependency-parsed corpus of L2 writing. The corpus is a 6-million word subset of EFCAMDAT, the Education First-Cambridge Open Language Database (Geertzen, Alexopoulou, & Korhonen, 2013), consisting of over 68,000 texts produced by L1 German and L1 Spanish learners at CEFR levels A1 through C1. We generated frequency-sorted VAC and verb-VAC lists for each level and L1 (e.g., German_A1) and extracted recurring multi-word clusters (spans 3, 4, and 5) around the 50 most frequent verbs in EFCAMDAT.

I will share selected results on verb construction development across learner proficiency levels. I will also discuss how insights from this and related studies can inform pedagogical practice and have a positive impact on second language teaching.

Lexical development in English school children's writing from six to sixteen

Philip Durrant, University of Exeter
Mark Brenchley, University of Exeter

This presentation will discuss part of an ongoing study which focuses on the age-related development of written language in school children in England. Our research team has established the largest corpus of authentic school writing in England to date, with the aim of providing a more thorough understanding of linguistic development in writing. The presentation will focus in particular on the lexical element of the study. Most previous work on the written productive vocabulary of L1 children has been on a relatively small scale and has concentrated largely on measures of diversity and on simple proxies for sophistication (e.g. word length; use of words not on 'high-frequency' lists; use of Latinate vocabulary). The starting point for the current research is the contention that theoretical developments in our understanding of vocabulary and technical advances in the form of larger, grammatically-tagged, corpora, enable analyses which are more sensitive and more informative than has previously been possible. The presentation will describe the analysis of a balanced subsample of our corpus, comparing the vocabulary used in writing by approximately 1,000 children at years two, six, nine and eleven in three different disciplinary groups (English, other Humanities and Science) and in two different genres (which we characterize as literary and non-literary). I will present and discuss the key patterns of vocabulary development at these ages and how these developmental patterns interact with genre-based variation. I will critically evaluate the validity and reliability of key measures of productive vocabulary, and discuss possibilities for the future development of these measures.

The Corpus of Irish Narratives

Chris Fitzgerald, University of Limerick

This paper describes the creation and analysis of a representative corpus of oral histories delivered by Irish English speakers in the 1940's and 1950's. The Corpus of Irish Narratives (COIN) is a 1,000,000 word corpus of language from the Irish Military Bureau archives, which was established by the Irish Minister for Defense in January 1947. Over the subsequent ten years 1773 witness statements were collected by the Bureau staff. They were given the task of travelling throughout Ireland to gather as much information as possible from those involved in the independence movement. This resulted in one of the largest oral history collections of its kind ever undertaken, comprised of over 36,000 pages of statements. Since becoming available to the public in 2003, these histories are being used as valuable sources of historical data relating to the 1916 Rising and war of independence, but have hitherto not been investigated linguistically. This study looks at the statements from a linguistic perspective, using methods of corpus linguistics and discourse analysis to evaluate their significance as sources of linguistic data. Though oral histories have been investigated from a linguistic perspective (Schiffrin, 2003), there remains to be a thorough examination of how they can be utilised within the field of narrative inquiry and how that may bridge the disciplines of history and linguistics. Narrative inquiry is increasingly being looked upon as a tool to further our understanding of not only language, but also language teaching and learning (Benson, 2014). This presentation will outline the potential contribution of oral histories to this field while establishing oral history as a genre of language by outlining the linguistic features that distinguish this type of narrative from others.

References:

Benson, P., 2014. Narrative inquiry in applied linguistics research. *Annual Review of Applied Linguistics*, *34*, pp.154-170.

Schiffrin, D., 2003. We knew that's it: Retelling the turning point of a narrative. *Discourse Studies*, *5*(4), pp.535-561.

The diagnostic potential of Contextual Prosodic Theory in establishing forensically the features of manipulative newspaper discourse in Russian

Marija Milojkovic, University of Belgrade

Contextual Prosodic Theory, developed by Bill Louw, investigates additional semantic layers in authorial collocations, both lexical and lexico-grammatical, against the background of reference corpora. It is one of the premises of Louw's theory that certain collocational clashes, unless they amount to a rhetorical device, may be considered indirect proof that the speaker is 'at pains to conceal' what he or she really thinks (Louw 1993). To date, Louw's theory has dealt mainly with semantic auras of authorial text, and has never been applied to investigating insincerity in natural language, apart from the language of literary personages or personas in lyrical poetry (Louw & Milojkovic 2014; Louw & Milojkovic 2016; Milojkovic 2013). This paper, for the first time, implements his theoretical stance to practical material with a view to proving that a particular statement is insincere. The material in question is an existing text of linguistic expertise which detects features of discrediting tactics in an article published in one of Nizhny Novgorod newspapers. Independently of the expertise results, the methodology founded on Louw's Contextual Prosodic Theory confirms the intention on the part of the journalist to manipulate public opinion. Contrary to expectations, this particular type of insincerity is not founded on collocational mismatches, but on using clichéd combinations in a manner that consistently deviates from their contextual norm of the Russian National Corpus.

References:

Louw, W.E. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In *Text and Technology: In Honour of John Sinclair*, M. Baker, G. Francis and E. Tognini-Bonelli (Eds.), 152-176. Amsterdam: John Benjamins.

Louw, W. E. and Milojkovic M. 2014. Semantic Prosody. In *The Cambridge Handbook of Stylistics*, P. Stockwell and S. Whiteley (eds), 263-280. Cambridge: CUP.

Louw, B. and Milojkovic, M. 2016. *Corpus Stylistics as Contextual Prosodic Theory and Subtext*. Amsterdam: John Benjamins

Milojkovic, M. 2013. Is corpus stylistics bent on self-improvement? The role of reference corpora 20 years after the advent of semantic prosody. In *Journal of Literary Semantics*. 42(1): 59-78.

Conceptualising context in corpus linguistics: moving beyond metadata

Michael Handford, Cardiff University
Dawn Knight, Cardiff University

An ongoing question in corpus studies concerns the role and analysis of context in understanding meaning. For instance, Hunston (2002) states that data from large corpora are, by definition, decontextualized. Researchers working on relatively smaller corpora and who bring together corpus linguistics with other approaches or methodologies, such as discourse analysis (Baker, 2006) multimodality (Knight, 2011) or professional genre analysis (Koester, 2006; Handford, 2010), have argued to varying degrees that context needs to be considered if we are to move beyond superficial meaning, and have attempted to show how context can indeed be contemplated when analyzing corpus data. While attempts to consider the role of context allow for greater understanding of patterns of meaning, the research has typically applied a 'top-down' approach, in that it defines context in terms of text-external categories, often in the form of meta-data, through which the discourse is interpreted. An example is corpus-informed research on culture (e.g. Stubbs, 1996; Connor et al, 2008), which typically treats cultural categories as a 'given'. This reified, common-sense view of culture contrasts with much discourse-influenced work into sociocultural identity (e.g. Bucholtz and Hall, 2005; Benwell and Stokoe, 2006) which see culture and (socio)cultural identities as constructed, emergent and negotiated in and through discourse. We argue that corpus studies also need to embrace a more emergent, dynamic, bottom-up approach to context, and one that can allow for several contributing factors to account for the meanings in the data. This is particularly apposite for spoken interactions, where context is often fluid and dynamic, and there are various modes for meaning-making that need to be considered. This talk will outline a methodology that allows for the analysis of context as discursively constructed and reflexive.

References:

Baker, P. (2006). *Using Corpora in Discourse Analysis.* London: Continuum.

Benwell, B. & Stokoe, E. (2006). *Discourse and Identity*. Edinburgh: Edinburgh University Press.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7 (4-5), 585–614.

Connor, U., Ruiz-Garrido, M., Rozycki, W., Goering, E., Kinney, E., & Koehler, J. (2008). Patient-directed medicine labeling: Text differences between the United States and Spain. *Communication & Medicine*, 5 (2), 117-132.

Handford, M. (2010). *The Language of Business Meetings*. Cambridge: Cambridge University Press.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Koester, A. (2006). *Investigating Workplace Discourse*. Routledge: London.

Knight, D. (2011). *Multimodality and Active Listenership: A Corpus Approach*. London: Bloomsbury.

Stubbs, M. (1997). *Text & Corpus Analysis*. London: Blackwell.

Let's talk about maths: using classroom data for collaborative professional development

Joanna Baumgart, University of Limerick

With approximately 180+ languages being spoken across the country and 16% of pupils being described as English as an Additional Language learners (Carson and Extra 2010), Irish classrooms have become multicultural and multilingual contexts where mainstream teachers have also taken the role of language instructors, whether consciously or not. This paper presents the results of a professional development project in a multilingual mathematics classroom in an Irish secondary school where five lessons were video recorded and critically analysed by both the teacher and the researcher. The effects of this interdisciplinary collaboration on the teaching and learning environment in this classroom were twofold. On the one hand, reflective meetings provided an opportunity to consider ways how to best respond to the dual challenge of teaching mathematics and language simultaneously. On the other hand, corpus linguistics analysis of the lessons helped to trace changes in the discursive patterns over the five weeks, which resulted from classroom interventions designed collaboratively by the researcher and the participating teacher. Overall, the results indicate that enquiry-based professional development activities, within a framework of dialogic reflection (Mann and Walsh 2013 and 2017) foster reflective processes and create an opportunity to respond more closely to an individual teacher's needs. Furthermore, the paper has brought to light the ways in which CL can be used as part of a combined data-led interdisciplinary approach to teacher development with a view to informing more effective classroom practice.

References:

Carson, L. and Extra, G. (2010) Multilingualism in Dublin: Home language use among primary school children, report on a pilot survey. Dublin: Centre for Language and Communication Studies.

Mann, S. and Walsh, S. (2013) "RP or RIP: A critical perspective on reflective practice", *Applied Linguistics Review* 4(2), pp. 291 – 315.

Mann, S. and Walsh, S. (2017) *Reflective Practice in English Language Teaching: Research-Based Principles and Practices*, London: Routledge.

Difficult Words: An Analysis of Terminology in a Corpus of Irish Education Materials

Mícheál J. Ó Meachair, Trinity College Dublin
Elaine Uí Dhonnchadha, Trinity College Dublin

It has been established that learners of Irish at primary (Pétarváry et al., 2014) and post-primary (Walsh, 2007) struggle to acquire and/or use terminology in studies of learner-generated language. These researchers found learners opting to use English equivalents of terms in a substantial number of cases. Mac Donncha et al. (2004) conducted a survey of parents and teachers which also found negative attitudes towards terminology, as well as reports of teachers not understanding the terminology in certain educational materials. In order to examine how terminology is acquired in the education system (i.e. which terms are used in the classroom, and at which levels of education are they used), I have created a corpus of educational materials. This corpus contains CEFR-GA materials for levels A1 through C2 (over 2 million words); materials for teaching Irish at primary, post-primary, and university level (over 1 million words); and materials for teaching other subjects through the medium of Irish (over 2 million words). The corpus has been automatically tagged for parts of speech and lemmas using Uí Dhonnchadha (2009). Terminology has been identified using the National Terminology Database for Irish (www.tearma.ie). In this paper I will describe the corpus design, the use of the terminology database, and the analysis developed. This analysis focuses on lexical frequency and results include meta information from the National Terminology Database with respect to domain(s) of knowledge and parts of speech.

References:

Fiontar & Scoil na Gaeilge, DCU and An Coiste Téarmaíochta. The National Terminology Database for Irish, www.tearma.ie, last accessed 12/01/2018

Mac Donncha, Ní Chualáin, Ní Shéaghdha, Ní Mhainín. (2004) Staid Reatha na Scoileanna Gaeltachta: Tuarascáil don Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta. URL: http://www.cogg.ie/wp-content/uploads/Staid-Reatha-na-Scoileanna-Gaeltachta-2004.pdf

Péterváry, Ó Curnáin, Ó Giollagáin, Sheahan. (2014) Analysis of Bilingual Competence: Language acquisition among young people in the Gaeltacht

Ui Dhonnchadha, E. (2009). Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar. PhD thesis, Dublin City University

Walsh (2007) Cruinneas na Gaeilge. Master's Thesis. URL: http://www.cogg.ie/wp-content/uploads/clare_walshe_cruinneas_na_gaeilge.pdf

Exploring markers of emergent professional identity and status in an international mobile community.

Margaret Healy, Mary Immaculate College, University of Limerick

Addressing some of the major themes of this conference, this presentation will consider the issues of community and mobility and, through the lens of Discourse Analysis, will also examine the application of corpus linguistics in the real world context of workplace discourse and academic discourse. Using data from the CLAS corpus, the one-million world spoken corpus collected at Shannon College of Hotel Management, Ireland, the discourse under consideration here is situated in a third-level academic institution where students are prepared for future careers in the global arena of international hotel management. Mobility is an integral aspect of this professional career development and, indeed, students at Shannon College are already engaged in bilateral mobility as approximately half of them are overseas students and the other half, mainly the Irish students, travel internationally on placement practices abroad as part of their degree programme. The College itself encapsulates quite a distinct, definable community, unique in academic institutional environments, presenting a microcosm of the larger community of international tourism and, specifically, of hotel management. This community of practice (Wenger, 1998) within the both the academic and professional education spheres is constructed through its discourse which builds the character and professional identity of its students. This presentation will examine the specific discourse of this micro world, using corpus linguistics methodologies, to mark some of the distinctive linguistic features found in the CLAS corpus data. Certain linguistic issues, for example pronunciation challenges for the non-native speakers, domain specific terminology and lexical profiles, and students' acquisition of international business discourse and practices, will be offered as instances of the intertwining of language and emergent identity in this professional mobile community which is central to the academic and professional development of the international hotel managers of the future. Keywords: academic discourse, workplace discourse, hotel management discourse, community of practice.

Reference:

Wenger, E. (1991). *Communities of practice: learning, meaning and identity*. Cambridge: Cambridge University Press.

On corpus linguistic approaches to study language: a critical reflection

Milena Hebal-Jezierska, University of Warsaw
Łukasz Grabowski, University of Opole

It is a bit of a cliche to say that the properties of a language corpus (its structure, size, representativeness, balance, annotation etc.) and of the corpus search query system to a large extent determine its potential for conducting linguistic research. Additionally, a corpus linguistic methodology is far from homogeneous and that is why before stating research questions or formulating research hypotheses, it is necessary to ascertain whether a given corpus or research methods are suitable for a task at hand and would allow to obtain reliable results. That is why in this presentation, we discuss certain opportunities and limitations of various corpus linguistic methods used for acquisition, analysis and interpretation of linguistic data found in language corpora, notably in the so-called national corpora (e.g. the BNC, National Corpus of Polish or Czech National Corpus). The inter-related problems under scrutiny pertain to the accuracy of annotation of linguistic data, too detailed tagsets, negative evidence in language corpora, precision of search queries, abundance of search results, limitations of in-built corpus search engines and concordance programs as well as to the use of tests of statistical significance when comparing frequencies of linguistic items across corpora, as compared with effect size metrics. Hence, using a number of small case studies (pertaining to English, Czech, Polish, Russian and Slovak languages and based on linguistic data extracted from the national corpora of those languages), this critical reflection paper is primarily designed to reflect upon the application of corpus linguistic methodologies to study language taking national corpora as a case in point.

Using a Multimodal Corpus in EFL Classrooms; an Action Research on Learner Attitude

Sara Aljohani, University of Limerick

Steady steps have been taken to facilitate the creation of pedagogic corpora and their use in ESL/EFL classrooms (Kohn et al. 2009; Braun and Kohn 2012), yet practical issues persist. Informed by the body of related literature, this research showcases a learner-centric and context-specific approach to designing a pedagogic corpus and preparing EFL learners to use it. This research also examines the corpus' efficacy through considering learners' perspectives. In particular, this paper investigates EFL learners' attitudes to corpus use and the potential influence of learners' profiles on their attitudes. For this, the researcher 1) created a web-based multimodal corpus of TED talks (MCOTT), and 2) implemented a five-week workshop to raise awareness of spoken academic discourse through the use of the MCOTT. Participants were intermediate and upper-intermediate EFL learners enrolled in the foundation year of their tertiary education and sought to improve their academic speaking. Data were gathered through two questionnaires (i.e. learner profile and learner attitude questionnaires), interviews, and classroom observations, and were analysed both quantitatively and qualitatively. To investigate learners' attitudes, this paper explored questionnaire results relating to participants' perceived usefulness, perceived ease of use, perceived behavioural control, attitude, and future intentions. Analysis of interviews and classroom observations also helped extending our understanding of participants' attitudes. To determine the possible influence of participants' profiles on their attitudes, this paper examined the correlation between participants' profiles (motivation, attitude to oral skills, attitude to autonomy and ICT competence) and their attitudes to using the MCOTT. Finally, while findings revealed the positive attitudes of the participants, questions are raised regarding whether the content of the corpus (i.e. TED) had a significant impact on participants' attitudes.

You can't achieve a shrimp: a corpus-driven approach to increased precision in noun-verb collocations

Jane Templeton, University of Leeds

Noun-verb collocations are a common problematic area for learners. In the high-stakes context of UK academic study, such errors can have impacts on performance in assessments. Although the vocabulary is often frequent, errors are common. Analysis of pre-undergraduate written assessment tasks suggests that a focus in the classroom on precision in choice of verb collocates for shell nouns would be profitable. However, given the relatively wide range of possible collocates, this is not necessarily easy to teach or learn. It is proposed that what learners really need is corpus research skills, so as to be able to find an appropriate verb collocate for any noun. However, corpus research is not an intuitive skill, so it needs to be taught, and this is problematic for teachers who have not found, or not looked for, effective methods of doing so. The aim of this presentation is to show how one online corpus tool, Mark Davies' www.wordandphrase.info, can be used, easily and with little preparation time, to teach students how to answer two key questions: can I use this verb here with this noun? and what verb should I use here with this noun? I will demonstrate two methods of integrating it into class activities: firstly, as a reference tool in response to production errors or questions; secondly, as part of a planned lesson, at the beginning of a new topic or prior to the initial processing of a text. I will show how an analysis of a shell noun using its collocates can provide not only a framework for analysing a topic conceptually, but also a language reference. I will also explain how this method can provide a way in to, and thereby unlock the benefits of, data-driven learning, in a way which sidesteps some of the issues inherent in the conventional model.

Populism in Italian politicians' social messages. A corpus-based study of three politicians' social media discourse.

Dario Del Fante, University of Padova
Virginia Zorzi, University of Padova

The rise of populist trends in Western politics has become increasingly relevant in media arenas (Rydgren 2017, Wodak and Krzyżanowski 2017). Populist views, enforced by the adoption of specific communication strategies, place "the people" as an unquestionable entity at the centre of political practices and decisions (Pelinka 2013). In this context, social media have been identified as a major site of populist communicative efforts (KhosraviNik, and Unger 2016). Populist tendencies have emerged in political groups inspired by radically different ideologies (Moffitt 2016). Considering language as an instrument of social interaction, in line with SFL (Halliday and Matthiessen 2004) and CDA (Fairclough 2001), we address the discourses of Italian politicians who are generally considered as embodying – at least partially - populist standpoints. Our aim is to understand whether the central concept of "people", in all its possible linguistic realisations (Chilton 2017: 586), is built differently across different politicians' discourses, and how it is constructed as an entity which is functional to different communicative purposes. Our analysis is focussed on Twitter and Facebook posts published between 2016 and 2017 by three Italian political leaders representing different political groups: M. Renzi, M. Salvini and L. Di Maio. Our method combines a corpus-driven approach aimed at identifying the main themes and linguistic structures in the analysed messages with a qualitative study of the most relevant examples (Partington, Duguid and Taylor 2013, Mautner 2016). A preliminary analysis highlights how different thematic focuses are exploited in the communicative styles specific to each politician.

References:

Chilton, P. (2017). "The people" in populist discourse. *Journal of Language and Politics*, 16 (4), 582 –594.

Fairclough, N. (2001). *Language and Power*. London: Longman.

Halliday, M. A. K., and Matthiessen, C. M. (2004). *An Introduction to Functional Grammar*. Third Edition. London: Hodder Arnold.

KhosraviNik, M., and Unger, J. W.(2016) Critical discourse studies and social media: power, resistance and critique in changing media ecologies. In R. Wodak, and M. Meyer, M. (eds.) *Methods of Critical Discourse Studies* (pp. 205–33). London: Sage.

Mautner, G. (2016). Checks and balances: How corpus linguistics can contribute to CDA. In Wodak, R., and Meyer, M. (Eds.) *Methods of Critical Discourse Studies* (pp. 154–179). London: Sage.

Moffitt, B. (2016). *The Global Rise of Populism: Performance, Political Style, and Representation*. Stanford: Stanford University Press.

Partington, A., Duguid, A., and Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies* (CADS) Amsterdam: John Benjamins Publishing.

Stories and 'story-likes': How lecturers use real and hypothetical events to scaffold knowledge in different EMI contexts

Siân Alsop, Coventry University
Hilary Nesi, Coventry University

Storytelling is recognized as an important rhetorical mode in sociolinguistic studies of spoken language. In the educational context, retelling events in story form is commonly used to convey information, construct new knowledge, and make sense of experience (McDrury and Alterio 2002). This paper examines the occurrence, linguistic characteristics, and pedagogical purposes of various forms of storytelling in lectures that are delivered in English-medium in different cultural/educational contexts. Data is drawn from the Engineering Lecture Corpus (ELC), which is a collection of 76 lectures recorded in three universities worldwide. Each lecture transcript has been manually annotated for pragmatic functions, including four types of story (cf. Martin 2008): anecdote, exemplum, narrative and recount. Categories were assigned by initially identifying sequence units following Labovian (e.g. 1967) rules, and then following Martin's choice pathway, which determines genre based on the presence (or absence) of complication, evaluation and resolution units. The start and end of each type of story is indexed in the transcript of each lecture. We have also identified a further category related to story, which we refer to as 'story-like'. In the same way as stories, story-likes re-tell events, frequently using analogies to explain engineering concepts. These events, however, are not situated in the past but instead use hypotheses or predictions of future events to make general claims of relevance to the subject matter. Using examples from the ELC, we will compare the use of stories and story-likes across three cultural subcorpora. Better understanding this usage may be of benefit to both staff and students who move between educational contexts, as well as more generally for supporting lecturer training and student listening skills.

References:

Labov, W. and Waletzky, J. (1967) 'Narrative Analysis: Oral Versions of Personal Experience'. in Sociolinguistics: *The Essential Readings* (2003). ed. by Bratt Paulston, C. and Tucker, G. R. Oxford: Wiley-Blackwell, 74-104.

Martin, J. R. (2008) 'Negotiating Values: Narrative and Exposition'. *Bioethical Inquiry* 5, 41-55

McDrury, J. and Alterio, M. G. (2002) *Learning through Storytelling*. Palmerston North: The Dunmore Press.

Using a learner corpus to develop English language teaching (ELT) materials targeted at Spanish learners

Julie Moore, Cambridge University Press

Although we live in an apparently globalized world, a global, one-size-fits-all solution to language-teaching materials may not be the best fit for all learners. Whilst there are undoubtedly some universal principles that apply across contexts and across language backgrounds, different learners approach the task of learning a language with different linguistic and cultural baggage. Amongst these factors, it is clear that the role of the learner's first language (their L1) has a significant impact. The issue of 'language transfer' or 'L1 interference' has long been recognized in language teaching. In this paper, I will look at how the Cambridge Learner Corpus has been used to help address the issue of L1 interference in ELT materials. I will explain how data from Spanish-speaking students was analysed to identify specific problem areas in lexis, grammar and spelling at different levels and was then used to feed directly into the development of teaching materials targeted at this group. I will look at how issues were identified in the data, both those which might be described as directly a result of L1 interference (interlingual errors) and those which could be seen as a general part of L2 development (intralingual or developmental errors), and how this information was related to the teaching syllabus. I will explore how analysis of the corpus data uncovered errors that may have otherwise been overlooked by teachers and materials writers. I will also present examples of the resulting materials and feedback on these from teachers.

Understanding the Interests of Communities of TripAdvisor Travellers through a Corpus-Based Discourse Analytic Approach

Phoenix Lam, The Hong Kong Polytechnic University

As one of the most important service-oriented industries in contemporary society, tourism has increasingly seen the influence of the Internet on all aspects of travelling. Travellers nowadays habitually research online before making travel-related decisions. One platform on which such research is conducted is destination forums. The emergence of such online destination forums in the last decade has allowed tourists to share their travel experiences quickly and easily with a large number of online users around the world. As such, these destination forums also provide invaluable data for tourism bodies to better understand travellers' views on their destinations. Collecting posts from the Hong Kong travel forum on the world's largest travel website TripAdvisor®, the present study identifies the top questions asked by TripAdvisor users about Hong Kong through a corpus-based discourse analytic approach. Based on questions posted on the forum and their associated meta-data gathered in a one-year period, the study examines the top questions asked by travellers around the world to identify the key geographical locations in which users have shown the greatest interest in the city. Questions raised by travellers from different geographical locations are also compared to see if traveller communities by location vary in terms of their areas of interest. This analysis involves the study of keywords and concordance of frequently-occurring items, and a close reading of representative examples in context. Findings from the present study show that travellers who asked the most questions about Hong Kong are from North America and Asia, and that travellers from different locations have different concerns and interests, which are clearly reflected in the language of the questions asked on the travel forum. These findings can therefore provide tourism organisations with useful information about the key markets that should be targeted for promotional purposes, and can also allow such organisations to design advertising campaigns which better address the specific needs of such markets. The present study thus demonstrates the value of applying linguistic knowledge and methodologies to the domain of tourism to address practical issues.

Science in the UK media: An analysis of two newspapers

Geri Popova, Goldsmiths, University of London

This paper draws on an international project which explores science representation in the media across a number of different contexts. We focus on the UK and compare science reporting in one broadsheet, The Times, and one tabloid, The Mirror. The project collected a large corpus of data from both newspapers spanning 20 years, which allows the use of corpus analytic, and also manual techniques of discourse analysis. In this paper we try to build a picture of differences and similarities between science representation in the two newspapers using corpus analytic techniques like keyword comparison and analysis of concordance lines. In some cases our findings are supported by the manual investigations of relevant texts. A keyword comparison using Antconc between The Mirror and The Times corpora identifies, for example, the pronouns you, your, my, I, her, she, me as significantly more frequent in the The Mirror than The Times. An examination of the contexts of use of these pronouns suggests a preponderance of health-related stories in The Mirror (cancer and hospital are also amongst the keywords), and a tendency to explore health-related stories as narratives of personal experience. Another tendency is to position the newspaper as a source of advice for the reader and in such cases to address the reader directly. Significantly more frequent in The Times is the word science itself. Together with some of the other keywords we found, e.g. company and companies, business, professor, student this suggests an increased tendency to view science as an institutionalised world with a more general economic and social significance. There are also similarities. In both newspapers science has the role to discover how the world is. We used Sketch Engine to look at the verbs that most frequently occur with scientist(s) and researcher(s), and in both newspapers we find that scientists find, claim, develop, think, discover, examine, investigate, etc. Interestingly, in both newspapers the most frequent verb associated with scientist(s) is the verb believe, suggesting that the media adopts a certain distance to the world of scientific discovery and makes explicit its role as mediator between that world and the reader.

Investigating advanced translation learners' problems of lexical cohesion: A corpus-based study of lexical repetition in English to Chinese translation

Jun Pan, Hong Kong Baptist University
Honghua Wang, Hang Seng Management College

Textual competence has been regarded an essential component in translator training (Nord, 2005). One important aspect of textual competence is reflected in the appropriate employment of lexical cohesive tools, which form the cornerstone of a text (see Halliday and Hasan, 1976). Nevertheless, students' use of lexical cohesion has rarely been addressed in translator training. This study thus aims to investigate advanced translation learners' problems of lexical cohesion, as illustrated by misused lexical repetition, in their English to Chinese translations. Two corpora were developed for the study: a translation learner corpus consisting of translation outputs of senior year translation major students, and a professional translation corpus including line-edited versions of the same translations done by professional translators / translation teachers. Lexical repetitions were annotated in both corpora. Then, the frequencies of lexical repetition were calculated and cross-examined. Comparisons were made between the translation learner corpus and professional translation corpus. Findings of the study show the differences in the employment of lexical repetition between advanced translation learners and professional translators. They also show the types and pattern of problems of lexical cohesion in semi-professional translation students. The study provides insights into the training of translation learners at an advanced stage. The corpora developed in the study also serve as valuable pedagogical resources for the enhancement of lexical cohesion in translation.

References:

Nord, Christiane. (2005). *Text Analysis in Translation: Theory, Methodology, and Didactic Application of a Model for Translation-oriented Text Analysis* (2nd Ed.). Amsterdam / New York: Rodopi.

Halliday, M.A.K. and Hasan, Ruqaiya. (1976). *Cohesion in English*. London: Longman.

A corpus-based comparison of registers in EFL school textbooks for secondary schools in France, Germany and Spain

Elen Le Foll, Osnabrück University

The majority of secondary school foreign language syllabi in Europe are based on commercial textbooks. It therefore follows that thorough evaluations of the authenticity of the linguistic content of these EFL textbooks is paramount. So far, textbook language evaluations have either considered textbook language as one register (e.g., Ljung, 1990), thus disregarding major intra-textbook register variations, or focused solely on textbook dialogues (e.g., Mindt, 1992; Römer, 2005). This study explores the lexico-grammatical specificities of a range of textbook registers and compares the language input students obtain through their textbooks in three European countries. The textbook corpus compiled to his end consists of nine series of EFL textbooks (42 textbooks in total) published between 2006 and 2017 and currently used in secondary schools in France, Germany and Spain. The textbook data has been manually tagged for text type in order to create spoken, narrative, informative and instructional textbook language subcorpora. In a first step, Multi-Dimensional Analysis (cf. Biber, 1986, 1988) is applied to map out the lexico-grammatical specificities of the different registers featured in EFL textbooks. Further, quantitative and qualitative analysis of the frequencies, functions and collocations of linguistic features typically associated with each register are carried out. The occurrences of these features in school EFL textbooks are compared to those occurring in a specifically assembled reference corpus designed to be as representative as possible of students' target learner language. The observed differences between the textbook registers and the corresponding reference sub-corpora inform recommendations to improve the authenticity of the language input students are exposed to via their school textbooks.

References:

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 384–414.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Ljung, M. (1990). A Study of TEFL Vocabulary. Stockholm Studies in English 78. Stockholm: Almgvist & Wlksell International.

Mindt, D. (1992). *Zeitbezug im Englischen: eine didaktische Grammatik des englischen Futurs*. Tübingen: Gunter Narr Verlag.

Römer, U. (2005). *Progressives, Patterns, Pedagogy: A corpus-driven approach to English progressive forms, functions, contexts, and didactics*. Amsterdam: John Benjamins.

A learner corpus research on the diagnosis of accuracy in grammatical morphemes by interlanguage users: A case of prospective English teachers in Turkey

Erdem Akbas, Erciyes University
Zeynep Olcu-Dincer, Erciyes University

This exploratory study aims to report the initial step of a longitudinal project, which seeks to unearth the developmental trajectories of L2 teacher candidates by comparing their texts as freshman and as senior students during pre-service education in EFL context. Within the scope of this paper, we scrutinized the language use proficiency and metalinguistic awareness of L2 freshman pre-service teachers of English by examining corpus data in relation to the acquisition of grammatical morphemes and analysing think aloud protocols conducted with purposefully selected participants. To this end, a learner corpus consisting of 136 exam scripts, produced by the freshmen prospective teachers of English, approximately 20,000 words has been compiled during the first semester of 2016-2017 academic year and analysed via UAM Corpus Tool. The corpus was manually annotated by following the labels and formula adapted by Pica (1984) and the target like uses for each grammatical morpheme were calculated. In general, the analysis of the corpus showed us the Turkish learners of English mastered using all grammatical morphemes to a greater extent; nevertheless, the learners seemed to have a relatively less accuracy in employing the marker of plural –s in their written texts, especially with the application of plural –s to the irregular plural nouns. It is surprising since the acquisition of the grammatical morpheme plural –s is generally known to be mastered by the native speakers of English at an early stage of their language development. The calculated TLU results also indicated that the acquisition of morphemes followed a considerably different developmental order from NS although various studies found that the order of these grammatical morphemes by L2 learners was quite similar (Murakami, 2013). This can be linked to the exposure to the explicit grammar instructions and importance of some morphemes over others in English major departments & curriculum.

Keywords: learner corpus research, second language acquisition, grammatical morphemes, prospective teachers of English, metalinguistic awareness, interlanguage

References:

Murakami, A. (2013). Cross-linguistic influence on the accuracy order of L2 English grammatical morphemes. In S. Granger, G. Gilquin, F. Meunier (Eds.), *Twenty years of learner corpus research. Looking back, moving ahead: Corpora and language in use*, *1*, Presses Universitaires de Louvain, Louvain-la-Neuve, Belgium (2013), pp. 325–334.

Pica, T. (1984). Methods of morpheme quantification: Their effect on the interpretation of second language data, *Studies in Second Language Acquisition, 6*, 69-78.

Corpus Analysis of Modal Verb unda in Georgian

Sophiko Daraselia, University of Leeds
Nino Sharashenidze, Ivane Javakhishvili Tbilisi State University

The concept of modality is extremely interesting in Georgian for several reasons. Firstly, the category of modality has developed at a relatively later stage in Georgian and it is not described in the grammars of the language. Secondly, there are very few studies about modality (Sharashenidze 2014). Hence, it is still a largely unexplored area of research providing opportunities to reveal new insights into the language. The paper presents the preliminary results of the corpus analysis of the core modal verb 'unda' "must, should" in the KaWaC corpus (Daraselia & Sharoff 2014). The 'unda' is the third person subject, singular form of an independent verb 'ndoma' meaning 'to wish, to desire'. The semantics of this modal verb can vary depending on the tense, aspect and mood of the main verb. For instance, when 'unda' co-occurs with the verb in Aorist (II) Subjunctive tense, it expresses necessity, desirability, whereas with verb in II Resultative (Pluperfect) it expresses unfulfilled obligation or duty. For the purposes of this research, we have analysed the collocations of the unda modal verb emphasizing three key characteristics of collocation: distance, frequency, and exclusivity (Brezina & McEnery 2015). In this collocation analysis, we used the Log-Likelihood Ratio and Mutual Information (MI) measures within 4 in the left and right span of the node. We paid particular attention to verb collocates, distribution of tense, aspect and mood forms and their modal semantics. As a preliminary result, we have identified the three main verb forms that can co-occur with 'unda'. They are: Present Subjunctive, II Subjunctive and II Resultative forms. The next stage of the research will involve the semantic analysis of how these forms affect the meaning of the modal verb.

References:

Brezina V., McEnery T. and Wattam S. (2015) Collocations in context. A new perspective on collocation networks International, in: *Journal of Corpus Linguistics* 20 [2], pp. 139-173.

Daraselia S. and Sharoff S. (2014) Towards Creating a Large Corpus for Georgian, 7th Biennial IVACS Conference, Newcastle University, 19-21 June, 2014.

Sharashenidze N. (2014). *Category of Modality in Georgian*, Ivane Javakhishvili Tbilisi University Press, pp.80-90.

The Effect of Indirect Coded Corrective Feedback and Error-Tagged Learner Corpora on the Reduction of Errors in L2 Writing Tasks

Chiachieh Tang, National Taiwan Normal University

Being a more explicit exemplar of indirect corrective feedback (CF), previous literature has suggested providing indirect coded CF as a way to help L2 learners reduce errors in writing. However, the effectiveness of indirect coded CF for all L2 learners has been in question due to its lack of metalinguistic explanations. In addition to that, EFL teachers and researchers alike have reiterated the need for a less time-consuming method of providing CF for written work. In order to provide feedback with more concise explanations, recent studies have suggested the use of error-tagged learner corpora as a more efficient alternative to indirect coded CF as a way to improve writing skills among L2 learners. To this end, the present study aims to compare the effects of providing error-tagged learner corpora and indirect coded CF in reducing errors and improving writing performance on a revision task. The participants consisted of 54 8th-graders from a private junior high school in Taiwan. Participants in the comparison group (N=27) had to complete a revision task after having been provided indirect coded CF based on the errors made on an initial writing task whereas those in the treatment group (N=27) worked on the same initial writing and revision tasks while consulting an error-tagged learner corpus as a reference resource. The results show that participants in the treatment group outperformed those in the comparison group in terms of both error reduction and overall writing performance. The findings of this topic and implications for future research are discussed. Finally, pedagogical applications for how indirect coded CF and error-tagged learner corpora can impact L2 writing are also highlighted.

The ELT Archive Textbook Corpus: How much has language teaching changed?

Rachel Allan, MidSweden University

This presentation describes and demonstrates some potential uses of a historical English Language Teaching (ELT) corpus, using a small pilot corpus of intermediate level ELT textbooks from the 1960s. There have been few systematic and objective historical accounts of language teaching methods and materials (Smith 2015: 84), and the history of ELT has typically been presented as a rather simplified 'procession of methods', with an emphasis on their difference rather than any similarities (Howatt and Smith 2014: 76). Here I will explore some ways in which a corpus-based approach can contribute to our knowledge in this area. The 1960s have been identified as a starting point for the corpus as a range of methodological influences were present in ELT in this decade. The books in the pilot have been selected to reflect this, including courses which situate themselves within structural, audiolingual and situational teaching paradigms. From the corpus, we can identify task-types, examine instructions and target language, and use this data to gain insights into the pedagogical approaches in use. The 1960s corpus is a pilot for a larger project that aims to investigate twentieth century English Language Teaching (ELT) materials diachronically. With the co-operation of the ELT Archive at Warwick University (see https://warwick.ac.uk/fac/soc/al/research/collections/elt_archive/), we aim to digitize a range of representative ELT textbooks from several decades of the last century, and build a corpus through which language teaching methods and language taught can be explored and compared.

References:

Howatt, A.P.R. and R. Smith. 2014. "The History of Teaching English as a Foreign Language, from a British and European Perspective." Language and History 57 (1): 75-95.

Smith, Richard. 2015. "Building 'Applied Linguistic Historiography': Rationale, Scope, and Methods." Applied Linguistics 37 (1): 71–87. doi:10.1093/applin/amv056.

Multimodal Corpus in Teaching Russian as a Foreign Language

Michal Kozdra, University of Warsaw

The article deals with possibilities of using a multimodal corpus in teaching Russian as a foreign language. The multimodal corpus within Russian National Corpus – Ruscorpora.ru will be analysed as an example [1]. The multimodal corpus can be defined as a collection of content on different modes of communication including speech, body language: gestures, facial expressions, as well as music, pictures and colours [2]. Therefore, it contains semiotically inhomogeneous objects: verbal (signs of natural language) and non-verbal, that is signs belonging to other semiotic systems, such as iconic (colour, font, video, gestural expressivity) [3]. Firstly, the use of a multimodal corpus in Russian language lessons encourages interest and motivation, provoke communication in a foreign language and improves the memorisation processes. Secondly, multimodal corpora help in understanding the meaning of words and in using them correctly in the speech. This is possible through multimodal semantisation which I define as identifying the meaning of the language units by perceiving different semiotic codes. Thirdly, multimodal corpora help to introduce and activate the real-life vocabulary of different semantic topics, express communicative intentions and emotions, as well as provide more natural communication. Finally, the usage of multimodal corpora contributes to the improvement of socio-cultural and "multimodal communicative competence" [4, 5, 6, 7].

Keywords: multimodal corpus, teaching, Russian as a foreign language, verbal, non-verbal communication

References:

1. Russian National Corpus, URL: http://ruscorpora.ru/
2. Foster, M.E. & Oberlander, J. (2007): Corpus-Based Generation of Head And Eyebrow Motion for an Embodied Conversational Agent. In: Proceedings of The International Language Resources and Evaluation Conference (LREC) 41 (3/4), 305-323.
3. Anisimova, E.E. (2003): Lingvistika teksta i mežkulʹturnaâ kommunikaciâ (na materiale kreolizovannyh tekstov): M.: Izdatelʹskij centr «Akademiâ».
4. Coccetta, F. (2009): Multimodal Text Analysis and English Language Teaching, URL: https://bit.ly/2G1u7zY
5. Harris T., Jaén M.M. (eds.) (2010): Corpus Linguistics in Language Teaching, Peter Lang AG, Internationaler Verlag der Wissenschaften; New edition.
6. Royce, Terry D. (2007): Multimodal Communicative Competence in Second Language Contexts. In: Royce, Terry D. and Bowcher, Wendy L. (eds.), New Directions in the Analysis of Multimodal Discourse. Mahwah and London: Erlbaum, 361-390.
7. Yen-Liang L. (2017): Co-occurrence of speech and gestures: A multimodal corpus linguistic approach to intercultural interaction. In: Journal of Pragmatics, Volume 117, 155-167, URL: https://bit.ly/2I4j9zO

Refugees in the Spanish written media: A corpus-based study from a semantic preference perspective. Towards a categorization of a negative portrayal

Gema Alcaraz-Maĕrmol, Universidad de Castilla-La Mancha
Jorge Soto Almela, Universidad Católica de Murcia

Although refugees have been arriving in Europe from the second half of the 20th century and into the start of 21st century, the number of asylum seekers has dramatically increased in the last years due to the conflicts in the Middle East, and especially the Civil War in Syria. This phenomenon has drawn the attention of the media, where daily news about this issue are published. Given the repercussions of the phenomenon of refugees, this study aims to explore – within the methodological framework of semantic prosody and semantic preference – how the term 'refugiado' (refugee) has been constructed in two major Spanish newspapers. The research is based on the analysis of a 1.8-million-word corpus of Spanish news articles about refugees that were extracted from the digital libraries of the two most widely read newspapers in Spain: El País and El Mundo. The time span for the articles selection covered 7 years, from 2010 to 2016, coinciding in this way with the years before and after the aggravation of the Syrian refugee movement. The corpus was run with the software WordSmith Tools, which enabled us to examine how 'refugiado' behaved in the corpus and to explore its most frequent co-occurrences. In our study, semantic prosody is understood in diachronic terms and it is then defined as an attached meaning or as a meaning which is transferred from one word to another during the course of time (Stewart, 2010). Semantic prosody – commonly classified as positive, negative or neutral – is to be distinguished from semantic preference, the latter being less abstract as it refers to the "habitual collocation of lexical items with linguistic expressions that belong to certain semantic fields" (Bednarek, 2008: 120). Within this framework, our preliminary results demonstrate that the negative prosody of 'refugiado' increases over the years and it is displayed through frequent co-occurrences that can be categorized into different semantic sets, thus showing a semantic preference of 'refugiado' for items indicating violent occupation and numbers, victimization, and objectification.

References:

Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory*, 4(2), 119-139.

Stewart, D. (2010). *Semantic Prosody: A Critical Evaluation*. London: Routledge.

The Use of Inferential Lexical Bundles in Academic Writing of Expert Writers: A Contrastive Study

Fatih Güngör, Afyon Kocatepe University

Due to the burgeoning growth and preeminence of English as the language of academia and publishing world with the belief that publishing research articles in English-medium prestigious journals indicates the academic productivity and status of researchers, a great deal of attention has been paid to the various features of research article genre such as moves, stance, engagement, metadiscourse markers, and lexical-grammatical choices following the pioneering genre analysis of Swales (1981) into the introduction sections of research articles. Many studies analyzed the certain parts of the research articles in terms of moves, interactional features, tense and lexical-grammatical choices. However, no study has touched upon the analysis of a certain function from a phraseological perspective due to the finding that some lexical bundles are overused by native and non-native scholars to fulfil a certain function. Two overused lexical bundle function are procedure and inferential lexical bundles in both native speaker and non-native speaker writing. This study reveals that the distinctive bundles by Turkish scholars are mostly (15 types-28% of all types and 5484-37% of all tokens) from the inferential lexical bundles category. Inferential lexical bundles are very frequent in the result section of the research articles, and they are of significance in research article genre as they signal interpretations, inferences and conclusions of the relevant data. Therefore, this study aims to investigate a contrastive analysis of inferential lexical bundles (as one of resultative signals) to help non-native scholars in their writing process. Considering the fact that the results and discussion sections of the research articles are challenging for writers due to its driving force to polish a manuscript, the results of this study are expected to contribute the research reporting strategies for undergraduate and graduate students and non-native scholars.

Reference:

Swales, J. M. (1981). Aspects of articles introductions. Aston ESP research reports No. 1. Birmingham: The Language Studies Unit, The University of Aston in Birmingham.

Rhetoric and Reader Pronouns in Economics Academic Writing: A Corpus-Based Contrastive Analysis of English, French and Spanish Research Articles.

Niall Curry, University of Cambridge

Research on the rhetorical functions of reader engagement in academic writing has largely focused on English with fewer studies focusing on languages other than English. Fewer still take a contrastive perspective, comparing the rhetorical features of research articles across languages. Furthermore, although research has identified both the advantage and need for contrastive research on more than two languages, contemporary studies continue to identify contrastive analysis as the comparison of two languages with few taking advantage of the perspective a third language can yield. My research, which centres on the analysis of reader pronouns in English, French and Spanish economics research articles, addresses these gaps by presenting a novel approach to contrastive analysis, synthesising core methods from the fields of contrastive analysis, corpus linguistics and evaluation. In this paper, I present a corpus-based contrastive analysis of reader pronouns in economics research articles in English and French, taken from the KIAP corpus (Fløttum et al. 2006) and a comparable Spanish subcorpus. The reader pronouns are categorised according to their pragmatic functions which act as the shared common ground, comparable across the three languages. Within each function, the pronouns are then contrasted, juxtaposing and analysing them in terms of multiple factors, such as their morpho-syntactic properties, location and formal typology with a view to measure equivalence across the three languages. The results of this study reveal some important similarities and differences which are investigated both quantitatively and qualitatively and have implications for the teaching of English for academic purposes, français langue académique and Español con fines académicos.

Reference:

Fløttum, K., Dahl, T. and Kinn, T. (2006) Academic voices across languages and disciplines, Amsterdam: Benjamins.

A corpus-informed exploration of the creation of intimacy in 'First Dates' (Ireland)

Brian Clancy, Mary Immaculate College, University of Limerick

'First Dates' is a reality TV show, the format of which features people on blind dates set up by the show's production team. The viewer sees the date itself, the participants discussing their previous romantic encounters, and, finally, the couples interviewed together at the end of the date. The show began in the UK and has since been franchised to, amongst other countries, Canada, Australia and Ireland. This paper discusses how the interaction in 'First Dates' (Ireland) can be analysed as a site of intimate discourse, albeit one attended by the complexities of the site itself. While the analysis of this interaction may be complicated by questions of authenticity, the interaction between the potential couples is unscripted and represents a valuable and interesting example of intimate discourse, a discourse type which typically presents more than the usual challenges in terms of data collection. Intimate discourse, that between couples, family and close friends, is ubiquitous yet can be difficult to capture, at least in large quantities. Usually, when this has been done successfully, it features previously established, long-term relationships such as between parents and children. However, the origins of intimacy are often the stuff of anecdote. Reality TV shows, such as First Dates, provide us with access to the beginnings of intimacy. This paper uses a corpus analytic approach to investigate linguistic patterns in these nascent encounters and compares them to naturally-occurring, established intimate discourse (see, for example, Clancy, 2016). Inter alia, questions of authenticity, the nature of performed intimacy and the harnessing of unscripted TV dialogue as a vehicle for linguistic analysis in an intimate context will be addressed.

References:

Clancy, B., 2016. *Investigating Intimate Discourse: Exploring the spoken interaction of couples, families and friends*. London: Routledge.

CorCenCC: applying the sociolinguistics of new speakers within a contemporary corpus of Welsh

Steve Morris, Swansea University
Dawn Knight, Cardiff University
Tess Fitzpatrick, Swansea University

The CorCenCC project (Corpws Cenedlaethol Cymraeg Cyfoes – The National Corpus of Contemporary Welsh) will create the first freely available corpus of the Welsh language, sampled from a range of real life contexts. In considering a representative sampling framework for the corpus, ensuring that all contemporary speakers and users of Welsh be included is a major challenge. In the context of a minoritised language such as Welsh, new speakers (including children from non-Welsh-speaking homes and adults who have acquired the language in later life) form a substantial percentage of the language community and a representative corpus must reflect this. This paper discusses the process by which these new speakers are included, and appropriate samples of their language captured, in the CorCenCC data collection framework. By including new speakers and tagging their data in a systematic way, the corpus will yield important information about the varieties of Welsh used by new speakers – at a lexical, grammatical and phonological level – and our data collection app also allows us to look at how, where and with whom they use Welsh within their communities. We will provide examples of how CorCenCC can inform the analysis of new speakers' language use in this way as well as offering a picture of how this could impact upon future corpus planning in Wales. This will enable us to identify and articulate the key challenges faced by language planners. CorCenCC's principled and robust inclusion of 'new speaker' information not only offers an opportunity for new insight into new speakers' language patterns, it also offers a template which could be employed in similar minoritised language communities to identify and scrutinise the contribution and impact of new speakers. The CorCenCC project is funded by the ESRC and AHRC.

The 'Research Article' event: Towards a more unified dress-code.

Maria N. Melissourgou, Hellenic Ministry of Education, Research and Religious Affairs
Laura Maruster, University of Groningen

Researchers have speculated that the growing international diversity of the writers and the global publishing context (Swales, 2004; Tardy 2016) are likely to affect the 'Research Article', one of the most conservative genres. There is, however, little research that has specifically examined possible instances of innovation. This paper attempts to investigate change in the Research Article over the past fifty years with a focus on informality. Building on work by Hyland & Jiang (2017) it explores three disciplines (Philosophy, Economics and Medicine) at three periods over the 50 past years. Three self-compiled and POS tagged corpora, representing research articles from each discipline and each time period, are analysed with a focus on informality markers used by Hyland & Jiang (2017). The set of features is based on previous work by Chang & Swales (1999). WordSmith Tools (Scott, 2015) is used for the analysis of texts. The results corroborate Hyland & Jiang's findings. It cannot be said that the 'Research Article' in general has become more or less formal/informal due to the observed variation in stylistic choices among different disciplines. The addition of more disciplines, however, and the comparison with previous results, has revealed a significant pattern of change: conventional norms among disciplines move away from the poles of 'soft' and 'hard' sciences and closer to the middle ground.

References:

Chang, Y.Y., & Swales, J. (1999). Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers? In C. Candlin, & K. Hyland (Eds.), *Writing: Texts, processes and practices*. London: Longman.

Hyland, K. & Jiang, F. (2017). Is academic writing becoming more informal? *English for Specific Purposes* 45, 40–51.

Scott, M., (2015). WordSmith Tools version 6, Stroud: Lexical Analysis Software. Available at http://lexi- cally.net/wordsmith/

Swales, J. M. (2004). *Research genres: explorations and applications*. Cambridge, U. K.: Cambridge University Press.

Tardy, C.M. (2016). *Beyond Convention: Genre Innovation in Academic Writing*. The University of Michigan Press.

Language Assessment in Ireland at Senior Cycle and the CEFR: A comparative analysis of assessment instruments across languages examined at Senior Cycle in relation to the CEFR.

Sorcha Ryan, Trinity College Dublin
Ann Devitt, Trinity College Dublin

The present paper is a corpus analysis of assessment instruments providing a comparative analysis across languages examined at the end of senior cycle in Ireland. Various instruments are used for assessment of languages at senior cycle. Students of Italian and Spanish prepare picture sequences and roleplays for the oral examinations and may answer an optional question on a prescribed novel for the written section of the examination, while students of French may prepare an optional document for the oral examination. Chief Examiner's reports on language students' performance in terminal examinations commonly cite an excessive dependency on rote-learning to answer examination questions as a problem (State Examinations Commission, 2009; 2010). The present paper aims to explore the range of higher order and lower order thinking skills that these various modes of assessment measure in language students by drawing on Bloom's revised taxonomy of educational learning objectives (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, & Wittrock, 2001). The study aims to determines the extent to which language examinations at senior cycle in Ireland are aligned with the Common European Framework of Reference for Languages (Verhelst, Van Avermaet, Takala, Figueras, & North, 2009) to measure students' language ability at a certain level (A1 – C2). The level descriptors of the Common European Framework of Reference for Languages (CEFR) are currently used by many European countries as a model for language instruction and assessment best practice. Recommendations are made for Irish Education stakeholders to consider how assessment methods for languages at the end of senior cycle may be more closely aligned with the CEFR. The possibility of a common level paper for Modern Foreign Languages to replace the current Higher and Ordinary Level divisions is also discussed.

References:

Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.

Anderson, L. W., Krathwohl, D. R., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., ... & Wittrock, M. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy.* New York. Longman Publishing.

The pragmatics of poverty in nineteenth century English pauper letters

Ivor Timmis, Leeds Beckett University

This paper discusses how early nineteenth century English paupers, many of whom were only semi-literate, used language for the pragmatic purpose of securing charitable relief. The paper is based on two historical sources: 1) Corpora of letters written by paupers applying for charitable relief 2) The Mayhew Corpus, a corpus of interviews with the destitute of London carried out by Sir Henry Mayhew in the 1850s. The paper focuses on certain grammatical differences between the language of the pauper letters and the language in the Mayhew Corpus. From this analysis, it emerges that the pauper writers made markedly less use of certain vernacular features than speakers in the Mayhew Corpus did. The features not used to any great extent in the pauper letters but present in the Mayhew Corpus are: vernacular relative pronouns (as and what), vernacular preterites and past participles; a-prefixing and non-standard verbal 's' ending. I argue that the infrequency of these features in the pauper letters indicates that the pauper writers were orienting towards the emergent notion of Standard English or at least to what they regarded as prestige forms. However, in contrast to this argument, we find that multiple negation, a low prestige vernacular feature, occurs with similar frequency in both the pauper letters and the Mayhew Corpus. The main argument of the paper, in the light of this apparent contradiction, is that, in some cases, the pauper writers' attempts to orient towards prestige forms faltered as they were dealing with the critical issues of health, welfare and money, the very issues which were at the heart of their letters.

"Box, Box Now": A Corpus-Based Analysis Of Formula One Radio Messages

Jukka Tyrkkö, Linnaeus University
Hanna Limatius, University of Tampere

Formula One is widely regarded as the pinnacle of all motor racing categories. With the highest team budgets, the most sophisticated automotive technology and an estimated worldwide television audience of more than 400 million for each Grand Prix, Formula One is one of the world's premier spectator sports (see Real 2012, Billings et al 2014, Caldwell et al 2016). During each GP, the drivers and pit crews communicate via team radio, and parts of their dialogues are publicly broadcasted. Linguistically, these messages present an interesting case of spoken performance under stressful conditions. We have compiled a corpus comprising transcripts of 5,500 individual radio messages, or all messages played on official broadcasts during the 2016 and 2017 seasons of Formula One; see Collantine (2013–) and Authors (in prep). Our analysis of the lexical and grammatical features of the messages will describe a clipped and syntactically reduced spoken variant. Taking into account individual variation and the stage of the race, we apply corpus-based quantitative and qualitative methods to investigate turn length, sentence length, syllable count, lexical density, phrasal complexity, the use of deictic and non-deictic references and stock expressions that are used by the drivers and race engineers to increase information density and ensure clear messaging.

References:

Authors. In prep. "When did I do dangerous driving then?": Structures and functions of Formula One race radio messages. In Callies, Marcus and Magnus Levin (eds.) (Multimodal) *Corpus Approaches to the Language of Sports*.

Billings, Andrew C. and Marie Hardin (eds.) 2014. *Routledge Handbook of Sport and New Media*. London and New York: Routledge.

Caldwell David, John Walsh, Elaine W. Vine and Jon Jureidini (eds.) 2016. *The Discourse of Sport: Analyses from Social Linguistics*. New York: Taylor and Francis.

Collantine, Keith. 2013–. Team radio transcript archive. On F1 Fanatic blog. Available online at <http://www.f1fanatic.co.uk/category/regular-features/team-radio-transcripts/>

Real, Michael. 2012. Theorizing the sports-television dream marriage: Why sports fit television so well? In Billings, Andrew C (ed.) *Sports Media: Transformation, Integration, Consumption*. Amsterdam: John Benjamins. 19–39.

Exploring disciplinary differences in academic writing using semantic tags

Michael Barlow, University of Auckland
Phuong le, University of Auckland

In this study we investigate differences in the language use and conceptualisation in research articles (RAs) in Applied Linguistics (Language Testing and SLA) and Mechanical Engineering (Fluid Mechanics and Heat Transfer). A sample of 20 RAs for each discipline are divided into Introduction, Methodology and Results-Discussion-Conclusion (RDC) sections and are annotated with semantic tags (USAS) using the Wmatrix website. The semantic tags categorise broad semantic concepts such as time, evaluation, cause/effect, thought/belief, etc. Using a concordancer, the tags can be counted and the associated words identified. Some results are obvious and relate to the content. However, it is possible to observe more subtle differences between the two disciplines and among the different sections. For example, focussing on the distribution of cause/effect lexis, we find, , that most occurrences occur in the Introduction and RDC sections and are quite common in all the sub-disciplines apart from SLA. If we examine particular lexical instantiations, we see a preferences for due to and induce, terms indicating a fairly direct connection between cause and effect, in Mechanical Engineering, as in example (1). (1) The high-temperature zone exhibited near the fuel and oxygen inlets is due to the exothermic coal combustion. On the other hand, the more indirect form, influence, as a verb or noun, is more prevalent in Applied Linguistics. A typical example is given in (2). (2) …. mother tongue, previous experience, and tolerance for errors have been found to influence the assessment of ESL writing. By investigating the distribution of semantic tags, such as cause/effect and thought/belief, along with their particular lexical instantiations, we explore differences in the conceptualisation and construction of RAs in the two disciplines.

Developing a CEFR-based and corpus-informed curriculum for English for Academic Purposes

Stergiani Kostopoulou, University College Dublin

Student mobility in secondary and tertiary-level education is a global phenomenon which poses a number of challenges for students and educators alike. Developing the language that is required for academic achievement appears to be one of the greatest difficulties faced by students for whom the language of the host educational system is a second language. In response to the need to support students' academic language development, schools and universities provide language support in various forms. The teaching content (curriculum) of these language support programmes must be rigorously defined to ensure that it accurately reflects the specific language learning needs of learners. The descriptive scheme of language use of the Council of Europe's Common European Framework of Reference for Languages (CEFR) is commonly used as a resource for curriculum development. The learners' target communicative repertoire can be described in the form of functional ('can do') statements of behavioural capacity embedded in specific domains of language use. Valuable as they may be, these descriptions do not capture the precise linguistic resources that learners need to acquire in order to perform the specified communicative tasks. This presentation demonstrates how the analysis of domain-specific language corpora can play an indispensable role towards achieving this aim. The paper reports on a corpus-based study which was motivated by the need to enhance the content of a CEFR-based curriculum that was developed for the English language support of immigrant students in Irish post-primary education. Firstly, the curriculum in question is briefly presented. Secondly, the development and analyses of 12 subject-specific language corpora are discussed. Then the wide range of quantitative and qualitative data are presented, illustrating how these enriched curriculum content and informed materials design. This empirical research represents the first of its kind in the Irish context and it responds to the calls for applied corpus linguistics research specifically into the language of secondary education to support students in their school-based studies (e.g. Coxhead, 2010: 466). It may also be argued that the proposed approach to informing curriculum content can be adopted in any other EAP context.

Reference:

Coxhead, A. (2010) What can corpora tell us about English for Academic Purposes? In McCarthy, M. and A. O'Keeffe (Eds.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge

'Sure where else is there? 'Pragmatic markers in radio advertising in Ireland

Joan O'Sullivan, Mary Immaculate College, University of Limerick

'Sure where else is there?' Pragmatic markers in radio advertising in Ireland The attempt by advertisement producers to replicate 'normal everyday communication' through the exploitation of language variation has been highlighted (Lee 1992; Kelly-Holmes 2005). More specifically, O'Keeffe (2006: 97) examines how pronouns, vocatives and pragmatic markers are used in media discourse 'to create an illusion of an interpersonal relationship between strangers by drawing on linguistic features that are normally used in casual conversation between real friends and intimates'. Schiffrin (2001: 67) observes how, in addition to their pragmatic and semantic functions, pragmatic markers (PMs) can provide information on 'the cognitive, expressive, social and textual competence of those who use them', suggesting an additional indexical function for PMs to be exploited in advertising. This study investigates how and to what extent PMs are employed in radio advertising through analysing a corpus of ads broadcast on an Irish radio channel between 1977 and 2017. PMs are examined in terms of their frequency and functions in the distinct components of the ad, the 'Action' (comprised of context-based dialogic interaction, designed to imitate discourses of 'everyday informal interaction' (Lee 1992:172-3)) and 'Comment' (commonly monologic, decontextualised and associated with the slogan or voice of authority (Sussex 1989)). The extent to which the use of PMs in the ad components reflects natural, unscripted discourse in the Irish context is investigated by comparing them with naturally occurring data in the Limerick Corpus of Irish English (LCIE). PMs are also examined as regards the extent to which their indexical function supersedes their semantic or pragmatic function.

References:

Kelly-Holmes, H. (2005) *Advertising as multilingual communication*, London: Palgrave Macmillan.

Lee, D. (1992) *Competing discourses: Perspective and ideology in language*, London: Longman.

O'Keeffe, A. (2006) *Investigating media discourse*, London: Routledge.

Schiffrin, D. (2001) 'Discourse marker: Language, meaning and context' In Schiffrin, D., Hamilton, H.E. and Tannen, D. eds., *Handbook of Discourse Analysis*, Malden, MA: Blackwell, 54-75.

Sussex, R. (1989) 'The Americanisation of Australian English: Prestige models in the media'. In Collins, P. and Blair, D., eds., *Australian English: The language of a new society*, Cambridge: Cambridge University Press, 158-170.

Language, mobility and climate change at the crossroad: A diachronic corpus-assisted discourse analysis of the climate-migration nexus in the press

Cinzia Bevitori, University of Bologna
Jane Helen Johnson, University of Bologna

While climate change has gradually and crucially become a 'defining symbol of our collective relationship with the environment' (Boycoff 2011: 1), an investigation of its role in complex, multi-causal phenomena of human mobility has only recently emerged as a salient policy-making issue (e.g. Bettini et al. 2016, Bettini 2017). This paper explores representations of migration and displacement in the context of anthropogenic climate change in newspaper discourse from a diachronic corpus-assisted discourse analytical perspective (Partington et al. 2013; Bevitori 2010, 2014).

For the purpose of this study, a domain-specific corpus of newspaper articles has been gathered from a selection of UK and US broadsheets dating from 2010 to 2017, to coincide with UNFCCC negotiations in Cancun, Paris and Marrakesh. Particular attention will be placed on whether any significant discursive shifts may be identified in newspaper discourse over the periods concerned to coincide with a change in focus from 'climate refugees' to migration as adaptation (Bettini et al. 2016; Bose 2016).

References:

Bettini, G., 2017, "Where Next? Climate Change, Migration, and the (Bio) politics of Adaptation", *Global Policy*, 8 (1), pp. 33–39.

Bettini, G., Nash S.L., and Gnoli, G., 2016, "One step forward, two steps back? The fading contours of (in)justice in competing discourses on climate migration", *The Geographical Journal*, doi: 10.1111/geoj.12192

Bevitori, C., 2014, "Values, Assumptions and Beliefs in British Newspaper Editorial Coverage of Climate Change, in C. Hart and P. Cap (eds.) *Contemporary Critical Discourse Studies*, London: Bloomsbury Academic, pp. 623-625.

Bevitori, C., 2010, *Representations of Climate Change. News and opinion discourse in UK and US quality press: A Corpus-Assisted Discourse Study.* Bologna: Bononia University Press.

Bose, P. S ., 2016, "Vulnerabilities and displacements: adaptation and mitigation to climate change as a new development mantra", *Area*, 48 (2), pp. 168–175.

Boykoff, M. T., 2011, *Who Speaks for the Climate? Making Sense of Media Reporting of Climate Change*. Cambridge: Cambridge University Press.

Partington, A., Duguid, A. and Taylor, C. 2013, *Patterns and meanings in discourse. Theory and practice in Corpus-Assisted Discourse Studies*. Amsterdam/Philadelphia: John Benjamins.

Teaching the "unlearnable" - A methodology for the use of corpora for Hungarian as a foreign language

Szilvia Szita, Sorbonne-Inalco Paris / University of Pécs

It is a common trait of all languages that some of their properties cannot be expressed by simple, yet comprehensive rules. Many such linguistic features emerge relatively early in the learning of Hungarian. Among these, one can list the use of verbs with and without prefix, the definite and indefinite conjugations and the use of cataphoric elements. Contrarily to the teachers of English, the vast majority of Hungarian teachers are native speakers. As such, they have an implicit, practical knowledge about the use of the language – a knowledge that has been acquired by numerous encounters. When presenting the above-mentioned linguistic features, they mainly draw on their intuition and either provide approximative guidelines for their use or formulate sets of rules that are so complex that they only serve to add to the learners' despair. The major deficiency of this approach is obvious. It lacks the stage in which learners have a chance to observe how the linguistic item in question actually works. This stage appears particularly relevant whenever a rule-based approach fails to provide clear, convincing answers. We claim that in these cases, learners must first acquire reliable knowledge about the linguistic items by studying them in authentic examples. Only after building up an extensive experience can a rule-based description be realised as a summary of the foregoing experience. To share good teaching practices for Hungarian as a foreign language, the International Studies Center of the University of Pécs (Hungary) organises a one-week workshop on methodology every year. One entire day is dedicated to the use of corpora to present linguistic items that have the reputation of being "unlearnable". According to our experience, most teachers are either unaware of the benefits of corpora or consider corpus analysis to be an extremely time-consuming activity that represents, above all, extra workload. Trainers must therefore give proper guidance and propose activities that make the benefits of using corpora very clear from the outset. After giving some background information, I will present a selection of such activities and a series of supplementary materials that have been proved to be effective in guiding teachers step by step towards the integration of corpus-based activities into their lessons. We believe that only teachers who are deeply convinced of the utility of corpora are capable of exploring them in meaningful ways. Guided by such teachers, students can, in turn, become autonomous corpora users themselves.

What "should" happen in a Modern foreign language class? A corpus-based analysis of secondary level Inspectorate reports in Ireland

Stéphanie O'Riordan, University of Limerick

In the post-2008 economic crisis era, increased expectations regarding accountability and transparency in all spheres of public services have strengthened the need for school inspections to be carried out and for subsequent Inspectorate reports to be published by the Department of Education in the Republic of Ireland (Sugrue, 2006). Such documents represent significant resources for all education stakeholders to review teaching standards in specific schools at a particular point in time. Thus, investigating a corpus of inspectorate reports is relevant not only to extract and analyse the linguistic features of this specific discourse genre but also to discover to what extent they can be used as pedagogic resources in teacher education. This paper presents a corpus-based analysis of Inspectorate reports in the area of Modern foreign language (French, German, Italian and Spanish) teaching and learning in secondary level education in Ireland. After a brief summary of the corpus construction methodology, the data analysis will be presented based on the significant frequency of the modal "should". In particular, the expression of directive language exemplified by the use of "should" will be the primary focus of this study. Although the corpus will be investigated primarily on the basis of quantitative criteria using Antconc (Anthony, 2014) qualitative analysis is necessary in order to extract and categorise the key pedagogic features of French teaching practices which were discussed in the reports. The contention of this research study is that corpus-based methodology is immensely valuable to draw an inventory of specific teacher needs with a view to informing language teacher education.

References:

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Sugrue, C. (2006). A Critical Appraisal of the Impact of International Agencies on Educational Reforms and Teachers' Lives and Work: the case of Ireland? *European Educational Research Journal*, 5 (3 & 4), pp. 181-195.

"Hey, can I borrow your…": Is sharing your corpus linguistic tools with others a good idea?

Jane Evison, University of Nottingham
Elaine Vaughan, University of Limerick

Despite criticisms of the use of semi-structured interviews in research, they remain common forms of data generation in education and the social sciences more broadly. In this presentation, we build on the arguments of others that interviews should be understood from an applied linguistic perspective as socially situated speech events with all their accompanying contextual 'baggage'. We do this by combining corpus and pragmatic techniques and perspectives. Specifically, we apply a personal pronoun framework to a corpus of interviews with practising language teachers. This framework was previously used as part of the pragmatic analysis of a corpus of English language teacher meeting talk in which groups of language teachers go about the business of getting language teaching done in their institution. In the current corpus, the business of teaching is reported rather than immediate, but many of the issues dealt with remain the same. However, there is an added layer of complexity in the interviews. The physical location of the participants varies, as does the mode of the interview (face-to-face and online). The relationships between the interviewers and interviewees are multilayered, and, in comparison with the meetings, there is less shared knowledge. The results of our analysis show that as teachers and their interviewers construct and negotiate pragmatic meanings that position themselves and their practice, they do so in observable ways which derive from the complexities of the interview setting. By treating the interviews as a collection of texts which can be subjected to corpus analytical techniques, we are able to understand some of the potential issues with interview data that are a direct result of the pragmatic affordances of the event. We also raise epistemological issues for social science researchers who want to add corpus techniques to their analytical toolkit.

Corpora for Remembering and Learning: The Second Amendment, the 2008 Supreme Court Ruling, and the Absolute Phrase

James Vanden Bosch, Calvin College, Grand Rapids, MI

The 2008 Supreme Court ruling concerning the meaning of the Second Amendment of the US Constitution provided an analysis of the relationship of the absolute phrase or absolute construction (in italics, below) that introduces the amendment to the main clause that follows it: *A well regulated Militia, being necessary to the security of a free State*, the right of the people to keep and bear Arms, shall not be infringed. [This is the form of the amendment as approved in 1792.] According to the majority opinion, written by Justice Scalia (District of Columbia v. Heller, 2008), that absolute phrase "does not limit or expand the scope of the operative clause." The Court's claim regarding the function of the absolute phrase is based on two significant problems—forgetting and not knowing. It has forgotten the long history of the absolute construction in English, particularly in the late 18th century, as well as its use in Supreme Court rulings; and it apparently does not know the current scholarship on the absolute construction, especially on its semantic variability. Corpus studies can provide the historical information that the ruling has forgotten, and corpus resources have already provided evidence of the construction's semantic variability, therefore allowing legal scholars and politicians to consider the limitations of the grammatical and semantic argument put forward by the 2008 ruling.

Notes

1. The history of the absolute construction in English:

- Timofeeva, O. (2010). "Non-finite constructions in Old English with special reference to syntactic borrowing from Latin." Helsinki: Société Néophilologique.

2. The Corpus of US Supreme Court Opinions (BYU, 2017).

3. Current work on the absolute construction:

- Fonteyn, L., & van de Pol, N. (2015). "Divide and conquer: the formation and functional dynamics of the Modern English ing-clause network." *English Language and Linguistics*, available on CJO2015. doi:10.1017/S1360674315000258;
- Kortmann, B. (1991). *Free absolutes and adjuncts in English: Problems of control and interpretation*. London and New York: Routledge;
- Stump, Gregory (1985). *The Semantic Variability of Absolute Constructions*. Netherlands: Springer.

From cross-cultural to superdiverse corpus linguistics: language ideologies and communities

Rachelle Vessey, Birkbeck University of London

Since corpus linguistic approaches can pinpoint patterns and are inherently comparative (Stubbs, 2001), they are useful in the study of language ideology in that they help to identify ideologies that are both explicit (i.e. manifest, often in language representations) and implicit (i.e. infrequent, unobservable, implied or absent) (Vessey, 2017). At the same time, corpus linguistic approaches can help researchers compare and contrast language data linked to specific communities, providing insight into local language ideologies. The problem emerges that traditional notions of community are being reconsidered in recent sociolinguistic work and new concepts are more difficult to align with existing corpus methods and theoretical conceptualisations of the "corpus". At this juncture, it becomes crucial to revisit the existing methods and the language principles and values that underpin corpus construction.

Drawing on examples from a range of different contexts, I show how (1) language ideologies underpinning the notion of "corpora" can compromise sociolinguistically oriented research and (2) not all mainstream corpus methods are suitable for sociolinguistic research. At the same time, I maintain that corpora and corpus methods can be powerful means by which we can explore important new dimensions of language ideologies and communities. The paper concludes by arguing that a closer alignment between corpus linguistic and sociolinguistic research will be mutually beneficial.

References:

Stubbs, M. (2001) *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Vessey, R. (2017). Corpus approaches to language ideology. *Applied Linguistics, 38* (3), 277-296.