



The University of  
**Nottingham**

UNITED KINGDOM • CHINA • MALAYSIA

# ICAME 35

## Corpus Linguistics, Context and Culture

Wednesday 30th April - Sunday 4th May 2014



**Book of Abstracts**  
hosted by CRAL, Department of English

# **ICAME 35**

30 April - 4 May 2014

## **Conference Organisers**

Michaela Mahlberg (chair)  
Gavin Brookes, Kathy Conklin, Rachele de Felice,  
Dave Evans, Kat Gupta, Kevin Harvey, Tony Fisher,  
Lorenzo Mastropiero, Rebecca Peck, Anna Pellicer-Sánchez,  
Tracy-Ann Stead, Viola Wiegand

## **Student Helpers**

Ylva Biri, Jacqueline Cordell, Samuel Clarke, Fiona Lake,  
Michele McIntosh, Katherine Norman, Jonny Pryn, Shatha Qaiwer,  
Stephanie Sewell, Lucia Scigliano Suarez, Eman Yonos

## **Student Editors**

Sam Clarke, Jonny Pryn, Stephanie Sewell

Conference bag design by Ana Jolly

## Contents

Opening talk and keynote speakers.....	3
Pre-conference workshops .....	10
Full papers .....	17
Work-in-progress reports .....	91
Posters .....	126
Software demonstrations .....	138
List of participants .....	142

# Opening talk and keynote speakers

## Opening talk and keynote speakers

### **Opening Talk: Corpus, context, culture: histories and futures**

*Ronald Carter*

*University of Nottingham, United Kingdom*

This opening talk involves a review of some key developments in corpus linguistics over the past 40 years and of what we might learn from them in both theory and practice for future developments of the field. The principal focus is on the main themes of the conference and examples are given which suggest how corpus linguistics can continue to address issues of interdisciplinary applications and involve other fields of study --- with particular reference to language learning, stylistics and professional and internet communication. Issues of context and culture will be explored with reference to a spoken corpus data and to a recent research collaboration between Nottingham University and Cambridge University Press in the development of an e-language corpus (the CANELC corpus) where (sometimes multimodal) language use is marked by forms associated with both spoken and written grammars. Exploration of data from this and other corpora suggests the need for a fuller recognition of key notions of 'emergent' cultures, 'dynamic' contexts and mixed research methods appropriate to different data streams.

# Opening talk and keynote speakers

## Place-making in Brooklyn, New York

Beatrix Busse

University of Heidelberg, Germany

In the model of urbanity (Busse and Warnke 2014), language and other semiotic modes function as parameters of both urban variation and of processes of urban variational place-making to create identity, belonging, attachment and place. These discursive practices are always historically motivated and mobile, and their linguistic patterning and social styling are highly declarative. In this paper, my focus will be on selected – so far mostly gentrified - neighbourhoods in Brooklyn, New York. I shall investigate contemporary and historical linguistic and multi-modal means of reflecting on people's sense of belonging as place-making activities (Cresswell 2004, Friedmann 2010) in Park Slope, Williamsburg and Brooklyn Heights. I will show when, how, and why not just dialect features, but practices on all levels of language as well repetitive linguistic and semiotic patterns index, that is, "enregister" (Johnstone 2009) social value and construe Brooklyn as a brand. These include, for example, comparisons and negation and range around interplaying themes like a generic reference to Brooklyn as a whole, a contrast between Brooklyn and Manhattan as well as a focus on aspects of home and on art. Methodologically, I will show that investigations of urban place-making are in need of (more) corpus-assisted approaches and, at the same time, should combine these with qualitative methodologies which also use more disparate linguistic practices or multimodal artefacts as their objects of analysis. Therefore, data to be analysed will be a corpus of semi-structured interviews I conducted with Brooklynites in February 2012 and – among others – narrative fiction, (historical) newspaper discourse and the semiotic landscape in these neighbourhoods.

## References

- Busse, Beatrix and Ingo H. Warnke. 2014. "Sprache im urbanen Raum – Konzeption und Forschungsfelder der Urban Linguistics. In: *Handbuch Sprachwissen* Ed. Ekkehard Felder and Andreas Gardt. Berlin: de Gruyter Mouton.
- Cresswell, Tim. 2004. *Place: A Short Introduction*. Oxford: Blackwell, 2004.
- Friedmann, John. 2010. "Place and Place-Making in Cities. A Global Perspective." *Planning Theory & Practice* 11.2: 149-165.
- Johnstone, Barbara. 2009. "Pittsburghese Shirts: Commodification and the *Enregisterment* of an Urban Dialect." *American Speech* 84.2: 157-175.

# Opening talk and keynote speakers

## The Contexts and Cultures of Interdisciplinary Research Discourse

*Susan Hunston*

*University of Birmingham, United Kingdom*

Academic discourse has long been a key focus of research in corpus linguistics as well as other forms of discourse analysis, and for good reason. There is a practical value in investigating a form of language that is crucial for students and professional users of English. Studies of academic discourse interrogate and demonstrate the link between language, culture and context. Academic discourse contributes to the social construction of knowledge. For the most part, studies have focused on distinct academic disciplines, for example contrasting features associated with stance. In the study reported in this paper we focus on interdisciplinary (ID) research discourse. We see this as an area of importance because ID research has a high priority in affecting societal change, yet ID discourse is often reported as constituting a site of difficulty. ('We talk different languages.') Aside from this utilitarian value, investigating ID texts, as opposed to those focused on a single discipline, has the potential to shake things up, methodologically and conceptually. Whereas there are established ways of comparing disciplinary discourses, methods of looking at an ID field are open to innovation. This paper presents three aspects of our study so far, and considers the contribution these might make to the broader study of discourse using corpus linguistics. The first is corpus design, including the relationship between a corpus and a community of practice, and the role of an interdisciplinary journal in constructing that community. The paper considers the usefulness of a topographic as opposed to a typographic representation of disciplines. Secondly, the paper considers the question of sameness or change in meaning, using phrases including the wordform *change* as an example and considering to what extent this is a contested site. Finally, the paper examines aspects of stance marking in selected texts.



## Opening talk and keynote speakers

### **The Corpus as Social History - Prostitution in the Seventeenth Century**

*Tony McEnery*

*University of Lancaster, United Kingdom*

To what extent can corpora illuminate the past? Can the close readings traditionally associated with the study of history gain from the tools and methods of corpus analysis? In this talk I will talk about work I have undertaken with historian Dr. Helen Baker looking at an issue in social history - prostitution in seventeenth century England. Social history in particular represents an interesting topic where the corpus might contribute - while the documentary sources and analyses associated with major historical events and figures are typically many and well analysed, the documents associated with the everyday, the unexceptional, are more sparse. In the case of marginalised or criminalised groups the documentary evidence outside of court proceedings is widely scattered and typically indirect. Prostitutes (we deal only with female prostitutes in this talk) are a good example of such a marginalised group - indeed in such a case the marginalization is enhanced by class and gender as well as criminality. I begin by considering what social historians have claimed about prostitution in this period. I then move to look at what the corpus shows us, using the latest version of the EEBO corpus available at Lancaster University - 1.5 billion words of lemmatised, POS tagged and spelling regularised written texts from the 15th, 16th, 17th and 18th centuries. Using corpus techniques to explore the texts, I show what a corpus may show the historian about prostitution in the period - and what historians can offer to corpus linguists who are approaching texts from this period.



# Opening talk and keynote speakers

## Corpus research for SLA: The importance of mixing methods

Ute Römer

Georgia State University, USA

Over the past few decades, the growing availability of native speaker and learner corpora has enabled Second Language Acquisition (SLA) researchers to study patterns in the linguistic input of learners as well as in their language output in a more empirical and systematic way than previously possible. Corpus linguists have contributed considerably to a better understanding of central aspects of second language (L2) learners' production and of differences between learner and native speaker English (see, for example, the influential work of Sylviane Granger and her team, and of Stefan Gries and Stefanie Wulff). This talk argues that corpus linguistics has a lot to offer to research and practice in SLA, especially if different methods and data types are combined in a "methodological pluralism" sense (McEnery & Hardie 2012). The talk also suggests that progress in corpus-based SLA research will depend to some extent on successful collaborations between corpus linguists and scholars from other fields. After a brief overview of some existing uses of corpora in SLA, the talk will present findings from two case studies that benefited from mixing methods and from the presenter's collaboration with researchers from neighboring disciplines, including a computational linguist, a genre expert, a psycholinguist, and a cognitive linguist. While not based on learner output, the first study examines language that can serve as a model to learners and has clear implications for second language teaching practice (see Wulff, Römer & Swales 2012). It combines quantitative and qualitative approaches to the distribution of attended and unattended *this* in successful student writing across disciplines and determines what learners need to know about *this* in this particular writing context. The second study examines the role that constructions play in second language acquisition. It uses data from a native speaker corpus, learner corpora, and psycholinguistic experiments to investigate what influences L2 learners' acquisition and processing of English verb-argument constructions (see Ellis, O'Donnell & Römer 2013; Römer, O'Donnell & Ellis Forthcoming; Römer, Roberson, O'Donnell & Ellis 2014). The talk discusses the implications of both studies before closing with thoughts on desiderata and future avenues for corpus-based SLA research.

## References

- Ellis, N. C., M. B. O'Donnell & U. Römer. 2013. Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning* 63(Supp. 1): 25-51.
- McEnery, T. & A. Hardie. 2012. *Corpus Linguistics. Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Römer, U., M. B. O'Donnell & N. C. Ellis. Forthcoming. Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions: Exploring corpus data and speaker knowledge. In: Nicholas Groom, Maggie Charles & Suganthi John (eds.). *Corpora, Grammar and Discourse: In Honour of Susan Hunston*. Amsterdam: John Benjamins.
- Römer, U., A. Roberson, M. B. O'Donnell & N. C. Ellis. 2014. Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. *ICAME Journal*.
- Wulff, S., U. Römer & J. M. Swales. 2012. Attended/unattended *this* in academic student writing: Quantitative and qualitative perspectives. *Corpus Linguistics and Linguistic Theory* 8(1): 129-157.

## Opening talk and keynote speakers

### Building onto the corpus-driven approach: a wider look on meaning

Wolfgang Teubert

University of Birmingham, United Kingdom

What makes the corpus-driven approach stand out in language studies is its appeal as a 'scientific' methodology. Using computational tools to identify, count and measure real language data, we obtain dependable findings. Scientific practice, however, is no different from any other social practice: it is discursively constructed. In the absence of a 'real' fundament, there cannot be a 'true' bottom-up approach. All corpus research presupposes a consensus on the arbitrary decisions underlying our research question, and the findings obtained have to be interpreted to make sense. Meaning is found only in discourse. In my investigation of the discourse object 'human rights,' I will move from ambiguous collocation profiles to what texts actually say about this object by assigning a meaning to this lexical item. The meaning of *human rights* is, as I see it, the entirety of what is said about this lexical item, i.e. of all the paraphrases we find in discourse. Yet what counts a paraphrase is a matter of interpretation. The corpus-driven approach offers candidates we can accept or reject. The study of paraphrastic content is thus a necessary extension of traditional corpus linguistics. It combines a methodological approach with an interpretive endeavour that is free from methodical constraints. In order to make sense of *human rights* in a specific text, or text segment, we have to uncover its intertextual links, thus revealing how it differs from what has been said before. We will not understand what *human rights* means in a specific context unless we have analysed those links. Texts thus can be seen as the nodes of dynamic, diachronically evolving networks held together by intertextual links. Therefore corpus linguistics has to concern itself with the diachronic dimension of discourse if it is to pave the way for interpreting what a given text (segment) or a lexical item means. Again, there is no method to capture intertextuality – it is up to the arbitrary decisions of an interpretive community. Language is not a natural phenomenon; it is a cultural artefact. Linguistics, including the corpus-driven approach, belongs to the human sciences.

# **Pre-conference workshops**

## Pre-conference workshops

### Cross-linguistic perspectives on verb constructions

*Signe Oksefjell Ebeling, Hilde Hasselgård*  
*University of Oslo, Norway*

The workshop is devoted to cross-linguistic perspectives on verb constructions. Much corpus-based contrastive work has already been done on various aspects of the verb phrase, including tense/aspect/modality constructions, multi-word verbs, verb complementation and phraseological patterns of verbs. By focusing on the (extended) verb phrase, we hope that the workshop will illuminate this area by means of comparisons with different languages and with different corpora. The workshop includes corpus-based studies comparing English with a range of languages, including Czech, Lithuanian, Norwegian, Spanish and Swedish. Topics covered are complementation of the verb (infinitival complements, the V +*ing* construction, and object omission), past time reference, voice, and semantic composition. There will be ample time for discussion.

Title	Speaker
Welcome	Signe Oksefjell Ebeling & Hilde Hasselgård
English translations of Norwegian infinitival complement constructions	Thomas Egan
Looking for the difference in our silences: A corpus-based approach to object omission in English and Spanish	Tania de Dios
Past-referring verb constructions in English, Norwegian and German: a contrastive look	Johan Elsness
Cross-linguistic perspectives on the verbs of putting in English and Swedish: Contrasts in construction and semantic composition	Åke Viberg
Evidential passive constructions in English and their equivalents in Lithuanian	Anna Ruskan
Come running: the construction 'lexical verb + -ing participle' in a contrastive perspective	Markéta Malá
Discussion and closing of workshop	Signe Oksefjell Ebeling & Hilde Hasselgård

# Pre-conference workshops

## Corpus-based approaches to discourse relations

*Kerstin Kunz, Ekaterina Lapshinova-Koltunski*  
*University of Heidelberg, Germany / Saarland University, Germany*

In recent years, more innovative methods of corpus-based analysis are used to investigate relations on the level of text or discourse. The workshop is organised with a focus on these new approaches and will include methods of corpus exploitation, e.g. (semi-)automatic identification of discourse relations, as well as new methods of statistical evaluation (e.g. multivariate statistical analysis). The papers presented will report on finalised studies or on work in progress. They will cover a variety of discourse phenomena such as cohesion and coherence, information structure and information packaging. The variation investigated will concern intra-lingual variation between registers and between written and spoken language, contrasts between several languages and differences between translations and originals. The workshop will end with a final discussion. It will mainly revolve around the question of how different aspects of discourse can be analysed in combination, both from the perspective of data exploration and data interpretation:

1. Which categories have to be defined? How many levels of analysis are needed? Which relations between which elements are concerned?
2. What are the dependent and what are the independent variables?
3. Which tools are best for semi-automatic extraction? Which tools can be combined?

Title	Speakers
Anaphoric Expressions in Parallel Czech-English Data	Michal Novák and Anna Nedoluzhko
Local Coherence by Abstract Anaphors - a Comparative Survey	Heike Zinsmeister
Discourse-level Features for Statistical Machine Translation	Thomas Meyer
Language Means Expressing Discourse Relations in Written and Spoken Czech	Magdaléna Rysová
Discourse Organization: Genre, Coherence, and Lexical Cohesion	Ildikó Berzlánovich
Diachronic register diversification in the use of interpersonal features in theme position	Stefania Degaetano-Ortlieb
Global discourse structures, rhetorical relations and thematic patterns in English and Spanish journalistic texts: a comparative study	Julia Lavid and Lara Moratón

## Pre-conference workshops

### Perfect and perfectivity re-assessed through corpus studies

*Elena Seoane, Cristina Suárez-Gómez, Valentin Werner*

*University of Vigo, Spain / University of the Balearic Islands, Spain / University of Bamberg, Germany*

While the area of the present perfect has always been a hotly contested ground, recent corpus-analyses have shown that grammatical variation in this realm in English is far more pervasive than has been assumed. This is particularly true when non-native and learner Englishes are taken into account (cf. Davydova 2011; Elsness 2009a, 2009b; Hundt and Smith 2009; Miller 2000, 2004; Seoane and Suárez-Gómez 2013; Suárez-Gómez and Seoane 2013; van Rooy 2009; Werner 2013; Yao and Collins 2012). These studies have addressed the issue from diverse theoretical perspectives and methodologies and using different approaches, both function-to-form and form-to-function, in an attempt to account for the envelope of variation under scrutiny. This workshop will include presentations on synchronic, diachronic and contrastive analyses on the expression of the present perfect and the perfective in both native and non-native varieties of English, which go beyond the traditional ascription of the perfect to the construction *have* + past participle for the expression of perfect meaning (Quirk et al. 1985: 192-195; Biber et al. 1999: 467; Huddleston and Pullum 2002: 143). The presentations will also observe variation determined by sociolinguistic variables such as mode, dialect, style, register, genre as well as linguistic factors, such as type of adverbial or the temporal distance and frame.

Title	Speaker
On the perfect-evidential-link in Continental Scandinavian	Björn Rothstein
The perfect in English-lexifier pidgins and creoles	Stephanie Hackert
The frequency of the present perfect in varieties of English around the world	Robert Fuchs
The present perfect in New Englishes: Common patterns in the situations of language contact	Julia Davydova
The sociolinguistics of the Australian English innovative present perfect: Methodological considerations	Sophie Richard
The perfect form and its meanings in Black South African English	Bertus van Rooy
Constraints on the use of the perfect in Ghanaian English: A comparative sociolinguistic approach	Gloria Otchere
The present perfect and perfectivity and Irish Standard(ised) English: A re-assessment	John Kirk
Perfects and perfectivity in two contact varieties of English	Markku Filppula
The impact of migration on the tense and aspect system of nineteenth-century Irish English correspondents	Marije Van Hattum
Speakers have a construction used – Speakers have used a construction: The construction HAVE + past participle in Old English – resultative or perfect?	Berit Johannsen

## Pre-conference workshops

### Statistics in corpus-based sociolinguistics: A practical workshop

*Vaclav Brezina*

*Lancaster University, United Kingdom*

This workshop will discuss different statistical procedures available for analysis of sociolinguistic data in large language corpora. I will demonstrate that the traditional approach of using aggregated data with the log-likelihood statistic is in principle unreliable. Instead, the workshop will offer suggestions for alternative methodologies and statistical procedures, which take into account within group differences and therefore produce more meaningful results. As part of the workshop, a new research tool BNC64 Search & Compare will be introduced. BNC 64 Search & Compare can carry out detailed analyses based on a socially-balanced spoken corpus BNC64 (1.5 million words). BNC 64 represents the speech of 64 speakers – 32 men and 32 women – extracted from the British National Corpus (BNC). BNC64 Search and Compare is a web-based environment that creates simple visualisations, calculates statistics and produces concordances. The website was created to allow for easy visualisations of complex corpus data and easy testing of a number of different sociolinguistic hypotheses. The workshop will be structured around a series of practical exercises guiding the participants through different types of analysis of corpus data and statistical procedures. The following areas will be covered:

1. Sociolinguistic data in language corpora
2. Descriptive and inferential statistics
3. Individual and social variation
4. The null-hypothesis testing paradigm and the “new” statistics
5. Statistics covered: Log-likelihood, Mann-Whitney U test, Spearman’s rank correlation, Confidence Intervals, Robust mean difference

The workshop does not require any prior knowledge of statistics. It will be of interest to anyone who wants to explore sociolinguistic data using language corpora.



## Pre-conference workshops

### The Corpus Stylistics Workshop

*Michaela Mahlberg, Peter Stockwell, Rein Sikveland*  
*University of Nottingham, United Kingdom*

This workshop aims to contribute to the growing area of research that employs corpus linguistic methods in the study of literary texts. The papers in the workshop address research questions in 'corpus stylistics' from a variety of angles: with a focus on literary effects, differences between genres, literary translation, cognitive approaches and the reading experience. The workshop will also present corpus software that is specifically designed to support the analysis of literary texts.

Title	Speakers
Serious vs. popular fiction: Contextualising commonalities and differences	Rocio Montoro
Verbs of Remembering and Forgetting in Dan Brown's Angels and Demons: A Corpus-Informed Stylistic Analysis	Ernestine Lahey
The application of principal component analysis to literary translations. Comparing the Italian translations of Joseph Conrad's Heart of Darkness	Lorenzo Mastropiero
Keyword and cluster analysis in Virginia Woolf's Mrs Dalloway and literary interpretation of the novel	Anna Cermakova
The corpus stylistic analysis of prose fiction	Dan McIntyre & Brian Walker
Psycholinguistic methods to study the reading of Dickens's characters	Kathy Conklin
CLiC Dickens: Towards a cognitive corpus stylistics of characterisation	Michaela Mahlberg, Peter Stockwell & Rein Sikveland

## Pre-conference workshops

### Health communication and corpus-linguistics: Examining health discourse through the use of specialist corpora

*Kevin Harvey, Gavin Brookes*  
*University of Nottingham, United Kingdom*

Electronic corpora are being harnessed by scholars of health communication now more than they ever have been. Corpus-assisted research into health communication has developed significantly over the last 15 years and continues to develop apace. This workshop will add to this burgeoning field of corpus-based health communication research through a series of talks which demonstrate how the analysis of corpora containing health-related communication from a wide variety of contexts (such as spoken interviews, email and electronic support group fora) can provide exciting and unique insights into the way people communicate about their subjective health experiences. A key theme addressed in this workshop is the power of the corpus to challenge researcher intuition and provide novel insights into people's subjective experiences of health-related issues; findings which can, and have in the past, help to inform clinical decision-making and treatment-planning in real-world therapeutic contexts.

Title	Speakers
A Journey into the unknown: Early insights into what corpus linguistics and the internet can tell us about Diabulimia	Gavin Brookes & Kevin Harvey
Evidence-based health communication?: Examples from a corpus of depression support group messages	Daniel Hunt
Using corpus annotation tools to explore the effects of a healthcare intervention for families of deaf children	Luke Collins
Linguistic analysis of the preschool five minute speech sample: What the parents of preschool children with early signs of ADHD say and how they say it?	Elvira Perez, Melody Turner, Anthony Fisher & David Daley
A corpus linguistics approach to the language of online self-help in mental health	Paul Bonham

# Full papers

### ***There's not a penny in your pocket, but we believe every single word you say: The extremes of hyperbolic synecdoche in the domains of money and language***

Annile Ädel  
Dalarna University, Sweden

There is often a discrepancy in the way in which the world is normally perceived to be—insofar as we can even agree on this—and the way in which we talk about the world. Such discrepancy has long been at centre stage in rhetoric and semantics, for example in the study of figurative language, exploring phenomena such as metaphor, irony and metonymy. Figurative language can be generally defined as language use “where speakers mean something other than what they literally say” (Gibbs & Colston 2012:1). The present study deals with a specific subset of figurative language; in fact, it investigates a combination of two basic types: hyperbole and synecdoche. What is labelled ‘hyperbolic synecdoche’, then, refers to cases in which “a lexeme which typically refers to part of an entity (a) is used to stand for the whole entity and (b) is described with reference to the end point of a scale” (Author forthcoming). The ‘part for an entity’ unit is placed at either end of the scale of NOTHING (0%) or ALL (100%), as exemplified in the title’s *not a penny* and *every single word*. Cases of hyperbolic synecdoche express affective meanings, which are often highly evaluative. In the study, hyperbolic synecdoche is investigated in two specific domains—those of verbal communication and money—through the lexical items *word* and *penny*. Empirical data was collected from the British National Corpus (BNC), resulting in over 500 examples of hyperbolic synecdoche involving *word* and *penny*. The category itself became salient through serendipity—characteristic of corpus-based work (cf. Fillmore 1992:35)—in the context of a larger study of metonymy and metaphor in WORD, in which hyperbolic synecdoche turned out to be a persistent pattern. In a large sample (total N=1,874) from the BNC, the noun *word* was found generally to be involved in metonymy to a great extent, and it was found to occur in hyperbolic synecdoche, specifically, in 20% of cases (Author forthcoming). The equivalent number for *penny* (total N=1,124) is even greater, with as much as 68% of the figurative examples counting as hyperbolic synecdoche. Metonymy is traditionally described in terms of single-word units, but the current material consists of a large proportion of multi-word units. Specifically, hyperbolic synecdoche was found to appear in a restricted range of lexicogrammatical patterns, with a high degree of conventionalisation. Another interesting pattern found concerned the end points of the scale, in that the negated NOTHING patterns were more common than the ALL patterns, accounting for some 80% of the cases for *word* and 60% for *penny*. This suggests that the discrepancies mentioned above are consistently patterned, such that for example negative evaluation—or the NOTHING end of the scale—predominates over positive evaluation—or the ALL end of the scale—in hyperbole (cf. Cano Mora 2009).

#### References

- Author. Forthcoming. ‘Not one word of it made any sense’: Hyperbolic synecdoche in the British National Corpus.
- Cano Mora, Laura. 2009. All or Nothing: A Semantic Analysis of Hyperbole. *Revista de Lingüística y Lenguas Aplicadas* (4). 25-35.
- Fillmore, Charles J. 1992. ‘Corpus linguistics or computer-aided armchair linguistics’, in Svartvik, Jan (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*. Berlin & New York: Mouton de Gruyter, 35-60.
- Gibbs, Raymond W. & Colston, Herbert L. 2012. *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.

### A variational pragmatic study of *actually* on the basis of four ICE-corpora

Karin Aijmer

University of Gothenburg, Sweden

This study could be regarded as a contribution to the budding new discipline of variational pragmatics concerned with “pragmatic variation between or across L1 varieties of the same language” (Barron and Schneider 2009; cf. Schneider and Barron 2008). Putting regional variation on the map has led to promising work particularly on speech acts across dialects or language varieties in many different languages. Variational studies on pragmatic markers are still scanty (but compare Columbus 1999 (invariant tags; Mair 2009) on *be like* in Jamaican English and Aijmer 2013 on general extenders. The aim of the current study is to investigate the distribution and functions of *actually* across some varieties of English on the basis of some selected ICE-corpora and to discuss the reasons for the variation. *Actually* has been extensively studied in British English but not across other national varieties (see e.g. Aijmer 2002, 2013; Clift 2001; Oh 2000). The inspiration for the study came from the observation that *actually* occurs with different frequencies and in different positions in national varieties of English. Table 1 shows the distribution of *actually* in British English (ICE-GB), New Zealand English (ICE-NZ) Australian English (ICE-AUS), and Singapore English (ICE-SIN) in private and public dialogue: In New Zealand English the total number of *actually* was less than in the other varieties. The table also shows that the distinction between private and public dialogue plays a role. *Actually* is most frequent in private dialogue in all the varieties except New Zealand English. There are also differences between the varieties with regard to the position of *actually* in the utterance. Medial position was most frequent in all the varieties except for Singapore English where the number of examples in initial position was dominant. In New Zealand English medial position was frequent while initial position was only found in 7.8% of the examples. In end –position on the other hand *actually* was less frequent than in the other varieties. A number of factors can account for the differences between the varieties. The high frequency of *actually* in initial position Singapore English can for instance be discussed in terms of the constraints imposed by real-time processing. The difference between New Zealand English and Australian English raises questions about grammaticalization and the spread of *actually* across the varieties. Pragmatic markers have been a challenge to scholars because of their multifunctionality and their formal properties as inserts. By studying pragmatic markers in many regional varieties (and activities) can we get a better picture of the extent of their variation which can be an input to a better description of pragmatic markers generally.

### References

- Aijmer, K. 2002. *English discourse particles. Evidence from a corpus*. Amsterdam and Philadelphia: John Benjamins.
- Aijmer, K. 2013. Understanding pragmatic markers. A variational pragmatic approach. Edinburgh: Edinburgh University Press.
- Barron, A. and K. Schneider. 2009. Variational pragmatics: Studying the impact of social factors on language use in interaction. *Intercultural pragmatics* 6(4) : 425-442
- Clift, R. 2001. Meaning in interaction: The case of *actually*. *Language* 77(2): 245-291.
- Columbus, G. 1999. A corpus-based analysis of invariant tags in five varieties of English. In Renouf, A. and A. Kehoe (eds) *Corpus linguistics: Refinements and reassessments*. Amsterdam and New York: Rodopi. 401-414.
- Mair, C. 2009. Corpus linguistics meets sociolinguistics: the role of corpus evidence in the study of sociolinguistic variation and change. In Renouf, A. and A. Kehoe (eds) *Corpus linguistics: Refinements and reassessments*. Amsterdam and New York: Rodopi. 7-32.
- Schneider, K. P. and A. Barron (eds.) 2008. *Variational pragmatics: A focus on regional varieties in pluricentric languages*. Amsterdam & Philadelphia: John Benjamins.

### The Engineering Lecture Corpus: Visualising cross-cultural difference in discourse function

Sian Alsop<sup>i</sup>, Hilary Nesí,<sup>i</sup> Joel Priestley<sup>ii</sup>  
Coventry University, United Kingdom<sup>i</sup> / University of Oslo, Norway<sup>ii</sup>

Although powerful, corpus query tools are not the most helpful for certain kinds of investigation into discourse features. The questions that drive the creation of specialised corpora often necessitate the use of specific systems of annotation, which can benefit from specific visualisation techniques. This is certainly the case with the Engineering Lecture Corpus (ELC), a small specialised corpus of lectures from the UK, Malaysia, and New Zealand. The ELC is designed to investigate the discourse of lectures, and to address whether discourse differences exist in lectures delivered in different parts of the world, but in the same language medium (English) from the same discipline (Engineering) at the same level of study (undergraduate). The ELC lectures are annotated for pragmatic features (c.f. Simpson-Vlach and Leicher, 2006), through inline XML annotation. Chunks of text within the transcriptions have been identified as elements (or discourse functions): explaining, housekeeping, humour, storytelling, and summarising. Subcategories are attributed to some; summarising, for example, identifies two types of preview and two types of review. One useful way of identifying and analysing patterns within ELC data is through a dashboard view across categories and cultural sub-corpora. We know that "[t]he power of the unaided mind is highly overrated" (Norman 1993: 43) and that computational corpus techniques can reveal certain patterns unseen by the naked eye (Flowerdew 2013: 161). Yet that the culture of visualisation has not been widely explored by corpus linguists (Rayson et al. 2009). We wanted to see the ELC annotation in a way that addresses the central research question of cultural similarity/difference. Building on an existing D3.js timeline library and the principle of small multiples, we therefore visualised dispersion and duration: where in the lecture each pragmatic function occurs and how many tokens it contains. The javascript uses normalised start and end indices to plot each category along the x axis, which can be displayed according to variables of element or subcorpus. This allows us to build a picture of what is happening and where in selected lectures, viewing the corpus in terms of any combination of pragmatic category or cultural component. Each lecture is presented in vertical alignment to facilitate comparison. This overview will also be interactive, allowing full text to be accessed via visual representation. Early findings suggest that significant differences occur in lectures cross-culturally. The narrative storytelling and bawdy and ironic humour types, for example, occur significantly more frequently in the UK lectures compared to those from Malaysia, whereas self-deprecating humour is most heavily used in the New Zealand component. Summaries of previous lectures cluster strongly towards the beginning of the speech events in general. Patterns of co-occurring summary types are also evident: previews of current lecture content are commonly immediately followed by previews of future content in the subcorpora from both New Zealand and the UK. By visualising our annotation system in this way, we notice that indications of cross-cultural difference and some phasal structuring (cf. Young 1994) are beginning to emerge.

#### References

- Flowerdew, J. (2013) *Discourse in English Language Education*. Routledge: New York
- Norman, D. A. (1993). *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Reading, MA: Addison Wesley
- Rayson, P. and Mariani, J. (2009) "Visualising corpus linguistics". In (Eds.) Mahlberg, M., González-Díaz, V. and Smith, C. *Proceedings of the Corpus Linguistics Conference (CL2009)*, University of Liverpool UK. <<http://ucrel.lancs.ac.uk/publications/cl2009>>
- Simpson-Vlach, R. and Leicher, S. (2006) *The MICASE Handbook: A Resource for Users of the Michigan Corpus of Academic Spoken English*. Ann Arbor: University of Michigan Press
- Young, L. (1994) "University lectures - macro-structure and micro-features". *Academic Listening*. (ed.) Flowerdew, J. Cambridge: Cambridge University Press, 159-176

### The double copula revisited

Gisle Andersen  
NHH, Norway

An intriguing new phenomenon in the grammar of English is the emergence of what has been called the ‘double’ or ‘reduplicative copula’ (Tuggy 1996; Shapiro 1993; Shapiro & Haley 2002; Author 2002; Curzan 2012), that is, sentences of the type *The truth is is that people are in the middle of all these things* (COCA SPOK 1995 ABC\_Brinkley) and *The reason was is that Janes (sic.) aunt had passed away* (COCA SPOK 2001 CBS\_48Hours). This study uses a range of different corpora to investigate this ‘bizarre syntactic phenomenon’ (Shapiro 1993: 9) and focuses particularly on how the double copula construction is constrained by social and linguistic factors. In an earlier paper (Auhor 2002), I have described the double copula as a discourse-organising construction that may have developed by the conflation of two focus structures, the *wh*-cleft and postponement of a clausal subject. It contributes to the information structure of utterances and is used by speakers to assign focus and weight to the following complement clause, in much the same way as cleft sentences and pseudo-clefts (Anderson 2002; Curzan 2012). The recent compilation of very large corpora such as COCA (Davies 2009) enables a more systematic account than the anecdotal evidence reported by most of the early studies (as seen from Curzan’s COCA-based study). This paper uses the methodology of short-term diachronic corpus analysis (Leech et al. 2009) in an analysis of variation across time and region. The corpus method involves the comparison of British and American corpora recorded at different periods, notably the LLC (BrE 1953-1987), the spoken component of the BNC (BrE 1991-1994), the spoken component of COCA (AmE 1992-2012), the Bergen Corpus of London Teenage Language (COLT, BrE adolescents 1993) and finally a more recent teenage corpus recorded in two sociolinguistic projects (LIC/MLE, BrE adolescents 2004-2006). Given the recency of this feature and the widely held view that AmE is the more progressive and BrE the more conservative of the two varieties (Leech et al. 2009), we can hypothesise that the double copula will be more frequent in the more recently recorded corpora and more frequent in AmE than in BrE. The study shows that the double copula construction occurs repetitively in both British and American speech in corpus data from the early 1990s onwards. Despite a few decades of existence, it is still largely confined to spoken language, but it occurs marginally in written representations of dialogue in news reportage in both BrE and AmE. The double copula construction is an atypical innovation in that it does not emerge ‘from below’ but is chiefly associated with educated speakers (Tuggy 1996). Further, there are indeed signs that the innovation first occurred in American, and, like other more well-established focus constructions it seem to thrive particularly in argumentative genres.

### References

- Anderson, Gisle. 2002. The best part is, is that you get to shoot your opponent: Corpora and the double copula. From the COLT’s mouth ... and others: Language corpora studies in honour of Anna-Brita Stenström, ed. by L.E. Breivik & A. Hasselgren. Amsterdam: Rodopi.
- Curzan, Ann. 2012. Revisiting the reduplicative copula with corpus-based evidence. *The Oxford handbook of the history of English*, ed. by T. Nevalainen & E.C. Traugott, 211-21. Oxford: Oxford University Press.
- Davies, Mark. 2009. The 385+ million word Corpus of Contemporary American English (1990-2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics* 14.159-90.
- Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith (eds) 2009. *Change in contemporary English*. Cambridge: Cambridge University Press.
- Shapiro, Michael & Michael C. Haley. 2002. The reduplicative copula *is is*. *American Speech* 77.305-12.
- Shapiro, Michael. 1993. Presidential address: The boundary question. *The American Journal of Semiotics* 10.5-26.
- Tuggy, David. 1996. The thing is is that people talk that way. The question is is why? *Cognitive linguistics in the redwoods: The expansion of a new paradigm in linguistics*, ed. by E. Casad, 713-52. Berlin: Mouton de Gruyter.



### “I differ with you, Madam” – verbal conflict in historical drama

Linnéa Anglemark  
Uppsala University, Sweden

The genre of drama texts is speech-purposed and speech-like in its nature (Culpeper and Kytö 2010:18), and includes many dialogic features. One type of interaction that is relatively common in drama texts is the verbal conflict. Its importance can be attributed to the nature of the genre; verbal conflicts create dramatic tension on the stage and can also be a means of increasing audience involvement (Herman 1995:137). This paper describes a comparative study on verbal conflict in historical drama in two languages. The corpus used for the investigation is the English and Swedish Drama Dialogue corpus, compiled and annotated by researchers at the English Department and the Department of Scandinavian Languages at Uppsala University. The corpus comprises excerpts from a total of 30 drama texts from the 18th, 19th and 20th centuries, in English and in Swedish. Every utterance in the texts is tagged for the speaker and recipient(s) of the utterance, and their respective age, gender and social status. The corpus can thus be searched to explore discourse patterns between interlocutors based on these variables. The plays include characters from different social strata. Previous studies on historical drama texts (e.g. Sörlin 2008) have indicated that the frequency of verbal conflicts is relatively similar in drama texts from different periods of time. The sociopragmatic tagging of the corpus makes it possible to investigate whether differences in gender, age, and social status make a difference in the frequency, duration, and type of conflict. The primary means of locating verbal conflicts in the material was through searches for conflict labels, lexical items such as fight and quarrel, which are associated with conflicts. This was supplemented with lexical searches for phrases such as with you which can be found in a variety of contexts, including verbal conflicts. Similar phrases were used for lexical searches in the Swedish-language text. The search hits were screened to exclude irrelevant hits, and to identify the beginning and the end of the verbal conflict. The investigation indicates developments over time, as well as differences between the two languages. Differences in social status are considerably more marked in the earliest dramas, in that social inferiors are more unlikely to instigate any form of verbal conflict with their social superiors in the 18th century dramas than in the 20th century texts; however. The difference over time is less pronounced in the English-language text than the Swedish-language texts.

### References

- Culpeper, Jonathan, and Merja Kytö. 2010. *Early Modern English Dialogues: Spoken Interaction as Writing*. Cambridge: Cambridge University Press.
- English and Swedish Drama Dialogue. Forthcoming. Corpus compiled by Linnéa Anglemark, Merja Kytö, Ulla Melander Marttala, and Mats Thelander (Uppsala University), and Sofia Gustafson Capková (Stockholm University).
- Herman, Vimala. 1995. *Dramatic Discourse: Dialogue as Interaction in Plays*. London: Routledge.
- Sörlin, Marie. 2008. Att ställa till en scen. Verbala konflikter i svensk dramadiolog 1725–2000. *Skrifter utgivna av Institutionen för nordiska språk* 73.

Normalising the *Corpus of English Dialogues (1560-1760)* using VARD2: Decisions and JustificationsDawn Archer<sup>i</sup>, Merja Kytö<sup>ii</sup>, Alistair Baron<sup>iii</sup>, Paul Rayson<sup>iii</sup>University of Central Lancashire<sup>i</sup>, United Kingdom / Uppsala University<sup>ii</sup>, Sweden / Lancaster University<sup>iii</sup>, United Kingdom

The development of (semi-)automatic tools such as the VARD (Baron and Rayson, 2008) has afforded compilers of historical corpora the opportunity to normalise variant spellings relatively quickly – following, that is, a dedicated period of manual training using relevant corpus samples (see, e.g., Lehto et al. 2010). In the case of VARD2, this period of manual training involves the user:

- (i) reading a given text, via the VARD interface,
- (ii) distinguishing variants within the text – via the tool's recommended list of (ranked) candidate replacements – or personally – by highlighting variant forms manually,
- (iii) choosing the most appropriate normalized form for each variant found – where relevant, being guided by the VARD's known variant list or f-score calculation (derived from , e.g., letter replacement rules, edit distance measures and/or phonetic matching algorithms),
- (iv) replacing the variant with the normalised form – but in such a way that the original spelling is retained in an XML tag (Baron and Rayson, 2008).

The corpus-linguistic argument for normalisation is that it helps improve automated techniques such as part-of-speech and keyword analysis, thereby allowing existing linguistic tools to be used unmodified (see, e.g., Archer et al. 2003; Rayson et al. 2007a/b; Rayson et al. 2009; Hiltunen and Tyrkkö 2013). But such normalisation needs to be handled sensitively: so that, for example, we can maintain - within the text - the original spelling of those forms which convey important morphosyntactic or orthographic information (as opposed to retaining these original spellings as part of the XML tag – see (iv)). Hence the inclusion of an IGNORE VARIANT facility within VARD. In this paper, we outline some of the decisions we have made, in respect to the *Corpus of English Dialogues* (CED), when determining which features required normalisation and which should be left as they were originally (and why). Compiled by Merja Kytö and Jonathan Culpeper, the CED covers a 200-year period (1560-1760) and contains speech-related texts representative of five genres – the courtroom, witness proceedings, comedy dramas, prose fiction and handbooks – as well as a group of texts subsumed under a miscellaneous category. In particular, we will discuss our treatment of: names; the genitive construction; auxiliaries and verbs; (open-hyphenated-closed) compounds; abbreviations; graphemes such as the tilde; terms which are now archaic, obsolete or rare; foreign terms; dialect terms; and personal pronouns. This work, although focussed on the CED, also has a wider aim: determining the feasibility of developing normalisation guidelines that are generalisable to other historical corpora such as ARCHER (*A Representative Corpus of Historical English Registers*) and EEBO (*Early English Books Online*). Hence, as part of the presentation, we will compare the normalisation decisions made in respect to the CED with those made in respect to the *Early Modern English Medical Texts* (see Lehto et al. 2010).

## References

- Archer, D., McEnery, A. M., Rayson, P. and Hardie, A. (2003). Developing an automated semantic analysis system for Early Modern English. In D. Archer, P. Rayson, A. Wilson and A. M. McEnery (eds.) *Proceedings of the Corpus Linguistics Conference 2003*. Lancaster: University of Lancaster. 22–31.
- Baron, A. and Rayson, P. (2009). Automatic standardization of texts containing spelling variation, how much training data do you need? In M. Mahlberg, V. González-Díaz and C. Smith (eds.) *Proceedings of the Corpus Linguistics Conference, CL2009*, University of Liverpool, UK, 20-23 July 2009.
- Baron, A., Rayson, P. and Archer, D. (2009). Word frequency and key word statistics in historical corpus linguistics. *Anglistik: International Journal of English Studies*, 20 (1), pp. 41–67.
- Baron, A. and Rayson, P. (2008). VARD 2: A tool for dealing with spelling variation in historical corpora. *Proceedings of the Postgraduate Conference in Corpus Linguistics*, Aston University, Birmingham, UK, 22 May 2008.
- A Corpus of English Dialogues 1560-1760*. (2006). Compiled under the supervision of Merja Kytö (Uppsala University) and Jonathan Culpeper (Lancaster University).
- Hiltunen, T. and Tyrkkö, J. (2013). Tagging *Early Modern English Medical Texts*. *Corpus Analysis with Noise in the Signal (CANS) 2013*. Lancaster University.
- Lehto, A., Baron, A., Ratia, M. and Rayson, P. (2010). Improving the precision of corpus methods: The standardized version of *Early Modern English Medical Texts*. In I. Taavitsainen and P. Pahta (eds.) *Early Modern English Medical Texts: Corpus Description and Studies*. Amsterdam: John Benjamins. 279-290.
- Rayson, P., Archer, D., Baron, A. and Smith, N. (2007a). Tagging historical corpora – the problem of spelling variation. In *Proceedings of Digital Historical Corpora, Dagstuhl-Seminar 06491*, International Conference and Research Center for Computer Science, Schloss Dagstuhl, Wadern, Germany, 3rd-8th December 2006.
- Rayson, P., Archer, D., Baron, A., Culpeper, J. and Smith, N. (2007b). Tagging the Bard: Evaluating the accuracy of a modern POS tagger on Early Modern English corpora. In: *Proceedings of the Corpus Linguistics Conference 2007*. Birmingham: University of Birmingham.

## A Corpus-Based Study of Intensifiers in Late Modern English

Tobias Bernaisch

University of Geissen, Germany

Intensifiers such as *very*, *so* or *extremely* are used to scale a particular descriptive quality upwards from an assumed norm (cf. Quirk et al. 1985: 590). These lexical items "are continually in need of replacement by stronger and more expressive forms" (Brinton & Arnovick 2006: 441) and thus represent a central area of grammatical change in English. Empirical studies, however, have so far predominantly restricted themselves to apparent-time/synchronic approaches (cf. e.g. Ito & Tagliamonte 2003; Fuchs 2013) or, when diachronic data was used, to present-day English (cf. e.g. Tagliamonte & Roberts 2005; Barnfield & Buchstaller 2010) with only few exceptions (e.g. Partington (1993) on Middle and Early Modern English or Peters (1994) on Early Modern English). In the light of this lack of empirical insights into the more recent development of intensifiers in the English language, the present paper focusses on intensifier usage in Late Modern English (LModE). This study examines to what extent the findings reported for intensifiers in present-day English also apply to earlier stages of English and, in case they do not, more closely scrutinises the patterns and agents of linguistic change involved. Further, quantitative perspectives on earlier (non-empirical) descriptions of intensifiers in the history of the English language are also provided. The following research questions are at the centre of attention in this paper:

- Which intensifiers are used most frequently in LModE and is this set of intensifiers subject to diachronic change within this period?
- Which speaker-related (e.g. social class, gender) and structural characteristics (e.g. collocations, colligations) accompany this possible diachronic change?
- Do female speakers use intensifiers more frequently than male speakers in LModE?
- Do lower-class speakers use intensifiers more frequently than upper-class speakers in LModE?

The Old Bailey Corpus, an approximately 14m-word-large and copiously annotated database covering spoken texts from 1720 to 1913 (cf. Huber et al. 2012), was used as the empirical resource, from which a total of 11,311 intensifiers was drawn. For each intensifier, structural (i.e. its adjectival complement, the corresponding adjective class (cf. Dixon 1977) and the type of predication) as well as sociolinguistic information (i.e. the year of production, speaker gender, speaker age and the speaker's social class) was documented. The results show that *very*, in line with Mustanoja's (cf. 1960: 326f.) observation, is the most frequent intensifier followed by *so* across LModE, but it is also in this period that *perfectly* and later *absolutely* become increasingly frequent as intensifiers representing instances of diachronic change. From a sociolinguistic viewpoint, it is noteworthy that, contrary to Jespersen's (cf. 1922: 250) and Ito & Tagliamonte's (cf. 2003: 277) descriptions, it is (upper-class) men who lead the changes of *perfectly* and *absolutely*, which are accompanied by structural effects such as the emergence of novel collocations (e.g. *perfectly safe*, *absolutely untrue*) and increases in adjectival complement classes. Further, the paper displays the diachronic validity of the present-day English finding that female speakers use intensifiers more frequently than male speakers (cf. Barnfield & Buchstaller 2010) and is able to show that intensifier usage - in analogy to present-day English (cf. Ito & Tagliamonte 2003: 277) - was not a linguistic marker of lower social class in LModE.

## References

- Barnfield, K. & I. Buchstaller (2010): "Intensifiers on Tyneside: longitudinal developments and new trends", *English World-Wide* 31(3), 252-287.
- Brinton, L.J. & L.K. Arnovick (2006): *The English Language: A Linguistic History*. Oxford: Oxford University Press.
- Dixon, R.M.W. (1977): "Where have all the adjectives gone?", *Studies in Language* 1, 19-80.
- Fuchs, R. (2013): "*So ladylike* and *particularly manly*? Determinants of variation in intensifier usage in Indian English". Paper presented at ICAME 34, Santiago de Compostela, Spain, May 2013.
- Huber, M., M. Nissel, P. Maiwald & B. Widlitzki (2012): *The Old Bailey Corpus. Spoken English in the 18th and 19th Centuries*. <www.uni-giessen.de/oldbaileycorpus>, 20 Nov 2013.
- Ito, R. & S. Tagliamonte (2003): "*Well* weird, *right* dodgy, *very* strange, *really* cool: layering and recycling in English intensifiers", *Language in Society* 32, 257-279.
- Jespersen, O. (1922): *Language: Its Nature, Development, and Origin*. London: Allen & Unwin.
- Mustanoja, T.F. (1960): *A Middle English Syntax*. Helsinki: Société Néophilologique.
- Partington, A. (1993): "Corpus evidence of language change: the case of intensifiers", *Text and Technology: In Honour of John Sinclair*, eds. M. Baker, F. Gill, E. Tognini-Bonelli & J. Sinclair. Amsterdam, Philadelphia: John Benjamins.
- Peters, H. (1994): "Degree adverbs in Early Modern English", *Studies in Early Modern English*, ed. D. Kastovsky. Berlin: Mouton de Gruyter. 269-288.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985): *A Comprehensive Grammar of the English Language*. London: Longman.
- Tagliamonte, S. & C. Roberts (2005): "So weird; so cool; so innovative: the use of intensifiers in the television series *Friends*", *American Speech* 80(3), 280-300.

**No fear(s) + complements clauses: a source of modal and emphatic polar markers**

Liselotte Brems

This paper investigates the path of semantic change and grammaticalization that led from complementation structures with lexical uses of *fear(s)* to grammatical markers containing 'no' *fear(s)* that convey (epistemic) modal and polar meanings. Based on extensive data from the PPCME, PPCEME, CLMETEV and *WordBanks Online corpus* (WBO), I will track the micro-developments of this pattern from 1150-present. Synchronically, *there is no fear*, *have no fear(s)* and *n/No fear(s)* can act as epistemic modal markers (2) and emphatic polar expressions (3) with scope over the entire utterance, expressing strong certainty and/or emphatic negation. In addition, lexical examples can still be found in which a CC merely fills in the contents of the fear (1):

- (1) "My voice trailed away as I succumbed to the **fear** that I was failing to convey my sincerity." (WBO)  
 (2) I **have absolutely no fears** that France will cave in to the strikers' action (WBO) ('they won't cave in')  
 (3) When I asked him if he wanted to change his mind, William said '**No fear**'. (WBO) ('certainly not')

Over time, the structural and discourse-semantic status of the CC has been reanalysed, similarly to related patterns such as *no doubt/question/wonder/way* (cf. Simon-Vandenberg 2007, Davidse & De Wolf 2012). In examples such as (2) and (3) the CC is reanalyzed as a modalized proposition. In the analysis I will take into account the sorts of CC (*to* vs. *that* cf. 'proposition' vs. 'desired actions'-contents of the CC, Wierzbicka 1988) and changes in the modification patterns of *fear(s)* by means of determiners, adjectives. In addition, I will look at the potential influence of phraseologies, e.g. *have/there is + fear(s)* and (negative) polarity sensitivity. Finally I will study specialization in terms of 'discourse contexts', e.g. types of subjects (impersonal *there*, first persons) and verb phrases in the matrix clause and in CCs. By means of a methodology that combines insights from construction grammar, grammaticalization research and diachronic corpus analysis this presentation makes an explicit stand for combined methodologies in linguistic analysis.

**References**

- Davidse, K., De Wolf, S. 2012. Lexicalization and grammaticalization: modal modifier constructions with "no question". *Text and Talk*, 32(5), 569-591.  
 Simon-Vandenberg, A. 2007. *No doubt* and related expressions: a functional account. In Hannay, Mike and Gerard J. Steen (eds.), *Structural-Functional Studies in English Grammar: In honour of Lachlan Mackenzie*. 9–34.  
 Wierzbicka, A. 1988. *The Semantics of Grammar*. Amsterdam: John Benjamins.

## Effect sizes in Corpus Linguistics: Keywords, collocations and diachronic comparisons

Vaclav Brezina

University of Lancaster, United Kingdom

Recently, there has been a considerable shift in the statistical paradigm in both natural and social sciences (Cumming, 2012; Stang et al., 2010; Anderson et al., 2000). This shift consists in a move away from the dichotomy of the null-hypothesis-significance testing (significant vs. non-significant) towards a more complex description of the magnitude of the effect observed (effect size). Cohen's  $d$  is one of the commonly used measures of effect size (Algina et al., 2005; Cohen, 1988) and can be compared across different studies. It is a measure of the difference between two means (frequencies in two corpora) expressed in standard deviation units (a measure of dispersion). Cohen's  $d$  when applied to corpus comparison has a number potentially desirable features: 1) It is a simple statistic that takes into consideration both the frequency of occurrence of a linguistic variable as well as the dispersion (cf. Gries, 2008). It can thus help evaluate with more precision the typicality of linguistic patterns. 2) Cohen's  $d$ , more precisely its robust variant, is suitable for analysis of linguistic data where the assumption of the normal distribution is often violated. 3) When using Cohen's  $d$ , no simple YES/NO, "all-or-nothing" decision is provided by the statistical test. The analyst, therefore, needs to take an extra interpretative step in order to evaluate the meaningfulness of the difference. In this study, the *BNC* (100M words), the Brown family corpora (6 corpora, 1M words each) and a small corpus of newspaper readers' comments (1.4M) were used to evaluate the application of Cohen's  $d$  in three analytical procedures in corpus linguistics: 1) keywords, 2) collocations and 3) diachronic corpus comparisons. The corpora of different sizes and genres, compiled at different points in time, were chosen in order to investigate the possibility of different types of comparison with Cohen's  $d$ . The results obtained for Cohen's  $d$  were compared with the results of analyses that use traditional statistics (log likelihood, MI, MI<sup>3</sup>) as well as newly proposed measures (Kilgarriff, 2009). This evaluative study addresses two research questions: How reliable is the use of Cohen's  $d$  when compared with other statistics? Does the size of the effect expressed by Cohen's  $d$  correlate with the linguistic/social importance of the pattern identified? The preliminary results show that Cohen's  $d$  has a large potential in corpus-based investigations. It can be seen as a statistic, which expresses the size of a difference between two corpora in meaningful terms/units and which allows an easy combining of results from different studies (meta-analysis).

## References

- Algina, J., Keselman, H., & Penfield, R. D. (2005). An Alternative to Cohen's Standardized Mean Difference Effect Size: A Robust Parameter and Confidence Interval in the Two Independent Groups Case. *Psychological methods*, 10(3), 317.
- Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null Hypothesis Testing: Problems, Prevalence, and an Alternative. *The Journal of Wildlife Management*, 64(4), 912-923. doi: 10.2307/3803199
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.
- Gries, S. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403-437. doi: 10.1075/ijcl.13.4.02gri
- Kilgarriff, A. (2009). *Simple maths for keywords*. Paper presented at the Corpus Linguistics Conference, Lancaster.
- Stang, A., Poole, C., & Kuss, O. (2010). The ongoing tyranny of statistical significance testing in biomedical research. *European Journal of Epidemiology*, 25(4), 225-230. doi: 10.1007/s10654-010-9440-x

## Toward a methodology of chunking: applications and extensions of Linear Unit Grammar

Ray Carey

University of Helsinki, Finland

The challenges of grammatical parsing of speech are well-known, with dynamic models of grammar seeking to describe the online production of speech (e.g. Brazil 1995, O'Grady 2005, Sinclair & Mauranen 2006). More recently, momentum has grown in corpus-based approaches to describing speech, with increasing interest in describing dysfluency features in spoken corpora (Gilquin & DeCock 2011, Götz 2013). This study advances and unites these interests by compiling a corpus of parsed, transcribed speech for examining features of (dys)fluency. This methodological paper engages the question of how the descriptive model of Linear Unit Grammar, or LUG (Sinclair & Mauranen 2006), can be systematized for compiling a fully LUG-annotated corpus of transcribed academic talk. Previous LUG-based studies have focused on a narrow range of organizing chunks (Mauranen 2009, NAME 2013), and fully LUG-annotated corpora are only now beginning to be compiled (Smart 2013). The present study extends LUG further into transcriptions from the ELFA corpus (2008), consisting of spoken English used as a lingua franca (ELF) in academic settings. As a descriptive model that draws on the analyst's intuitions of chunking a linear stream of text, this study details the operationalization of LUG toward systematizing and partially automating these analytical judgments. Rather than being a liability, it is argued that applying LUG to transcriptions of spoken interaction is an advantage toward consistent analysis. With its robust ability to handle the ordinary dysfluencies of speech as automatic chunk boundaries, a great deal of LUG analysis can be partially automated. This paper further details the principles behind a partially automated LUG chunking algorithm developed to aid in corpus compilation. Following the proposal of Sinclair & Mauranen (2006: 158), a data-derived list of Organizing (O) chunks are first identified in a transcript and various dysfluencies (e.g. false starts, repetitions, and filled pauses) are located and automatically segmented. Then, using a similarly data-driven chunk parser, Provisional Unit Boundaries (PUBs) are inserted. A fully LUG-annotated transcript is output in XML format, which is then manually proofread and revised in an XML editor by a human analyst. Finally, the revised transcript is automatically parsed to determine the boundaries of the Linear Unit of Meaning (LUM), the clause-like sequences forming the only hierarchical level in LUG analysis. The paper concludes by anticipating the type of research invited by a corpus of LUG-annotated spoken texts. Drawing on the first 30,000 words of LUG-parsed ELFA corpus transcripts, concordance lines of emerging chunk patterns are discussed as a means of describing (dys)fluency features in spoken interaction. At the same time, a methodological shift is demonstrated in which traditional searches for co-occurring words is extended to co-occurring chunks of language.

## References

- Brazil, David (1995) *A Grammar of Speech*. Describing English Language. Oxford: Oxford University Press.
- NAME (2013) On the other side: formulaic organizing chunks in spoken and written academic ELF. *Journal of English as a Lingua Franca*, 2(2), 207-228.
- ELFA (2008) The Corpus of English as a Lingua Franca in Academic Settings. Director: Anna Mauranen. <http://www.helsinki.fi/elfa/elfacorporus>. (9.12.2013).
- Gilquin, Gaëtanelle & De Cock, Sylvie (2011) Errors and disfluencies in spoken corpora: Setting the scene. *International Journal of Corpus Linguistics*, 16(2), 141-172.
- Götz, Sandra (2013) *Fluency in native and non-native English speech*. Studies in Corpus Linguistics, vol. 53. Amsterdam: John Benjamins.
- Mauranen, Anna (2009) Chunking in ELF: Expressions for managing interaction. *Intercultural Pragmatics*, 6(2), 217-233.
- O'Grady, William (2005) *Syntactic carpentry: an emergentist approach to syntax*. London: Lawrence Erlbaum Associates.
- Sinclair, John McH. & Anna Mauranen (2006) *Linear Unit Grammar: Integrating Speech and Writing*. Studies in Corpus Linguistics, vol. 25. Amsterdam: John Benjamins.
- Smart, Cameron (2013) *The role of discourse reflexivity in a linear description of grammar and discourse: the case of IMDb message boards*. Unpublished doctoral dissertation, University of Birmingham.

Distinguishing and subclassifying *there*-clefts

Kristin Davidse, Ditte Kimps

University of Leuven, Belgium / University of Leuven, Belgium

*There* clefts have received little attention in the English grammar tradition and their very existence is debated. Huddleston (1984:469) characterized example (1) as "a kind of cleft construction... descriptive of an event that took place", and Huddleston & Pullum (2002:1396) hold that the case is weak for structurally analysing (2) as specifying a value for a variable designated by the relative clause, i.e. as in some way 'specificational' like an *it*-cleft.

- (1) There was one man (that) kept interrupting.
- (2) Who might be able to help? Well, there's John you could try.

Against this, I will present grammatical and semantic-pragmatic arguments based on qualitative and quantitative corpus-analysis for recognizing **specificational** existential *there*-clefts, which contrast with **presentational** *there*-clefts. I distinguish two types of **specificational** *there*-clefts, whose 'binary' value-variable structure is prosodically reflected by the value carrying information focus, *lone* (3), *Hlocking*, *Hlerman* (4), and the variable featuring a post-nuclear rise (3) or occupying a separate information unit (4). **Cardinal** specificational *there*-clefts, e.g. (3), indicate the absolute quantity, e.g. *one* (3), *few*, *not many*, etc. of instances corresponding to the variable, e.g. *who didn't lecture*, which restricts determiners in the focal constituent to absolute quantification (Milsark 1977). **Enumerative** specificational *there*-clefts (4) enumerate one or more instances, e.g. *Hocking*, *Herman*, as corresponding to the variable *who is ... known*, which, I hypothesize, requires the 'enumeration' to be started off by definite NPs, even though indefinite ones may appear later in the list, e.g. *a few others* (5).

- (3) // think the majority of them d/o // per^haps there was lone who didn't lecture // (LLC)
- (4) // well ^first of 'all // there's a ^man called " ! Hlocking // who ^has I ^think taken his degree // ^in this department // and is ^known // who ^seemed // to ^be !fairly 'strong // ^and there is " ! Hlerman // who is ^also 'known // (LLC)
- (5) They can lie, so can I. It's the internet. There's only me and a few others who could tell you if I was telling the truth or not. (Google)

**Presentational** *there*-clefts are not specificational. In at least a number of them, *there's* serves to introduce a whole **proposition**, e.g. (6), as reflected by the single information unit and sentence focus (Lambrecht 1988). In others, however, it's the **entity** referred to by the postverbal NP, *the two men* (7), that seems to be presented and foregrounded (Lambrecht 1988).

- (6) A: which ^on the 'whole are a good 'thing#  
C: I'm ^sure there's a 'lot could be 'done# (LLC)
- (7) There's the two men downstairs called to see you again. (LDC)

The characteristics of these two basic types, and further subtypes, of *there*-cleft will be fleshed out in terms of the observed determiner structure of their post-verbal NPs, the allowed relative markers (*wh*-pronoun, *that* or zero), and their information structure, as deducible from the larger context (Kaltenböck 2005). This systematic descriptive study will be based on all relevant instances of syntagms containing existential *there* followed by lemma 'be', up to 3 intervening words, noun, up to two intervening words and finite VP from the 'brspok' (1995-1999) subcorpus of Wordbanks and from the transcribed parts of COLT and LLC.

## References

- Huddleston, R. 1984. *Introduction to the Grammar of English*. Cambridge: CUP.
- Huddleston, R. & G. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: CUP.
- Kaltenböck, G. 2005. It-extraposition in English: a functional view. *International Journal of Corpus Linguistics* 10(2): 119-159.
- Lambrecht, K. 1988. Presentational cleft construction in spoken French. In Haiman, J. & S. A. Thompson (eds), *Clause combining in grammar and discourse*. Amsterdam : John Benjamins. 135-179.
- Milsark, G. 1977. Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis* 3: 1-29.



**'What else did we do?': Interrogatives in learner interviewee speech**

Sylvie De Cock

Universite catholique de Louvain, Belgium

The Louvain International Database of Spoken English Interlanguage (LINDSEI) contains informal interviews with intermediate to advanced level learners of English as a foreign language. However informal these interviews may be, they do not share two of Clark's (1996) typical features of face-to-face conversation, namely self-determination and self-expression. While the free exchange of turns is a fundamental organising factor of conversations, in interviews the participants do not determine for themselves what actions to take when. Instead of being 'locally managed' as in conversations (Lazaraton 1992), the turn-taking system is pre-specified: interviews are organised according to a question-answer format. Besides taking actions as themselves (Clarke's self-expression) the participants in an interview also take actions as 'interviewer' or 'interviewee'. As Fiksdal (1990) points out, the participants have rights and obligations as interviewer or interviewee: the interviewer has the right and obligation to ask questions and the interviewee has the obligation to answer these questions. This paper sets out to explore the extent to which interrogative clauses and more specifically *Wh*-questions and *yes/no*-questions (Biber et al 1999) are used by the learners in the various subcorpora included on the LINDSEI CD-ROM (Gilquin et al. 2010) in spite of the fixed turn-taking format. The Concord tool in WordSmith Tools 4.0 (Scott 2008) is used to retrieve the instances of *wh*-words and primary and secondary auxiliaries from the interviewee turns of all the subcorpora on the LINDSEI CD-ROM (spoken productions by learners from eleven mother tongue backgrounds: Bulgarian, Chinese, Dutch, French, German, Greek, Italian, Japanese, Polish, Spanish and Swedish). The automatic retrieval is followed by careful analysis of the concordances to identify the *Wh*-questions and *yes/no*-questions in the data. The various corpora used contain between 60,000 and 100,000 words of interviewee speech and the interviews all follow the same set pattern. The use of *Wh*-questions and *yes/no*-questions is also investigated in a comparable corpus of informal interviews with native speakers of English, i.e. the Louvain Corpus of Native English Conversation (LOCNEC). The focus of the study is more specifically on the discourse/pragmatic functions of questions uncovered in the data: e.g. genuine questions to the interviewers, direct reported questions, rhetorical questions which enable speakers to carry on talking while thinking about what to say next (*what else did we do*) and/or to establish 'rapport' with the interlocutor (*do you know what I mean*), direct appeals for assistance (*how do you say that*; cf. communication strategies, Tarone et al. 1983). The interviewer variables recorded in the LINDSEI database (especially status, mother tongue and knowledge of other foreign languages) are also taken into consideration in the analysis of the discourse/pragmatic functions of the questions under study as they may be seen to affect the frequency of the type of questions used.

**References**

- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999), *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Clark, H. H. (1996) *Using Language*. Cambridge: Cambridge University Press.
- Fiksdal, S. (1990) *The Right Time and Pace: A Microanalysis of Cross-cultural Gatekeeping Interviews*. New Jersey: Ablex Norwood.
- Gilquin, G., De Cock, S. & Granger, S. (eds) (2010), *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Lazaraton, A. (1992) The Structural Organization of a language Interview: A Conversation Analytic Perspective. *System* 20/3, 373-386.
- Scott, M. (2008), *WordSmith Tools 5*. Oxford: Oxford University Press.
- Tarone, E., Cohen, A., Dumas, G. (1983) A Closer Look at Some Interlanguage Terminology: A Framework for Communication Strategies. In Faerch, C., Kasper, G. (eds) *Strategies in Interlanguage Communication*. London: Longman, 4-14.

## Variation in native and learner informal interviews

Sylvie De Cock, Pascual Perez-Paredes

Universite catholique de Louvain, Belgium / Universidad de Murcia, Spain

This paper takes as its starting point a multidimensional analysis of informal interviews with native speakers of English and upper-intermediate to advanced EFL learners (Pérez-Paredes and Sánchez-Tornel, 2014b). The native data were taken from The Louvain Corpus of Native English Conversations (LOCNEC) and the British component of the Contrastive Analysis of Orality in Spoken English Corpus 2 (CAOS-E Corpus 2, Aguado et al. 2012); the learner data were taken from the Louvain International Database of Spoken English Interlanguage (LINDSEI, Gilquin et al. 2010). The informal interviews used in the analysis all follow the same set pattern and are made up of three main tasks: a personal narrative based on a set topic (an experience that taught them a lesson, a country that impressed them, or a film or play they liked/disliked), a free discussion mainly about university life, hobbies, foreign travel or plans for the future and a picture description. In Pérez-Paredes and Sánchez-Tornel (2014a; 2014b) both the native and the learner data were tagged for part of speech at the University of Northern Arizona and scores were subsequently computed for each interview and for each task for the following five dimensions of linguistic variation identified by Biber (1988): Dimension 1: involved versus information production, Dimension 2: narrative versus non-narrative concerns, Dimension 3: explicit versus situation-dependent reference, Dimension 4: overt expression of persuasion, Dimension 5: abstract versus non-abstract information. The multidimensional analysis revealed that, overall, the learners tend to score higher on the dimensions under scrutiny than the native speakers performing the same tasks. The analysis also shed light on variations in dimension scores between the various learner populations investigated and between the three interview parts. This research methodology has been used by Hardy and Römer (2013) to show variation in successful student writing in different disciplines represented in the MICUSP corpus. This paper is intended as a follow-up to Pérez-Paredes and Sánchez-Tornel (2014a) as it seeks to explore why the learners in the LINDSEI corpus scored higher than the native speakers in the same communicative situation on Dimension 1. In other words, the aim of the paper is to answer the following main research question: what makes the learner interviews more involved, interactional and affective than the native interviews? The study concentrates on four learner populations, namely the Chinese, French, German and Spanish components of LINDSEI. Both positive and negative features associated with Dimension 1 (Biber 1988, Conrad 2001) are scrutinized in the POS tagged learner and the native data. Positive features include for example reduced lexical content markers like demonstrative and indefinite pronouns and affective aspect of personal involvement markers like private verbs and possibility modals. Negative features include nouns, prepositions, attributive adjectives and the type/token ratio. The qualitative study also sets out to investigate differences between the learner varieties (e.g. what makes the interviews with the Spanish learners even more involved than the other learner interviews?) and between the three tasks in the interviews (e.g. is the free discussion more involved than the other parts?).

## References

- Aguado, P., Pérez-Paredes, P. & Sánchez, P. 2012. Exploring the Use of Multidimensional Analysis of Learner Language to promote Register Awareness. *System*, 40,1: 90-103.
- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Conrad, S. 2001. Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In *Multi-dimensional studies of register variation in English*. S. Conrad & D. Biber (eds.), 94-107. Harlow, England: Pearson Education.
- Gilquin, G., De Cock, S., & Granger, S. 2010. *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Hardy, J. & Römer, U. 2013. Revealing disciplinary variation in student writing: a multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 8,2: 183-2017.
- Pérez-Paredes, P. & Sánchez-Tornel, M. 2014a. A multidimensional analysis of learner language during story reconstruction in oral proficiency interviews (OPI). In Callies, M. & Götz, S. (eds) 2014. *Learner Corpora and their application in language testing and assessment*. Amsterdam: John Benjamins. Forthcoming.
- Pérez-Paredes, P. & Sánchez-Tornel, M. 2014b. Understanding the role of speaking tasks in oral proficiency interviews (OPI): a Multidimensional Analysis of native speaker language. Forthcoming.

**Can response tokens be speech acts? A corpus pragmatics investigation.**

Rachele De Felice  
UCL, United Kingdom

Sometimes, single word response tokens (i.e. not followed by a further utterance) appear to have a pragmatic function that goes beyond mere acknowledgement, and takes on the status of a full speech act, as in the following examples from the spoken component of ICE-GB:

- |                                         |         |                 |
|-----------------------------------------|---------|-----------------|
| A: Sorry could you start again          | B: Ok   | (S1A-001-003/4) |
| A: Are you going to be in there at five | B: Yeah | (S1A-040-382/3) |
| A: Can you open a little more           | B: Yes  | (S1a044 -001/2) |

Response tokens (also known as backchannels, listener responses, continuers, among many other terms) have received a vast amount of attention in areas of linguistics such as conversation analysis, sociolinguistics, and interactional linguistics with respect to their role in structuring interaction, showing affiliation or stance, displaying listener engagement, acknowledgement, maintaining the flow of conversation, and so on (cf. Yngve 1970, Sacks et al. 1974, Schegloff 1982, Drummond and Hopper 1993). With the advent of large corpora of spoken language, this has become an even richer field of inquiry, allowing researchers to examine the behaviour of these tokens on a much larger scale and identify regular functions (cf. McCarthy 2003, O’Keeffe and Adolphs 2008). For example, Gardner (2001) concludes there are four main uses of response tokens such as yeah, mm, and okay: as continuers, acknowledgements, newsmakers (marking something as noteworthy), and change-of-activity tokens (starting a new topic in the conversation). However, in examples such as the ones above, the response tokens implicitly convey the speakers’ intentions to carry out a particular action, and therefore can be considered comparable to traditional commitments, which, in Searle’s definition, “commit the speaker to a future course of action” (Searle 1976:11). The speech act function of response tokens has to date not received much attention in the literature (though see McCarthy 2003 and Aijmer and Altenberg 2012 for some comments on this issue), but, in seeking to establish how speech acts are performed in spoken language, the role of these minimal utterances cannot be ignored. This paper looks at the use of single-word response tokens in the spoken component of ICE-GB, in particular yes, yeah, and ok, to establish whether a commitment function can be clearly identified, and how the distribution of the single-word form of this speech act compares to its occurrence in longer-sentence form (e.g. I will do it). By examining how traditional categories of pragmatics can apply to the forms typical of spoken language, this research contributes to the growing field of corpus pragmatics (cf. e.g. Romero-Trillo 2013) and to our understanding of the many roles of response tokens.

**References**

- Aijmer, Karin and Altenberg, Bengt (2012). The Swedish adverb *gärna* and its English correspondences. Paper presented at *ICAME 33*, Leuven.
- Drummond, Kent, and Robert Hopper (1993). Back Channels Revisited: Acknowledgment Tokens and Speakership Incipency. *Research on Language and Social Interaction* 26, 157-177.
- Gardner, Rod (2001). *When Listeners Talk: Response Tokens and Listener Stance*. Amsterdam: John Benjamins.
- McCarthy, Michael (2003). Talking Back: “Small” Interactional Response Tokens in Everyday Conversation. *Research on Language in Social Interaction* 36 (1), 33-63.
- O’Keeffe, Anne and Adolphs, Svenja (2008). Response tokens in British and Irish discourse. In Schneider, K. and Barron, A. (eds.), *Variational Pragmatics. A Focus on Regional Varieties in Pluricentric Languages*. Amsterdam: John Benjamins, 69-98.
- Romero-Trillo, Jesus. (ed.) (2013). *The Yearbook of Corpus Linguistics and Pragmatics: new Domains and Methodologies*. Dordrecht: Springer.
- Sacks, Harvey, Schegloff, Emanuel A., Gail, Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.
- Schegloff, Emanuel (1982). Discourse as an interactional achievement: some uses of “uh huh” and other things that come between sentences. In Tannen, D. (ed.), *Analyzing Discourse Text and Talk*. Washington DC:Georgetown University Press, 71-93.
- Searle, J.R. (1976). A classification of illocutionary acts. *Language in Society* 5(1), 1–23.
- Yngve, Victor (1970). On Getting a Word in Edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, 567-557.

### NP structure and NP distribution as an indication of the development of advanced Dutch EFL writing

Pieter de Haan

Radboud University, Netherlands

Advanced Dutch learners of English are generally considered to be highly proficient EFL users. However, in the discussion of what is an appropriate command for Dutch academics there is a tendency to ignore the specific language needs of *EFL professionals* (de Haan & van der Haagen, 2012), who are employed as teachers, translators, and editors of English. In the Netherlands, these EFL professionals are trained at universities, in full-time BA programmes of English language and culture. Dutch departments of English all aim to turn out graduates with a near-native command of English. It has been argued (de Haan & van der Haagen, 2012, 2013) that the current CEFR descriptors have not (yet) been operationalized properly in tests that assess students' EFL levels. For one thing, there appears to be a discrepancy between receptive and productive skills. Advanced university students of English who have achieved CEFR level C2 (usually in the course of their first or second year) are perfectly able to present a complex written argument (as C2 dictates), but their writing has, on the whole, a distinctly non-native ring to it. It is not clear whether this means that there is still a considerable gap between C2 and native usage, or that test results place students on C2 level too easily. This research aims to get a better understanding of EFL writing development at advanced levels; it is carried out within the LONGDALE project (Granger, 2009). My focus is on the syntactic make-up of EFL writing. What distinguishes much of my students' writing from native English writing is that it is syntactically made up as narrative rather than expository writing. Expository writing has distinctly fewer subject NPs than narrative prose. It is important for future EFL professionals to be aware of these differences as a narrative prose style is often not appropriate in academic writing. At the same time, there are noticeable differences between more successful and less successful EFL writers, even if their academic achievements and their general knowledge of English are comparable. Studying these differences longitudinally offers a better insight into the successful students' behaviour and strategies, which the less successful students can be encouraged to adopt. The specific research question that my presentation addresses is whether NP distribution and NP structure in EFL writing can tell us something about its level. Longitudinal data have been collected from a small random selection of Dutch students of English at Radboud University Nijmegen, of the 2011 cohort. These data were analysed for NP structure and distribution. Individual development paths were studied. NP distribution patterns displayed by more successful students were seen to correspond to those in native expository prose. Those of less successful students did not correspond to expository prose, and were characterised by a high frequency of NPs in verb complement positions, suggesting relatively simple sentence structures.

### References

- de Haan, Pieter, and Monique van der Haagen (2012). Modification of Adjectives in Very Advanced Dutch EFL Writing: A Development Study, in *The European Journal of Applied Linguistics and EFL*, 1(1): 129-142.
- de Haan, Pieter, and Monique van der Haagen (2013). Assessing the Use of Sophisticated EFL Writing: A Longitudinal Study, in *Dutch Journal of Applied Linguistics*, 2(1), 16- 27.
- Granger, Sylviane (2009). LONGDALE. Retrieved 22 October, 2010, from <http://www.uclouvain.be/en-cecl-longdale.html>

### Scalarity in the development of focus adverbs with quantifying origins. The case of *mostly* and *at least*.

Tine Defour

Ghent University, Belgium

While the notion of scale is traditionally associated with degree modifiers and adjectives, scalarity also plays a determining role in the semantic-pragmatic description of (particularizing) focus adverbs. Particularizers – such as *particularly*, *especially* or *at least* – limit the application of an utterance *predominantly* (though not exclusively) to a focused constituent (cf. Quirk et al. 1985; Nevalainen 1991). In doing so, they allow speakers to include clear manifestations of a relevant property in the adverbial focus, and exclude implied ‘alternative’ values that may not manifest the property as clearly, ranking all relevant values on a speaker-based scale of semantic strength (König 1991; Traugott 2009). This type of subjective grading results in different degrees of inclusion for different particularisers, ranging from a strong focus (e.g. *specifically*) to a weak one (e.g. *at least*) (cf. Ernst 1984). The focus adverbs selected for this paper, i.e. *mostly* and *at least*, both derive from Old English adjectives with quantifying meanings (*læst* and *mæst*, i.e. ‘smallest vs. greatest’ in size or degree). Despite similarities in their source meanings, the adverbs follow distinct semantic-pragmatic and structural trajectories. *At (the) least* mainly modifies quantitative constituents in Middle English data, indicating clear measures or numbers (1). Later focus modification also includes non-measurable constituents, with increasingly scalar readings indicating that the focused constituent forms the lower limit on a scale of acceptable values (2). *Most(ly)* moves from a concrete adverbial use (signifying ‘to the greatest extent’) with limited focus modification (3), to a scalar marker indicating that a specific value is applicable to the *greater part* of a broader group (4).

(1) I beseche your Grace to send more money, **at the leeste** viij Mli [4,000]. (CEECS 1418-1529)

(2) whosoever should search his study (or **at least** his Cabinet) should like enough finde all the libells that ever were made against the Queene. (CED 1603)

(3) the partes that shall be **mostly** touched, be the hart, chest, brayne. (OED 1580)

(4) They were **mostly** good guys, just fellows down on their luck. (OED 1987)

Our aim is to trace the forms’ semantic and structural diversifications back to their respective source contexts, and interpret them within a broader frame of grammaticalization and subjectification theories. This enables us to attest a) whether hypothesized shifts from a limited to broader range of modified focus constituents, and from concrete to scalar meanings coincides with higher frequencies and more advanced levels of grammaticalization and subjectivity (Nevalainen 1991; Traugott 2009), and b) in which ways scalar aspects in early quantifying meanings can trigger later developments towards scalar focus modification (Traugott 2009). The material for this paper is taken from a selection of historical corpora, including the *Helsinki Corpus*, the *Corpus of Early English Correspondence (Sampler)*, *A Corpus of English Dialogues 1560-1760*, the *Corpus of Late Modern English Texts* and ARCHER. Results are based on a detailed analysis taking into account frequencies, semantic-pragmatic changes and structural properties (focus constituents, scope).

### References

- Ernst, Thomas Boyden (1984) *Towards an Integrated Theory of Adverb Position in English*. Bloomington: Indiana University Linguistics Club.
- König, Ekkehard (1991) *The meaning of focus particles. A comparative perspective*. London: Routledge.
- Nevalainen, Terttu (1991) *But, only, just. Focusing adverbial change in Modern English 1500-1900*. Helsinki: Société Néophilologique.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- Traugott, Elizabeth Closs (2009) [2006] The semantic development of scalar focus modifiers. In *Handbook on the History of English*, Ans van Kemenade and Bettelou Los (eds.). Oxford: Blackwell, 335-359.

### Investigating the use of ‘violence’ metaphors by patients, family carers and healthcare professionals involved in end-of-life care

Jane Demmen<sup>i</sup>, Andrew Hardie<sup>i</sup>, Veronika Koller<sup>j</sup>, Paul Rayson<sup>i</sup>, Elena Semino<sup>i</sup>, Zsófia Demjén<sup>ii</sup>  
Lancaster University<sup>i</sup>, United Kingdom / Open University<sup>ii</sup>

‘War’ and ‘military’ metaphors, and ‘violence’ metaphors in general, are known to be used conventionally to talk about illness, and particularly terminal illness (e.g. “[someone] is battling cancer”). Such metaphors have also been criticized (e.g. by Sontag 1978) for their potentially negative implications, and in recent years they have been deliberately removed from policy documents on end-of-life care (EOLC) in the UK (e.g. the 2008 End of Life Care Strategy for England and Wales). In this paper we present findings on ‘violence’ metaphors from the ‘Metaphor in End-of-Life Care’ project, funded by the UK’s Economic and Social Research Council. The project combines quantitative semi-automatic corpus methods with manual qualitative analysis to investigate metaphor patterns in a 1,500,000-word corpus of data, split into equal sections from three stakeholder groups involved in EOLC: patients, family carers and healthcare professionals. The corpus comprises 300,000 words from semi-structured interviews and 1.2 million words from online forum contributions. Using a well-established analytical method (Pragglejaz Group 2007), a 90,000-word sample dataset was first analysed manually to identify metaphorical expressions, assisted by the online collaborative annotation tool eMargin. Lexical items deemed metaphorical were then allocated to broader semantic domains using a data-driven approach. These semantic domains were matched with established categories in the UCREL Semantic Analysis System tagger in an adapted version of the Wmatrix corpus software tools (Rayson 2008). This enabled us to explore the most relevant semantic domains in the whole corpus, and to study metaphor patterns systematically across the data. In our paper we will present quantitative and qualitative findings concerning patterns of use of ‘violence’ metaphors among the three stakeholder groups, to address three research questions:

- (1) What are the relative frequencies and types of ‘violence’ metaphors used by patients, carers and professionals involved in EOLC?
- (2) What similarities and differences are there in the ‘violence’ metaphors used by members of each stakeholder group?
- (3) What evidence is there for positive or negative implications of the ‘violence’ metaphors used?

We will demonstrate that ‘violence’ metaphors are used freely by patients, carers and professionals, and that the patients’ and carers’ language exhibits similarities in the frequency and type of metaphors employed, whilst the professionals’ language is somewhat different both in type and effect. These findings are to be expected to some extent given the role of the professionals in this context. However, our approach enables us to identify a wide range of ‘violence’ metaphors, and to make a nuanced assessment of the implications of their use by members of different stakeholder groups in different EOLC contexts. We will show that in some cases ‘violence’ metaphors seem to be self-empowering, for example in expressing personal determination or solidarity with others going through similar experiences. Our findings have implications for communication in medicine and healthcare, in that an appreciation of the different types and uses of ‘violence’ metaphors, in different contexts and by different stakeholders, is likely to be more helpful than a general condemnation of such metaphors in the context of (terminal) illness.

#### References

- Pragglejaz Group. 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1-39. Rayson, P. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519-549.
- Sontag, S. 1978. *Illness as a metaphor*. Harmondsworth, Middlesex: Penguin.

**Exploring the EFL-ESL paradigm gap: A two-step regression approach***Sandra Deshors, Stefan Gries**New Mexico State University, USA / University of California, USA*

The last 40-50 years have seen a large number of studies on alternations such as those traditionally referred to as dative movement, particle movement, preposition stranding, and many others. As the names of those alternations suggest, for many years, those studies adopted theoretical perspectives anchored in (transformational) generative grammar, which entails that (i) virtually all alternation studies focused on native speakers' mental grammars and (ii) the predominant methodology of study was acceptability judgments of (typically decontextualized) sentences.

The present study takes a diametrically opposite stance. We explore the dative alternation - the choice of *John gave Mary the book* or *John gave the book to Mary* - from the perspective of usage-/exemplar-based approaches, which hold that native and non-native speakers' mental grammars are based on the extraction and representation of large amounts of co-occurrence information on linguistic expressions and their (linguistic and non-linguistic) contexts. Our study involves native English speakers, but also French and German learners of English as a foreign language (EFL) as well as speakers of post-colonial second-language English (ESL) from Hong Kong, India, and Singapore. Finally, our study is based on data from different corpora that were annotated for twelve linguistic features previously shown to be correlated with the dative alternation in native Englishes (however, as previous work has shown, those correlations vary across native English varieties; cf. Bresnan & Hay 2008). Ultimately, we aim to emphasize how the reliance of much traditional learner corpus research on over-/underuse counts is problematic as its extreme decontextualization of linguistic items cannot easily be related to usage-based approaches to language acquisition and learning. In this paper we extend Gries & Deshors' (to appear) and Gries & Adelman's (to appear) methodological approach. Specifically, we identify 18 verbs that previous work (Gries & Stefanowitsch 2004) has shown to either prefer (i) the ditransitive or the prepositional dative or (ii) neither of the two constructions. We retrieved examples of those verbs in both dative constructions from the following spoken and written corpora: ICLE-FR/GE, LINDSEI-FR/GE, LOCNESS and LOCNEC as well as from the student essay/class lessons subsections of ICE-HK, ICE-IND, and ICE-SIN. These examples were manually annotated for parameters such as transitivity, voice, recipient/patient length, animacy, accessibility, pronominality, etc., and submitted to the MuPDAR approach which involves fitting a multifactorial mixed-effects regression model - i.e. a regression that can take speaker-, register-/mode- and verb-specific variation into consideration - on the native speaker data, and applying the resulting regression equation to the non-native speaker data. This allows us to determine where EFL/ESL users make non-nativelike choices. In a second multifactorial mixed-effects regression model, we study which characteristics of the dative alternation explain non-native speakers' deviation from the target variety. The results not only reveal systematic deviation patterns of EFL/ESL speakers from the native variety, but also showcase how taking usage-based theory and its implementation into learner corpus research seriously can help us contrast EFL and ESL speakers' mental grammars at an unprecedented level of granularity and identify how those grammars differ from those of native speakers.



### Corpora and context: New perspectives on the English of newspapers in the Caribbean

Dagmar Deuber

University of Muenster, Germany

Newspapers as a common text genre usually representing standard usage have been an important source in studies of written Standard English in the Caribbean (see Bruckmaier and Hackert 2011, Hänsel and Deuber 2013; also Mair 2002). However, previous studies suffer from a lack of comparability with regard to the data selected and phenomena investigated and they fail to consider the various islands of the Eastern Caribbean that are small on their own but non-negligible collectively. The present paper presents a comparative study, based on matching corpora, of the English of newspapers in Jamaica – population-wise the largest anglophone Caribbean country – and the small island nations of St. Kitts & Nevis, Dominica, and St. Vincent & the Grenadines. The focus is on features associated with American English, which previous studies have identified as a major influence. The corpus investigation is supplemented with considerations of the specific context of the corpus texts in terms of both text producers' and consumers' perspectives. Research questions are:

- (a) To what extent do American spellings and vocabulary items occur in the corpora?
- (b) What are text producers' and consumers' beliefs and attitudes concerning American spellings and vocabulary items?
- (c) Are distinctive American (dis)preferences for certain grammatical constructions, specifically that-relative clauses, contractions, and be-passives (Leech et al. 2009) discernible in the corpora?
- (d) How does the Jamaican corpus compare with the corpora for the small nations and what factors are relevant to this?

The paper combines quantitative corpus analyses and qualitative approaches to the context. The Caribbean corpora, of a size of 180,000 words each, have been compiled from online newspapers. The press sections of the BE06 and AE06 corpora are used for comparison. The context is investigated by telephone interviews with newspaper staff and content analysis of letters to the editor. Based on findings so far the following results are anticipated:

- (a) American spellings hardly occur in the Jamaican data but to varying degrees in the corpora for the small nations. Spelling is a conscious choice on the part of the text producers; some consumers are critical.
- (b) The corpora reveal a common preference for American vocabulary but text producers mostly believe to use British English; consumers rarely comment on this.
- (c) Distinctive American (dis)preferences for certain grammatical constructions are discernible only to a small degree.
- (d) Contributors to the newspapers of the small nations are more likely to have lived in the US for extensive periods; some still live there.

Overall the results somewhat temper previous findings on the role of American English in the region. With regard to vocabulary a clash between ideology and language use is revealed. Spelling is variable in the data from the small nations but hardly so in the Jamaica data, suggesting that the standard is less fixed in the former; otherwise common tendencies are observed. Methodologically the paper emphasizes the importance of considerations of both corpora and context where questions of standard language are concerned.

### References

- Bruckmaier, Elisabeth & Stephanie Hackert. 2011. "Bahamian standard English: A first approach." *English World-Wide* 32: 174-205.
- Hänsel, Eva Canan & Dagmar Deuber. 2013. "Globalization, postcolonial Englishes, and the English language press in Kenya, Singapore, and Trinidad and Tobago." *World Englishes* 32: 338-57.
- Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.
- Mair, Christian. 2002. "Creolisms in an emerging standard: Written English in Jamaica." *English World-Wide* 23: 31-58.

## Starting Skype conversations: Pragmatic features and strategies in an ELF context

Stefan Diemer, Marie-Louise Brunner, Selina Schmidt  
University of Saarland, Germany

This paper is concerned with conversation organisation features documented in CASE, the Corpus of Academic Spoken English, compiled at Saarland University in 2013/14, comprising roughly 150 hours of Skype conversations between EFL speakers from four different countries (Germany, Bulgaria, Italy and Spain). By describing the pragmatic features and strategies used in this multimodal academic context we endeavour to contribute to the exploration of spoken English as a lingua franca. English in an international academic context has developed distinct features in lexis, syntax and pragmatics (Diemer & Schmidt 2013, Conrad & Mauranen 2003, Mair 2003, Meierkord 1996). Several corpora have been compiled in this field, such as the English as Lingua Franca in Academic Settings corpus (Mauranen 2008) and the Vienna-Oxford International Corpus of English (Seidlhofer 2013). Recently, international spoken varieties have been attracting attention (e.g. Jenkins et al. 2001) as English plays a central part in academic and private computer-mediated communication (CMC), frequently between non-native speakers. CMC research has not yet explored the domain of Skype conversations due to the lack of respective specialised corpora. We aim at providing a resource that allows research in such varied areas as English as an international language, CMC, error analysis of learner language, pragmatics and lexical innovation. The particular focus in this paper is on pragmatic features and strategies in Skype conversation starts, as documented in CASE. We conceive of conversation starts as not only openings as categorised by Schegloff (1968), but also the ensuing introduction of the first topic. Conversation openings have been extensively researched in telephone conversations (Schegloff 1986, 2004, Couper-Kuhlen 2001) and, recently, in some computer-mediated communication (CMC) settings, such as vlogs (Frobenius 2011); however, not in a dialogic CMC setting using Skype. Openings in CASE generally follow Schegloff's classification, but have to be adapted to the international CMC environment. Topic introductions were researched by Geluykens (1993) whose structural categories can also be identified in CASE. Additionally, we propose a preliminary content-related classification of topic introductions in CMC spoken settings, distinguishing between assigned, task-oriented, and random topics, as well as hybrids. Another key issue in first contact situations between non-native speakers of English is rapport management. While several researchers comment on face-to-face rapport (Spencer-Oatey 2002), a CMC situation introduces several new factors, e.g. novelty and situational delicacy caused by CMC features. In the international English context, pragmatic competence is also essential. Learner mistakes and errors may influence conversational sequence and organisation, leading to misunderstandings and resulting in repair or accommodation. Finally, Skype as a medium influences both conversational content and structure, as speakers have to deal with echos, lags and interferences.

## References

- Couper-Kuhlen, Elizabeth. "Constructing reason-for-the-call turns in everyday telephone conversation." *InLiSt: Interaction and Linguistic Structures* 25 (2001). Web. 10 June 2013.
- Diemer, Stefan; et al. *CASE Sofia-Saarbrücken: Corpus of Academic Spoken English in an international context*. Saarbrücken: Saarland University / Sofia: St Kliment Ohridski University, forthcoming. Web.
- Diemer, Stefan, and Selina Schmidt. "Ja, also, can you see me now? - Designing and compiling a corpus of computer-mediated international academic English." *ICAME Journal* 38, 2014. Print.
- Frobenius, Maximiliane. "Beginning a monologue: The opening sequence of video blogs." *Journal of Pragmatics* 43 (2011): 814-827. Print.
- Geluykens, Ronald. "Topic introduction in English conversation." *Transactions of the Philological Society* 91.2. (2011): 181-214. Print.
- Conrad, Susan, and Anna Mauranen. "The corpus of English as lingua franca in academic settings." *TESOL Quarterly* 37.3 (2003): 513-527. Print.
- Jenkins, Jennifer, and Markus Modiano, and Barbara Seidlhofer. "Euro-English." *English Today* 17.4 (2001): 13-19. Print.
- Mair, Christian. (ed.). *The Politics of English as a World Language*. Amsterdam: Rodopi, 2003. Print.
- Mauranen, Anna. *ELFA. The Corpus of English as a Lingua Franca in Academic Settings*, 2008. Web. 23 November 2013.
- Meierkord, Christiane. *Englisch als Medium der interkulturellen Kommunikation. Untersuchungen zum non-native-/non-native speaker-Diskurs*. Frankfurt: Lang, 1996. Print.
- Seidlhofer, Barbara. *VOICE. The Vienna-Oxford International Corpus of English* (version 2.0 Online), 2013. Web. 23 November 2013.
- Spencer-Oatey, Helen. "Managing rapport in talk: Using rapport sensitive incidents to explore the motivational concerns underlying the management of relations." *Journal of Pragmatics* 34 (2002): 529-545. Print.
- Schegloff, Emanuel A. "Sequencing in conversational openings." *American Anthropologist* 70 (1968): 1075-1095. Print.
- Schegloff, Emanuel A. "The routine as achievement." *Human Studies* 9 (1986): 111-151. Print. Schegloff, Emanuel A. "Answering the phone." *Conversation analysis: Studies from the first generation*. Ed. Gene H. Lerner. Amsterdam/Philadelphia: John Benjamins, 2004. 63- 129. Print.

### Geo-referenced collocation analysis: tracking evolving emotional responses to place and landscape

Christopher Donaldson, Ian Gregory, Andrew Hardie, Patricia Murrieta-Flores  
Lancaster University, United Kingdom

This paper exemplifies a corpus-based/geographical approach to the analysis of meaning within text corpora. We address an active issue in the study of writers' aesthetic responses to landscapes in the eighteenth and nineteenth centuries, with particular reference to the Lake District. Literary scholars and historians have often remarked that Lake District tourists of this period moved from viewpoint to viewpoint, collecting sensations in much the same way that modern tourists collect snapshots and souvenirs. Most such social histories (see e.g. Nicholson 1955, Walton and Wood 2013) assume, largely on qualitative-impressionistic grounds, that as aesthetic sensibility evolved – from the late-C18 fashion for the picturesque, to subsequent fashion for more concrete landscape aesthetics, and ultimately in the later C19 to differentiated religious and scientific appreciations of the natural world – the language used to describe these sensations likewise shifted. We test this hypothesis of an evolution in the aesthetic terminology used in response to place and landscape in a custom-built 1,500,000-word corpus of Lake District travel literature, 1622-1900. This corpus contains 80 texts (by authors including Wordsworth and Coleridge among other less famous names), stratified according to sub-period. Our (generalisable) method represents an elaboration of the techniques of (\*\*\*\*), who report a basic procedure in which statistical collocation analysis is used to extract links between mentions of place-names in a corpus and co-occurring elements of lexis or annotation that are associated with some concept of interest. Subsequently techniques of Geographical Information Systems (GIS) are employed to link each such mention to mappable coordinates (identified via *geo-parsing*), allowing a range of visualisations and spatially-aware statistical analyses. The three main stages here are (1) selection of search terms, (2) use of collocation statistics to identify meaningful links to mentions of place-names; and (3) use of the geo-references annotated on all such mentions in the corpus to create a GIS from which maps may be plotted and then analysed spatially and statistically. Each of these stages raises methodological questions. First, how should a suitable set of conceptually-linked terms be chosen? Second, what operationalization – in terms of window/statistic – is most suitable in this context? Third, what GIS methods lend themselves best to analysing the resulting collocations? Literary theory may suggest an initial core of relevant search terms (e.g. *sublime*, *majestic*, *picturesque*, *romantic*, *beautiful*, *charming*); however, our data suggests that better results emerge if a broader set of node words, determined by reference to some ontology (in our case, the USAS lexicon: Rayson et al. 2004), is deployed. In terms of collocation methodology, we explore the effect of “bandwidths” beginning with the traditional +/-4 tokens and then extending outwards, alongside the Mutual Information and Z-score measures. Initial exploration of the set of geo-referenced collocations suggest that the technique of density smoothing is the most appropriate visualisation for general analysis. Finally, an application of the resulting techniques across the stratification of texts by period within the corpus allows us to assess the extent to which quantitative corpus/GIS analysis supports the impressionistic findings of literary scholars.

### References

- Nicholson, N. (1955) *The Lakers: The Adventures of the First Tourists*. London: Robert Hall.
- Rayson, P., Archer, D., Piao, S. L., McEnery, T. (2004). The UCREL semantic analysis system. In: Proceedings of the workshop on *Beyond Named Entity Recognition Semantic labelling for NLP tasks*, in association with 4th International Conference on Language Resources and Evaluation (LREC 2004), 25th May 2004, Lisbon, Portugal, pp. 7-12. Available online at [http://www.comp.lancs.ac.uk/~paul/publications/usas\\_lrec04ws.pdf](http://www.comp.lancs.ac.uk/~paul/publications/usas_lrec04ws.pdf)
- Walton, J. K. and Wood, J. (2013) *The Making of a Cultural Landscape: The English Lake District as a Tourist Destination*. Ashgate.

## Representations of the FUTURE in English language blogs on climate change

Kjersti Fløttum<sup>i</sup>, Øyvind Gjerstad<sup>i</sup>, Anje Müller Gjesdal<sup>i</sup>, Nelya Koteyko<sup>ii</sup>, Andrew Salway<sup>iii</sup>  
 University of Bergen, Norway<sup>i</sup> / University of Leicester, United Kingdom<sup>ii</sup> / Uni research, Norway<sup>iii</sup>

This paper investigates how the notion of “future” is represented in English-language blogs related to climate change, with an overarching theme of “gloom and doom” perspectives versus more positive perspectives of a “sustainable low carbon society”. The relation between the threats of climate change, proposed solutions to the challenges, and the potential shape of future societies is of great societal importance. The blogosphere has been a major site for large-scale and complex discourses about these issues. Even though the question of the future is raised in numerous public debates on climate change, its representations have so far received little attention in corpus linguistic research. Thus, our research interest lies around three main questions: 1) To what extent do representations of the future exist in public debates related to climate change? 2) How are the representations expressed linguistically? 3) What meanings are the representations conveying? Additionally we are interested in what contexts future representations are present (local, national, international), the extent to which they are related to emotions such as hope and fear, and the extent to which they are related to culture, values and ethical questions. Our approach is an exploratory content analysis of blog material, using corpus-assisted discourse analysis methods/techniques. We are working with material from the NTAP blog corpus (Salway, Hofland and Touileb 2013): this comprises 1.5m English-language blog posts from around 3,000 blogs related to climate change. For our initial analysis we created a corpus of the sentences containing ‘climate change’ (209,107 sentences; 3.6m words) and ‘global warming’ (124,092 sentences; 3.0m words). The analysis concentrated on the use of the word ‘future’ and other frequent semantically related words. A semi-automated analysis using sorted advanced concordances and word clusters in the AntConc tool (Anthony 2011) gave about 40 simple patterns containing these words. Our ongoing research is analyzing the cotexts and contexts of these patterns with regards to our research questions. Patterns around ‘future’ point to a variety of representations including ‘future’ compounds, which are more or less value-laden (e.g. a bleak|uncertain|low carbon future), some including time perspective (in the near|not so distant|distant future), some indicating “ownership” (the|our|their future). In order to examine to what extent and how action/inaction is insisted on in relation to the future, patterns around ‘should’ and ‘must’ (in their deontic meaning) are being analyzed: here we see interesting variation between who ‘should|must’ act and what ‘should|must’ be done. Further, to examine the extent to which the notion of future is related to negative and positive perspectives, we look to patterns around ‘danger|risk|threat’ and ‘opportunity’. In these patterns we also see variation between the use of verbs such as ‘combat|fight|tackle’ and ‘address|assess|understand’. Having identified a set of ‘future representations’ from all these patterns, we will then investigate how they are attributed to different voices and blog communities (e.g. skeptic and non-skeptic), and how their occurrences change over time (the blog corpus spans 2006-2012).

## References

- Anthony, L. 2011. AntConc (Version 3.2.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>
- Kurz, T., Augoustinos, M. and S. Crabb. 2010. Contesting the ‘national interest’ and maintaining ‘our lifestyle’: A discursive analysis of political rhetoric around climate change. *British Journal of Social Psychology* 49, 601–625.
- Leiserowitz, A. 2006. Climate change risk perception and policy preferences: the role of affect, imagery, and values. *Climatic Change* 77, 45–72.
- Salway, A., Hofland, K. and S. Touileb. 2013. Applying Corpus Techniques to Climate Change Blogs. Procs. Corpus Linguistics 2013, University of Lancaster.

### Quantitative behaviour of signalling nouns in academic discourse

John Flowerdew

City University of Hong Kong, China

#### Research problem

Abstract nouns like *argument*, *fact*, *idea*, and *problem* (referred to variously as carrier nouns (Ivanič, 1991), shell nouns (Schmid, 2005), and signalling nouns (SNs) (author)) are pervasive in academic language and barely a sentence goes by which does not contain one or more of these nouns. Such nouns are essential to the cohesion of a text, on the one hand, and to the process of academic argumentation, on the other. While previous work by the author has shown various aspects of the behaviour of SNs in academic discourse from a qualitative perspective, quantitative data has not been available to date. This paper, accordingly, represents a response to a range of research questions regarding the behaviour of SNs in academic discourse from a quantitative perspective.

#### Research questions

1. What is the overall frequency of occurrence of SNs across the corpus?
2. What is the relative frequency of the most frequent SN types?
3. What is the frequency of occurrence of SNs across the three sub-corpora of lectures, book chapters and research articles?
4. What is the relative frequency of the most frequent SN types across the three sub-corpora of lectures, book chapters and research articles?
5. What is the comparative total frequency of SNs across natural sciences and social sciences?
6. What is the comparative frequency of SNs across the ten disciplines represented in the corpus?

#### Corpus and method

Consisting of 650,000 words of text drawn equally from lectures, textbooks and research articles and matched for content across ten disciplines, the corpus used in the study can be described as a very large small corpus. All SNs in the corpus were annotated according to a set of semantic categories derived inductively from the data and according to the various syntactic patterns which they occur in. The annotation was done by a combination of automated and manual techniques. Manual techniques were needed because automated techniques are not available to date for SNs. In order to ensure reliability of the annotation, three trained research assistants participated in the annotation process, along with the researcher. Software used in the study was Antconc (Version 3.2.2) (Anthony, 2011) and word frequency and concordancing functions were used in the analysis.

#### Findings

The findings include the following: a ranking of the most frequent SNs in the corpus in order of frequency; a ranking of the most frequent SNs across the three sub-corpora of lectures, book chapters and research articles; comparative total frequency of SNs across natural sciences and social sciences; and comparative frequency across the ten disciplines represented in the corpus. Implications of each of these findings will be discussed. The paper represents a contribution to our knowledge of how SNs function in academic discourse.

#### References

- Anthony, L. (2011). AntConc (Version 3.2.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>
- Ivanič, R. (1991). Nouns in search of a context: a study of nouns with both open- and closed-system characteristics. *International Review of Applied Linguistics*, 29 (2), 93–114.
- Schmid, H-J. (2000). *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin: Mouton de Gruyter.

### Learner Corpus Research in ESP/EAP: Some core issues and future directions

*Lynne Flowerdew*  
*HKUST, China*

The issues of corpus size, corpus-based vs. corpus-driven enquiries and interpretation of corpus data have been much discussed in textbooks and studies on corpus linguistics in general. However, these issues have been less debated in the literature on learner corpus research. My aim in this presentation is to critically examine these three issues with respect to key studies in the area and also map out some directions for future work in learner corpus research of ESP/EAP written text. The size of a learner corpus is dependent on a number of factors. As a general rule, more broad-based quantitative studies make use of corpora of around 500,000 words. However, when qualitative methods are employed, smaller corpora, ranging from 50,000 to 150,000 words, tend to be used. Moreover, the theoretical perspective of the analyst, the type of item under investigation, the level of delicacy of the analysis, how specialized the corpus is, and whether the corpus is manually tagged or items are retrieved automatically are other considerations determining the size of a learner corpus. Most studies of learner corpora adopt a corpus-based approach, for the key reason that error tagging is employed. Those studies which are purely corpus-driven tend to be based on learner corpora which are relatively small. This is not so surprising given that such types of investigation are conducted inductively requiring close attention, with the linguistic categories arising from the corpus data. The question of possible explanations for interlanguage features in learner corpora is not an easy issue as there are a number of variables to take into consideration and analysts are usually quite cautious in offering explanations. Factors accounting for interlanguage features include L1 transfer, developmental aspects, teaching input, and genre/register and cultural influences. Often it is difficult to pin down an interlanguage feature to a single factor. Neither is the question of L1 transfer as straightforward as it might at first appear and a contrastive interlanguage comparison of L1 and L2 learner corpora may not be sufficient to indicate L1 transfer. There are a number of newly-emerging developments which are already in train, as can be seen by way of comparison with past practices. Whereas previously cross-sectional learner corpus studies dominated the research, there is now more alignment between learner corpus research and SLA with more focus on longitudinal studies. Statistical procedures are also undergoing refinement. In the past learner corpus data tended to be aggregated, but it is now recognized that it is better practice to consider individual data such that the number of students contributing to the token count is known. Learner corpus data is performance-based with a focus on product. However, initiatives are underway to incorporate a more ethnographic perspective, with a view to capturing writing as process. While the field of learner corpus research continues to thrive, it would be timely to take stock of some core theoretical and methodological issues and also ongoing and future developments, as raised in this abstract.

## Recent Change in the Use of the Progressive in Nigerian English: An Apparent Time Study

Robert Fuchs, Gut Ulrike, Presley Ifukor  
University of Münster, Germany

Based on an apparent time approach and using data from the ICE-Nigeria, this paper traces recent change in the use of the progressive in Nigerian English (NigE). Gut and Fuchs (2013) showed that in NigE the progressive is used with stative and punctual verbs, and is particularly frequent in registers that involved discussion and an exchange of arguments ("persuasive registers"), while in British English (BrE) the progressive is frequent in informal registers. The progressive has also witnessed a dramatic increase in frequency in British and American English (Leech et al. 2009, Mair and Hundt 1995). However, little is known about recent change in the progressive in African and Asian varieties of English. The present paper addresses this lack of research, based on a novel approach that applies the apparent-time method to corpus data. It is the premise of apparent-time studies that differences between older and younger speakers observable at one point of time can be interpreted as evidence of language change: older speakers represent older stages, and younger speakers more recent stages in the development of a language variety. ICE corpora have been used as the basis for indirect, apparent-time studies previously - Collins (2009) and Van der Auwera et al. (2012) have extrapolated directions of ongoing change from differences between the spoken and written parts of ICE corpora, based on the assumption that changes tend to be more advanced in spoken than in written texts. However, ICE corpora have not been used intensively yet to explore age-grading effects (but see Höhn 2012 for the usage of quotatives). The data for this study is taken from the ICE Nigeria (Wunder et al. 2010), comprising language produced by 1191 educated Nigerians. A total of 4,813 progressive constructions in Nigerian English were thus analysed, and factors determining (1) the frequency of the progressive and (2) the frequency of extended uses were analysed with logistic regression models. Results indicate that the progressive is overall becoming less frequent in NigE, but that it is becoming more frequent in persuasive registers. Among older speakers there are differences in the use of the progressive between ethnic groups, but younger speakers tend to converge in their usage regardless of ethnic group. The use of the progressive with punctual and stative verbs is particularly frequent in persuasive registers, but overall punctual progressives are becoming less frequent. Only the use of the progressive with stative verbs appears to be a stable feature of NigE. Comparing our results with previous research on other varieties, NigE appears to be an exception to the general trend of the increasing frequency of the progressive. Another difference appears to be the continuing preference for the progressive in persuasive registers. By contrast, with regard to another stable feature, the use of the progressive with stative verbs, NigE may be at the forefront of a trend that has also been documented for other varieties of English (Leech et al. 2009).

## References

- Collins, P. 2009. Modals and quasi-modals in world Englishes. *World Englishes* 28, 281-92.
- Gut, U. and Fuchs, R. 2013. Progressive aspect in Nigerian English. *Journal of English Linguistics*
- Höhn, N. 2012. 'And they were all like 'What's going on?': New quotatives in Jamaican and Irish English. In Hundt, M. and Gut, U. (eds), *Mapping Unity and Diversity of New Englishes World-Wide.*: John Benjamins, 263-290.
- Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English. A grammatical study.* Cambridge: Cambridge University Press.
- Mair, Christian & Marianne Hundt. 1995. Why is the progressive becoming more frequent in English? *Zeitschrift für Anglistik und Amerikanistik* 2. 111-122.
- Smitterberg, Erik. 2005. *The progressive in 19<sup>th</sup>-century English.* Amsterdam: Rodopi.
- Van der Auwera, J., Noël, D. and De Wit, A. 2012. The diverging need (to)'s of Asia Englishes. In Hundt, M. and Gut, U. (eds), *Mapping Unity and Diversity of New Englishes World-Wide.*: John Benjamins. 55-7.
- Wunder, Eva-Maria, Holger Voormann & Ulrike Gut. 2010. The ICE Nigeria corpus project: Creating an open, rich and accurate corpus. *ICAME Journal* 34. 78-88.

**Life and death, living and dying, alive and dead: The prevalence of super-categorial relations**

Gregory Garretson  
Uppsala University, Sweden

Collocation is one of the central concepts of corpus linguistics, and yet it is defined in a multitude of ways. While some researchers (e.g. Sinclair 2004, Hoey 2005) maintain that the associations of every form of a lexical item (e.g. *die*, *died*) should be considered separately, others (e.g. Benson *et al.* 1997, Nesselhauf 2005) normally consider inflected forms together (e.g. *die*, *dies*, *died*, *dying*). Yet even these scholars seldom investigate related lemmas with different parts of speech, such as *die* (V), *death* (N), and *dead* (ADJ). However, if we do examine the syntagmatic relations of such “families”, the results are striking. The examples in (1), from the British National Corpus (Aston & Burnard 1998), demonstrate the super-categorial association between the lexical families “*conclude*, *conclusion*” and “*satisfy*, *satisfaction*, *satisfying*, *satisfactory*, *satisfactorily*”:

- (1) resolved to bring the whole matter to a satisfactory conclusion  
 bringing any of the characters to a satisfying conclusion  
 come to a **conclusion** which satisfied  
 I reach that **conclusion** with no satisfaction  
 for the negotiations between our two parties to be **concluded** satisfactorily

But how frequent is this phenomenon? This question has an important bearing on our ideas about collocation. This talk reports on a study aiming to investigate the prevalence of super-categorial co-occurrence relations in English—specifically, a study of 100 families of related lemmas (including inflected forms), such as the one in (2).

- (2) die/dies/died/dying  
 death/deaths  
 dead/deader  
 deadly/deadlier/deadliest etc.

While individual members of such families are known to collocate with members of other families (e.g. *life+death*), this study shows that there are many associations between *families*: *life+death*, *live+die*, *alive+dead*, *living+dead*, *life+die*, *living+death*, etc. As can be seen, many of these collocating pairs are not of the same syntactic category. In fact, super-categorial relations are found to exhibit many configurations of parts of speech. The study used as its data the BNC in its entirety. A computer program was developed for grouping lemmas in the BNC into families (e.g. *probable*, *probably*, *probability*, *improbable*) based on a combination of spelling similarities and morphological rules; the resulting families were manually edited, and 100 were chosen for the study, based on their frequency. A second program was developed that took these families (including 500+ lemmas, 1000+ forms) and searched for co-occurrences. The results, grouped by family, include counts for all observed co-occurring form-pairings (e.g. *clearly+see*, *clear+see*, *clarify+see*, etc.). The results demonstrate that super-categorial relations are remarkably frequent: The 100 families co-occur in the corpus with an average of 7.1 different form-pairings per pair of families, and at times as many as 50. The method used here, because it considers groups of related forms (some of which are infrequent and would not normally be noted), yields an unusually broad coverage and reveals patterns of association in lexis that most approaches to collocation obscure. I will conclude by discussing the repercussions for theories of syntagmatic relations.

## References

- Aston, G. & Burnard, L. 1998. *The BNC handbook: exploring the British National Corpus with SARA*. Edinburgh textbooks in empirical linguistics. Edinburgh: Edinburgh University Press.
- Benson, M., Benson, E., & Ilson, R. 1997. *The BBI dictionary of English word combinations (Rev. ed.)*. Amsterdam; Philadelphia: John Benjamins.
- Hoey, M. 2005. *Lexical priming: a new theory of words and language*. London: Routledge.
- Nesselhauf, N. 2005. *Collocations in a learner corpus*. Studies in corpus linguistics; v. 14. Amsterdam: John Benjamins.
- Sinclair, J. M. 2004. *Trust the text: language, corpus and discourse*. London: Routledge.



**The use of the high-frequency verb *make* by German learners of English: Shifting the focus to individual learners and L1 varieties of native speakers**

Caroline Gerckens

Leibniz Universität, Germany & Université catholique de Louvain, Belgium

This paper focuses on the use of the high-frequency verb *make* in the interlanguage of intermediate-advanced German learners of English based on data from the German sub-corpus of the *Longitudinal Database of Learner English* (LONGDALE), which is one of the few existing longitudinal learner corpora. Researchers agree that high-frequency verbs are problematic for learners, even at advanced levels (e.g. Hasselgren, 1994; Altenberg & Granger, 2001; Juknevičienė, 2008; Gilquin & Viberg, 2009), and particularly when part of multi-word units. To sound native-like, a phraseological competence is essential since “native speaker linguistic competence has a large and significant phraseological component” (Howarth, 1998: 29). However, the development of the learner phrasicon, which can only be studied by using longitudinal data, is a “greatly under-researched field” (Granger, 2009: 62) with only few exceptions (e.g. Chen, 2013). This paper contributes to filling this “collocational gap” (Barfield & Gyllstad, 2009: 1) by tackling the phraseological performance of intermediate-advanced German learners of English over a period of three years. The learner data for this study comes from the first cohort of LONGDALE-GE and is of two types: learner narratives (three sub-corpora) and argumentative essays (three sub-corpora) which have been collected since October 2011. As reference corpora for the argumentative data, all three sub-corpora of LOCNESS were used in order to show that a study’s outcome is influenced by the choice of reference corpus. The learner narratives, however, have been analysed longitudinally only and have not been compared to native speaker data in the absence of a comparable corpus. The main emphasis of this paper is on the verb *make* and its different uses such as delexical or causative structures. All occurrences of *make* were automatically retrieved with WordSmith Tools 6.0 (Scott, 2012) and then manually checked. In general, a tendency for an underuse can be detected. More apparent, however, is the considerable variation in the native-speaker data. The study shows that German learners of English significantly underuse *make* when compared to data from LOCNESS-US, while they use it with a similar frequency to the British students. An in-depth qualitative analysis was then carried out following the example of Altenberg and Granger’s study (2001). The results of this study contribute to learner corpus research in several ways. First, it sheds light on the longitudinal development of intermediate-advanced learners of English. There is, for instance, a decrease in errors in delexical uses. Second, inter-learner variability is highlighted (cf. Callies, 2013; Mukherjee, 2009) because the data is often unevenly distributed. The study shows that a small number of students are responsible for a considerable amount of data. Third, the issue of corpus comparability (cf. Callies, 2013; Chen, 2013; Granger & Paquot, 2009) is addressed because the choice of reference corpus strongly influences a study’s outcome.

## References

- Altenberg, B. & S. Granger. 2001. The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22 (2), 173-194.
- Barfield, A. & H. Gyllstad. 2009. *Researching Collocations in Another Language. Multiple Interpretations*. Palgrave Macmillan: Basingstoke & New York.
- Callies, M. 2013. Agentivity as a determinant of lexico-grammatical variation in L2 academic writing. *International Journal of Corpus Linguistics*, 18 (3), 357-390.
- Chen, M. 2013. Overuse or underuse. A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics*, 18 (3), 418-442.
- Gilquin, G. & A. Viberg. 2009. How to make do with one verb: a corpus-based contrastive study of DO and MAKE. *Revue française de linguistique appliquée*, 14(1), 67-82.
- Granger, S. & M. Paquot. 2009. Lexical verbs in academic discourse. A corpus-driven study of learner use. In: M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic Writing. At the Interface of Corpus and Discourse*. London/New York: Continuum, 193-214.
- Granger, S. 2009. Commentary on Part I. Learner Corpora. A Window onto the L2 Phrasicon. In: Barfield, A. & H. Gyllstad. ed(s). *Researching Collocations in Another Language. Multiple Interpretations*. Palgrave Macmillan: Basingstoke & New York, 60-65.
- Hasselgren, A. 1994. Lexical teddy bears and advanced learners. A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4, 237-259.
- Howarth, P. 1998. Phraseology and Second Language Proficiency. *Applied Linguistics* 19(1), 24-44.
- Juknevičienė, R. 2008. Collocations with high-frequency verbs in learner English. Lithuanian Learners vs. Native Speakers. *Kalbotyra*, 59(3), 119-127.
- Mukherjee, J. 2009. The grammar of conversation in advanced spoken learner English. Learner corpus data and language-pedagogical implications. In: K. Aijmer (Ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, 203-230.
- Scott, M. 2012. *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software.

## Periphrastic causative constructions in EFL and ESL: The role of acquisition context

Gaëtanelle Gilquin

University of Louvain, Belgium

English as a Foreign Language (EFL) and English as a Second Language (ESL) differ in their context of acquisition – mainly instructional in the case of EFL (classroom setting) and mainly naturalistic in the case of ESL (everyday interactions). Given a usage-based perspective on language acquisition, which views linguistic competence as the result of one's "accumulated experience with language across the totality of usage events in [one's] life" (Tomasello 2001: 62), we may expect this difference to have an effect on learners' knowledge of English. In particular, it can be hypothesized that the higher degree of exposure to (authentic) language in the ESL context will lead to a better and more idiomatic knowledge of the language than is the case in the EFL context. This hypothesis was tested by means of a corpus-based analysis of periphrastic causative constructions, which have been studied in EFL (e.g. Gilquin 2012) and ESL (e.g. Ziegeler & Lee 2009) but whose behaviour in the two contexts has never been compared. Using data from the *International Corpus of Learner English* for EFL and from the *International Corpus of English* for ESL, as well as a corpus of native English as a reference, I examined the frequency and well-formedness of periphrastic causative constructions in EFL and ESL, but also their idiomaticity through a collocation analysis (Stefanowitsch & Gries 2003) of the lexemes occurring in the non-finite verb slot. The results show that, from a formal point of view, ESL students do not necessarily use periphrastic causative constructions more accurately than the EFL students, but the non-standard constructions they use tend to follow some general tendencies of the English language. Thus, the construction [X MAKE Y V<sub>to-int</sub>], which is frequently found in ESL, corresponds to the general pattern of non-finite complementation, in which *to*-infinitive clauses are more common than bare infinitive clauses (see Biber et al. 1999: 699). By contrast, the EFL data display more varied, and apparently random, non-standard constructions, which often go against the general preferences of the English language, e.g. the use of an *ing*-form rather than a *to*-infinitive with the verb *cause*, while *ing*-clauses are generally less common than *to*-infinitive clauses in English (see Biber et al. 1999: 754). Although they do not exactly confirm the initial hypothesis, these results still support a usage-based view of language acquisition, in which ESL students, thanks to the larger (and richer) amount of input they receive, can better approximate construction schemas found in the English language. From a phraseological point of view, preliminary results suggest that periphrastic causative constructions are more idiomatic in ESL than in EFL, which might indicate that phraseology benefits more from exposure to authentic language than formal aspects. Finally, it appears that EFL and ESL also share some features, such as the use of a redundant *be* verb with *make* (e.g. *Our illusions and creative imagination make us be different*). These similarities underline that both EFL and ESL are non-native varieties, which are likely to involve common cognitive principles of language acquisition like redundancy or explicitation.

## References

- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education Limited.
- Gilquin, G. 2012. Lexical infelicity in English causative constructions. Comparing native and learner collocations. In J. Leino & R. von Waldenfels (eds) *Analytical Causatives. From 'give' and 'come' to 'let' and 'make'*, pp. 41-63. München: Lincom Europa.
- Stefanowitsch, A. & S. Th. Gries. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8.2: 209-243.
- Tomasello, M. 2001. First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11(1/2): 61-82.
- Ziegeler, D. & S. Lee. 2009. A metonymic analysis of Singaporean and Malaysian English causative constructions. In K.-U. Panther, L. L. Thornburg & A. Barcelona (eds) *Metonymy and Metaphor in Grammar*, pp. 291-322. Amsterdam & Philadelphia: John Benjamins.

### The most underused method in corpus linguistics: Multi-level and mixed-effects models

Stefan Gries

University of California, USA

For several decades, the statistical analysis of especially experimental data in psycholinguistics was characterized by the recognition that the data points collected in an experiment are not independent of each other because (i) each subject provides more than one judgment/reaction time/... and (ii) each experimental stimulus is judged/reacted to more than once. This interrelatedness of the data was taken into consideration typically by computing two repeated-measures ANOVAs, one averaging across subjects, one averaging across experimental stimuli. In the last 10 years or so, psycholinguists have turned to mixed-effects modeling (MEM) as a superior analytical tool. Interestingly enough, much corpus-linguistic work has not even embraced multifactorial modeling in general, let alone MEM, although corpus linguists stand to benefit from these tools even more than psycholinguists. This is for three reasons: (i) corpus data exhibit the same interrelatedness of data points (speakers/writers and lexical items) as psycholinguistic data; (ii) corpus data are usually much messier/noisier than psycholinguistic data; (iii) corpus data often come with a hierarchical sampling structure such as the ICE-GB's structure. Crucially, as corpus linguists we know that corpora come in these structures and that nearly every phenomenon we study will exhibit differences on the levels of speakers/lexical items *and* on the levels of the mode and/or the register and/or the subregister – however, while some corpus-linguistic studies now use MEM to address the former level of variability, there is virtually no work at all that also accounts for levels of corpus organization using multi-level modeling (MLM). In this paper, I discuss the constituent order alternation of particle placement in the ICE-GB – *Mary gave up [DO smoking]* vs. *Mary gave [DO smoking] up* – and show how we can relatively straightforwardly study corpus data correctly by taking into account all the levels of variability that we know exist in corpora but that we – and that includes much of my own earlier work – routinely ignore. In particular, I will show how the results of this approach are far superior to those of traditional regression modeling in terms of classification accuracy (a highly significant increase of 9%) and  $R^2$  (an increase in 0.06), but especially regarding the precision and interpretation of the results. For example, the results show how the (hypothesized) effects of the length of the DO and the type of the head of the DO need to be looked at differently at the level of the subregister (rather than the mode or the register). In particular, the subregisters do not differ much with regard to how length affects particle placement with non-lexical DO heads, but differ considerably with lexical DO heads, which not only makes the whole analytical process more precise, but also indicated that subsequent analysis of this alternation could benefit from subregister-based analysis (e.g., of a Biberian kind).

### ***Corpus-based dialectology: Regional grammatical variation in a corpus of written American English***

*Jack Grieve*

*Aston University, United Kingdom*

This presentation describes the methods and results of the first large-scale corpus-based dialect survey of regional variation American English, which is also the first grammatical dialect survey of American English and the first dialect survey of written American English. The results of this survey show for the first time that systematic patterns of regional grammatical variation exist in written American English. The basis of this survey is a 37 million word corpus of letters to the editor, published in major American newspapers between 2000-2013, representing the language of 240 cities from across the United States. In order to identify patterns of regional grammatical variation in this corpus, the values of 184 grammatical alternation variables were measured across the 240 cities in the corpus. Each of these grammatical alternations consist of two or more variant grammatical forms that have the same basic referential meaning in written American English, including alternations involving morphology (e.g. strong vs. weak verb forms), contractions (e.g. full vs. contracted *not*), function words (e.g. *among* vs. *amongst*, *although* vs. *though*, *may* vs. *might*), and word order (e.g. active vs. passive sentences). To measure an alternation variable for a particular city in the corpus, the frequency of each of its variant forms was counted in that city sub-corpus using specially designed computational linguistic software. The percentages for each of the variant were then computed for that city sub-corpus based on these frequency counts. These percentages were then mapped across the cities in the corpus and subjected to spatial (i.e. spatial autocorrelation) and multivariate (i.e. factor analysis and cluster analysis) statistical analysis to identify individual and common patterns of regional grammatical variation in the corpus. Overall this analysis identified clear patterns of regional variation in the values of many of the individual variables, as well three common patterns of regional variation that characterize the regional patterns exhibited by the majority of these variables. Specifically, three basic common patterns of regional grammatical variation were identified. The first pattern contrasts the language of Eastern United States with the language of the Western United States, with the transition zone between these two regions running roughly along the Mississippi river. The second pattern contrasts the language of the Northern United States with the language of Southern United States, with the transition zone between these two regions running roughly along the Mason-Dixon Line. The third pattern contrasts the language of the Central United States with the language of the Coasts. Finally, based on these three common patterns of regional variation, 5 dialect regions are identified: the Northeast, the Southeast, the Midwest, the Central States, and the West.

### "He told us he was going to become a girl after Christmas": Lucy Meadows, her pronouns, and the UK mainstream media

Kat Gupta

University of Nottingham, United Kingdom

In March 2013, a woman named Lucy Meadows was found dead at her home. Meadows, a primary school teacher, was transitioning from male to female. In December 2012, the school announced her decision to return to work after the Christmas break as Miss Meadows. This was reported in the local press and quickly picked up by the national press. Three months later, Meadows was found dead. Her death prompted discussions of responsible media reporting, press freedom and the contributions of trans\* people to society. Meadows' death can be contextualised in a history of press 'monstering' of transgender people. Trans Media Watch, an organisation that aims to improve media coverage of trans\* and intersex issues, made two written submissions to the Leveson inquiry (Trans Media Watch 2011, 2012) into the culture, practices and ethics of the British press. In it, they identified strategies used by the press to negatively represent trans\* people: routine use of previous names, routine use of "before" photos, demeaning and intimidating language for comic effect, and misgendering. In this paper, I focus on misgendering through pronoun use. I collected two corpora of newspaper articles: an 110,000 word corpus of 166 articles mentioning Lucy Meadows and a four million word reference corpus of 7000 general news articles. Both corpora are composed of articles published between 31<sup>st</sup> October 2012 and 31<sup>st</sup> October 2013. These two corpora are used to identify keywords – words that occur more frequently in the Lucy Meadows texts than might be expected from examining the collection of general news texts. The female pronouns *she* and *her* emerged as key; in this paper I look at these more closely using approaches drawn from corpus linguistics and critical discourse analysis (Baker 2006; Baker et al 2008). A major finding to emerge from this data is the different use of pronouns used before and after Meadows' death. *He* is used almost exclusively before Meadows' death, comprising 76% of pronoun use in reference to Meadows before March 2013. In contrast, *she* comprises 96% of pronoun use in reference to Meadows after her death in March 2013. I argue that these different usages only reflect deliberate tabloid misgendering to an extent. Instead, a more subtle form of misgendering is enacted through the selective use of quotations, which is exacerbated through the repeated use of selected quotations across different articles. This research offers an insight into media representation of trans\* people, particularly with regard to tabloid misgendering. I explore one of the patterns of pronoun use in the media representation of Lucy Meadows, and argue that press misgendering can take more subtle forms than the reporter's use of "quotation marks to dismiss the veracity of the subject's identity inappropriate pronouns or placing the person's identity in" (Trans Media Watch 2011: 11). Instead, I claim that reporters use quotations to evade direct responsibility for misgendering while continuing to produce the effect of undermining a trans\* person's gender identity.

#### References

- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, P., Gabrielatos, C., Khosravini, M., Krzyzanowski, M., McEnery, T., & 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 and Society*, 19(3), 273.
- Scott, M. (2012). *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software.
- Trans Media Watch. (2011). *The British Press and the Transgender Community: Submission to The Leveson Inquiry into the culture, practice and ethics of the press*. Retrieved from <http://www.levesoninquiry.org.uk/wp-content/uploads/2012/02/Submission-by-Trans-Media-Watch.pdf>
- Trans Media Watch. (2012). *Trans Media Watch – Additional Submission To The Leveson Inquiry – February 2012*.

## Statistical identification of keywords, lockwords and collocations as a two-step procedure

Andrew Hardie

Lancaster University, United Kingdom

The extraction of keywords from a pair of corpora under comparison is traditionally accomplished using either Log Likelihood or Chi-Squared. Both are statistics for significance testing: if the statistic is high enough, we reject the null hypothesis of no difference between the corpora. Such statistics are, however, arguably suboptimal for the purpose of keywords analysis. First, keyword lists are typically sorted by the test statistic, and analysts are (always implicitly, sometimes explicitly) encouraged to treat this as an order of importance, with the “strongest” keywords – i.e. those most distinctive of the one of the two corpora – at the top. But ranking by significance highlights those items for which there is *most evidence of a difference*, not those with the *greatest difference*. Second, an increasing interest in the study of similarity between corpora has led to the introduction of the concept of lockwords (Baker 2011), words whose frequency is similar between corpora; however, the logic of significance testing means that significance-testing statistics cannot be used to identify or to rank lockwords. Both these considerations suggest that we should move towards some form of effect size statistic for the extraction of keywords and lockwords (in keeping with a wider move towards the “new statistics”; see Cumming 2012, 2013) and indeed Gabrielatos and Marchi (2012), among other recent studies, move in this direction. The parallel case of collocation statistics is illustrative here, since, indeed, the process of extracting collocates can be conceptualised as a keyword analysis for the subcorpus made up of the set of “windows” around the node compared to the rest of the corpus. The range of statistics used for collocation has always been much wider than for keywords, and has always included effect size measures such as Mutual Information (MI) as well significance statistics (see Evert 2004). MI is a popular measure, but has the known flaw, shared by other pure measures of effect size, that it gives unwarrantedly high scores to low-frequency co-occurrences. Thus, many measures intermediate between effect size and significance have been tried out and efforts have been made to identify the optimal measure (e.g. Wiechmann 2008; Bartsch and Evert 2013). I argue that a simpler approach is to reconceptualise the procedure into two steps: first filter, then sort. For sorting items, a pure effect size statistic such as MI is to be preferred; in fact I suggest the binary log of the relative risk, which is similar to MI but is asymmetric i.e. directional, a property posited by Gries (2013) as desirable. For filtering items, either a significance test or effect-size confidence intervals should be used to identify and remove all items where we are not adequately certain that the effect exists at all. The confidence level for the filtering step should be subject to Bonferroni’s correction in order to address Kilgarriff’s (2005) concern that Log Likelihood and similar tests identify too many keywords as being statistically significant. I illustrate an implementation of this single approach for keywords, lockwords and collocations.

## References

- Baker, P. 2011. “Times may change, but we will always have *money*: Diachronic variation in recent British English”. *Journal of English Linguistics* 39 (1): 65-88.
- Bartsch, S. and Evert, S. 2013. “Exploring the Firthian notion of collocation.” In Hardie, A. and Love, R. (eds) *Corpus Linguistics 2013 Abstract Book*, pp. 31-33. Lancaster: UCREL.
- Cumming, G. 2012. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.
- Cumming, G. 2013. “The new statistics: why and how”. *Psychological Science*: online first.
- Evert, S. 2004. *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. Unpublished Ph.D. thesis. University of Stuttgart. (Published 2005; available online at <http://www.stefan-evert.de/PUB/Evert2004phd.pdf> )
- Gabrielatos, C. and Marchi, A. 2012. “Keyness: Appropriate metrics and practical issues”. Presentation at CADS 2012 conference. Available online at <http://repository.edgehill.ac.uk/4196/1/Gabrielatos%26Marchi-Keyness-CADS2012.pdf>
- Gries, S. Th. 2013. “50-something years of work on collocations: What is or should be next ...” *International Journal of Corpus Linguistics* 18(1): 137-165.
- Kilgarriff, A. 2005. “Language is never, ever, ever, random”. *Corpus Linguistics and Linguistic Theory* 1(2): 263–76.
- Wiechmann, D. 2008. “On the computation of collocation strength: Testing measures of association as expression of lexical bias”. *Corpus Linguistics and Linguistic Theory* 4(2): 253-290.

### From mother bombardment to mortar bombardment: unsupervised post-correction of OCR errors in diachronic newspaper data

Sebastian Hoffman, Andrea Sand  
Universität Trier, Germany

In recent years, institutionalised second-language varieties of English such as, for example, Indian or Jamaican English have received ample scholarly attention. However, still fairly little is known about the actual development of such “New Englishes”. Although various models have been proposed – most prominently Schneider’s (2003, 2007) evolutionary model of variety-formation – the lack of suitable corpus data has meant that so far virtually no empirical evidence for the validity of these frameworks could be provided (see Borlongan et al. 2012 for an exception). Our paper is part of a project that aims to remedy this gap by creating a diachronic corpus of a number of South-East Asian Englishes that can be located in different phases of Schneider’s model. During the course of our endeavour to compile a diachronic corpus of Singaporean English, we have been confronted with the less-than-optimal quality of the newspaper data that we have been able to obtain. Although millions of words of suitable data are available for the time-periods we aim to cover (1951 to 2011 in 10-year intervals), the fact that they were digitised via an OCR-process that clearly neither involved optimisation nor any kind of post-processing means that the data needs to be corrected before they can be integrated into the final corpus. While some level of manual correction will no doubt still be required, our current approach is to improve the quality of the data via unsupervised post-correction first. For this purpose, we are adapting and optimising a tool described in Niklas (2010), which has been provided to us by its author. In our paper, we will report on the progress made so far in improving the quality of our Singapore newspaper data. We will provide a brief overview of the internal workings of the tool and discuss possible ways to further improve its accuracy. In particular, we will show how an approach that is informed by corpus-linguistic methods (e.g. the calculation of collocations, the detection of phraseological units) can contribute to this undertaking. At the moment, our main attention focuses on the correction of non-word errors, i.e. OCR errors that result in tokens that are not contained in any of the dictionaries used for our purposes (e.g. *otllicial* instead of *official*). However, we will also evaluate the reliability of correction proposals for real-word errors of the type exemplified in the title of our paper. Once completed, our tool will be freely available to the research community and thus hopefully be of benefit beyond our own project.

### References

- Borlongan, A. M., Lim, J., Collins, P., & Yao, X. (2012). The subjunctive mood in Philippine English: A diachronic analysis. Paper presented at the 17th International Conference on English Historical Linguistics, August 20-25, 2012, Zurich, Switzerland.
- Niklas, Kai (2010) *Unsupervised Post-Correction of OCR Errors*. Unpublished diploma thesis, Hannover University, Germany. Available at [http://www.l3s.de/~tahnasebi/Diplomarbeit\\_Niklas.pdf](http://www.l3s.de/~tahnasebi/Diplomarbeit_Niklas.pdf).
- Schneider, Edgar W. (2003) The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79: 233-281.
- Schneider, Edgar W. (2007) *Postcolonial English: Varieties around the world*. Cambridge UP.

## Stylistic variation in an emerging New English. Evidence from ICE-Ghana

Magnus Huber

University of Giessen, Germany

The aim of this paper is to investigate the place of stylistic variation in Schneider's (2003, 2007) Dynamic Model of the evolution of New Englishes. In this model, it is not until the emergence of a new variety is complete that an internal diversification starts in Phase 5 "Differentiation". This last stage sees the birth of dialects in the New English and, most importantly, "marks the onset of a vigorous phase of new or increased, sociolinguistically meaningful internal diversification" (Schneider 2003: 254). Nevertheless, the Dynamic Model allows for some sociolinguistic variability as early as Phase 3. This is essentially conceived of as social class variation, with higher social classes of the indigenous population approximating the English of the (former) colonisers. Focusing on Ghanaian English (GhE, Phase 3 in Schneider's model), an Outer Circle variety spoken in Britain's former Gold Coast colony in West Africa, this paper will take a sociolinguistic-variationist approach and examine two morpho-syntactic variables: 1. auxiliary and negative contraction (e.g. *you have* ~ *you've*, *will not* ~ *won't*), and 2. noun/emphatic pronoun + non-emphatic pronoun constructions (e.g. *I'm not really dressed* ~ *me I'm* ...). The study will analyse the distribution of these variants in GhE text types of different degrees of formality and compare them to that in the historical input variety British English (BrE). The databases are the Great Britain and Ghana components of the International Corpus of English (ICE-GB, ICE-GH). The written part of ICE-GH has already been compiled and about half of the spoken part has been transcribed, providing enough tokens for a quantitative analysis. Previous research on auxiliary and negative contraction has shown that there is a clear stylistic stratification in British English, with contracted forms especially frequent in spoken and informal texts. By contrast, contracted forms are much less frequent in GhE and it will be interesting to see how they are stylistically distributed. Noun/emphatic pronoun + non-emphatic pronoun constructions are particularly frequent in conversational GhE but as good as absent in ICE-GB. In GhE, they are mostly found with 1st person singular pronouns (*me I*), but they also occur in other combinations:

- (1) *Oh, me I think they are exploiting us.* (ICE-GH, S1A-049)  
 (2) *I think it will be interesting because you guys you like football.* (ICE-GH, S1A-034)

The findings demonstrate that, in addition to the social class variation envisaged by Schneider for Phase 3, there is robust stylistic variation *within* the group of higher status users of GhE, as documented in ICE-GH. That is, there is a deeply rooted "sociolinguistically meaningful internal diversification" in the pre-Differentiation phases of a New English - considerably earlier than hitherto assumed. While the stylistic distribution of some variables parallel that in the historical input variety BrE (contraction), other variables (the pronoun sequences) are quintessentially Ghanaian and independent from BrE. Again, this is noteworthy on a theoretical level since the emergence of such local, sociolinguistic norms would not have been expected until Phase 5.

## References

- Schneider, E. (2007). *Postcolonial English. Varieties around the world*. Cambridge: Cambridge University Press.  
 Schneider, E. (2003). "The dynamics of New Englishes: From identity construction to dialect birth". *Language* 79 (2), 233–281.



### Identity representations during global sports events: Nation, race and gender in media reporting on the London Olympics in 2012 and the 2010 FIFA World Cup in South Africa

Sylvia Jaworska, Sally Hunt

Reading University, United Kingdom / Rhodes University, South Africa

In our globalised world with fractured identities, weaker nation states and mass migration, sport has become one of the social forces that maintain a sense of national belonging. As Smith & Porter (2004: 2) observe, increasingly, national identities are defined through and are “inextricably linked to what happens on the field of play”. However, recent thinking about social categories in Sociology and especially the paradigm of intersectionality (Crenshaw 1991) have demonstrated that socially constructed categories such as national identity cannot be treated in isolation, as they are always bound together with other constructs. Sport, with its historical emphasis on whiteness and masculinity, presents an interesting context in which to examine the intersection of nation, race and gender (Wheaton 2000). Our paper presents results from a project that investigates the ways in which national, racial and gendered identities are constructed in media reporting on global sports events. Sports media have been shown to play a pivotal role in constructing and mediating identities (Meân & Halone 2010: 255), which is particularly significant when most people experience such events only through the media. Thus the specific questions which the research project addresses are:

Q1: which discourses around identity are constructed and supported by media coverage during major global sports events?

Q2: which identities are foregrounded and positively valued and which are backgrounded, negatively valued or absent?

Q3: do global sports events have a major impact on the discursive ways in which national, racial or gendered identities are constructed, and, if so, what is the nature of this effect?

Our study is therefore an analysis of media reporting surrounding both the London Olympics 2012 and the 2010 FIFA World Cup in South Africa. A comparison of media representations concerning two international events allows for generalisations to be made about patterns of how identities are constructed discursively, as opposed to a study of a single event in a single geographical location. Our methodology combines the tools of Critical Discourse Analysis and Corpus Linguistics (Baker 2008, Baker et al. 2008). Answers to Q1 and Q2 will be offered by an analysis of two corpora that comprise all articles published in 6 major British national newspapers during the London Olympics 2012 and all articles from 8 major national South African newspapers from the time of the 2010 FIFA World Cup. This totals 24,535 articles. To test the impact of these two global sport events on the construction of identities (Q3), data from two control corpora encompassing articles from the same publications with the same time span, but one year before and one year after the global event have been collected for comparison. By scrutinising frequency lists and studying collocational profiles of lexical items denoting national/ethnic/regional identities in all six data sets, we identify patterns in terms of the representation of identity and draw links between the global sport events, textual tendencies and the discourses which emerge from the newspaper data. Our study also offers new methodological insights of relevance to comparative discourse analysis across time and cultures

#### References

- Baker P. (2008) *Using Corpora in Discourse Analysis*. London: Continuum.
- 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.
- Crenshaw, K. (1991) Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. In: *Stanford Law Review*, 43 (6): 1241–1299.
- Meân, L. J & Halone, K. K. (2010) Sport, Language and Culture: Issues and Intersections. In *Journal of Social Psychology*, 29 (3): 253–260.
- Smith, A. & Porter, D. (2004) *Sport and National Identity in the Post-War World*. Routledge, London & New York.
- Wheaton, B. (2000) 'New Lads'?: Masculinities and the "New Sport" Participant. In: *Men and Masculinities*, 2 (4): 434–456.

### Exploring Object Control in English: on a Class of Exceptions to Bach's Generalization

Mark Kaunisto, Juhani Rudanko  
University of Tampere, Finland

Consider sentences (1a-d), from Rizzi (1986, 503):

- (1)
- a. This leads people to the following conclusion.
  - b. This leads to the following conclusion.
  - c. This leads people to conclude what follows.
  - d. This leads to conclude what follows.

Sentence (1a) illustrates object control: the matrix predicate *lead* selects three arguments, and the direct object argument controls the reference of the understood subject. As regards the ill-formedness of (1d), it is accounted for by Bach's Generalization. Rizzi formulates it as follows: "in object control structures the object NP must be structurally represented" (Rizzi 1986, 503). The availability of large electronic corpora affords new opportunities to investigate exceptions to Bach's Generalization. For this study the Corpus of Historical American English, COHA, has been selected for investigation, because of its size and variety of text types. The study examines the verb *advise* from the perspective of Bach's Generalization. The reason for this choice is that besides object control constructions that observe the Generalization, as in (2a), it also permits constructions that violate the Generalization, as in (2b):

- (2)
- a. ... friends advised her against marrying him (1955, COHA, FIC)
  - b. ... military supply experts for the allied powers advised against bringing Russia into the war ... (1955, COHA, News)

The paper tracks the incidence of "regular" object control constructions of the type of (2a), which are consistent with Bach's Generalization, in COHA, and then focuses on violations of the Generalization, as in (2b). Besides recording the incidence of such violations in relation to the regular pattern during the entire time span covered by COHA, the paper seeks to characterize the nature of such violations, in order to test whether the emphasis placed in earlier work on general or even generic interpretations is justified in the light of authentic data, and in order to compare the findings based on *advise* to other recent work on object control. The paper also uses corpus evidence to identify the factor or factors that may have promoted the use of the pattern.

### References

Rizzi, Luigi. 1986. Null Objects in Italian and the Theory of *pro*. *Linguistic Inquiry* 17, 501-557.

**Comment is free? A corpus-based analysis of reader comments on *The Guardian* website**

Andrew Kehoe, Matt Gee  
Birmingham City University, United Kingdom

Launched in March 2006, 'Comment is Free' is a section on the website of the UK newspaper *The Guardian* where non-journalists can, by invitation, write a blog post on any subject of their choosing. Readers are encouraged to comment on these blog posts and take part in discussions, with some posts generating over 1000 comments. A fortnight after the launch of Comment is Free, *The Guardian* began to allow reader comments on conventional news articles across all sections of its website. Hermida & Thurman (2008: 6) report that five other UK newspaper websites were allowing reader comments on news articles by the end of 2006. The integration of blogs and reader comments – so-called 'user-generated content' – across such websites has led to a blurring of the boundaries between opinion and hard news, and between professional and non-professional writing. Indeed the blog posts in the Comment is Free section of *The Guardian* look almost identical to the conventional news articles elsewhere on its website. This paper presents a corpus-based analysis of comments across *The Guardian* website since their introduction in 2006. It is based on a wider corpus of articles from *The Guardian* covering the period 2000 to present, searchable through the WebCorp Linguist's Search Engine. We describe how this pre-existing corpus has been augmented with reader comments on over 500,000 articles and blog posts published since March 2006. Previous studies of the WebCorpLSE *Guardian* corpus (e.g. Kehoe & Gee 2009, Renouf 2013) have not differentiated between blog posts in Comment is Free and professionally-authored news articles elsewhere on the *Guardian* website. In the first part of this paper, we adopt a keywords-based approach to explore the differences between the two sections (distinguished by URL) and whether or not these differences are becoming less pronounced over time. In the second part of the paper, we explore the distribution of reader comments across blog posts and articles (henceforth referred to collectively as 'articles'). Our initial analysis has suggested that comments are permitted on around 40% of articles and, where commenting is permitted, the vast majority of articles (85%) have at least one comment. *The Guardian's* commenting policy is rather vague, stating only that comments are not allowed on 'stories about particularly divisive or emotional issues'. In this paper, we are able to identify sub-sections of the newspaper's website where commenting is most prevalent and where it is most likely to be banned outright. Taking the analysis further, through the extraction of keywords we identify the specific topics which are most likely to generate debate, often relating to politics, religion and social issues. Moreover, we are able to identify keywords indicative of particular styles of writing which encourage the most reader discussion. Examples presented in detail include the keywords *simply*, *far*, *even* and *seem(s)*. We also present our most recent work analysing the corpus at author level. We look both at whether there are particular authors whose articles generate the most comment and at whether the more prolific commenters tend to focus their responses on specific topics or comment across the *Guardian* website. Overall, our work offers insights in to changing newspaper practices and reader behaviour through lexical analyses of a large corpus of articles and comments. As we illustrate, with the continued growth of user-generated content, our work has potential benefits in the refinement of automated spam detection and moderation procedures.

**References**

- Hermida, A. & N. Thurman (2008) 'A clash of cultures: the integration of user-generated content within professional journalistic frameworks at British newspaper websites'. *Journalism Practice* 2(3), 343-356.
- Kehoe, A. & M. Gee (2009) 'Weaving Web data into a diachronic corpus patchwork' in A. Renouf & A. Kehoe (eds.) *Corpus Linguistics: Refinements & Reassessments*, Amsterdam: Rodopi, pp. 255-279.
- Renouf, A. (2013) 'A Finer Definition of Neology in English: the life-cycle of a word', in H. Hasselgård, S. Oksefjell Ebeling & J. Ebeling (eds.). *Corpus Perspectives on Patterns of Lexis*. Amsterdam: John Benjamins, 177-208.

## The Progressive and Durativity in Irish Standard English

John Kirk

Independent Researcher, Northern Ireland

The syntax and semantics of the Progressive in English have been well studied in descriptive reference grammars and in recent monographs (e.g. Kranich 2010 on their historical development, and Römer 2005 on their didactic implications). Recent studies have shown the construction to be expanding in frequency generally in British English (e.g. Mair and Hundt 1995 and numerous other studies culminating in Leech et al. 2009), and in its acquisition of pragmatic functions (cf. Smith 2005). However, corpus-based studies specifically of Irish English (e.g. Filppula 2001, 2003, 2004, 2012b) have attested that the frequency of the Progressive in Irish English is yet higher. Similar claims are made for Scottish English (e.g. Kirk 1986, Miller e.g. 2004, etc.) In each case, remarkably, the frequency is attributed to use (or greater use) with private, cognitive-state verbs. The data for this paper comes from the spoken component of the ICE-Ireland Corpus (Kirk et al. 2011, cf also Kallen & Kirk 2008). A first and prima facie research question, using well-tried corpus-linguistic quantitative techniques, is how far the fairly standardised language represented in the ICE-Ireland Corpus confirms those findings about the Progressive: it does. However, a more fine-grained analysis involving particular verbs, harmonic adverbials, meanings and pragmatic functions shows a more complicated situation. In Irish English, the tense-aspect system is not merely identical with exocentrically-imported standard English but combines elements which have transferred endocentrically from the Irish language. A second research question thus seeks a fresh adjudication of the patterns of variation within the Progressive construction in the data with regard to these two superstratal and substratal determinants. Although the high frequencies of the standard Progressive in as attested in the ICE-Ireland Corpus may be ascribed to seemingly new linguistic contexts (such as the occurrence of progressive with infinitives after modals (cf. Kirk et al. 2008, Filppula 2012a), the analysis shows that, from earlier contact with Irish (cf. Henry 1957), the Progressive also expresses durative meanings. This substratally-derived expression of durativity through stative verbs goes some way towards explaining the high frequency of progressive statives in Irish English – and so the higher frequency when compared with ICE-GB or other data sets. To further the main theme of the conference, substratal influence from Irish provides a cultural context for a re-interpretation of the Progressive.

### References

- Filppula, Markku. 2001. 'Irish Influence in Hiberno-English: Some Problems of Argumentation'. In *Language Links: The Languages of Scotland and Ireland*. Eds. John M. Kirk and Dónall P. Ó Baoill. Belfast: Cló Ollscoil na Banríona. 23–42.
- Filppula, Markku. 2003. 'More on the English Progressive and the Celtic Connections'. In *The Celtic Englishes III*. Ed. Hildegard L.C. Tristram. Heidelberg: Carl Winter Universitätsverlag. 150–168.
- Filppula, Markku. 2004. 'Irish English'. In *A Handbook of Varieties of English, Volume 2: Morphology and Syntax*. Eds. Kortmann, Bernd, Kate Burridge, Rajend Mesthrie, Edgar W. Schneider, and Clive Upton. Berlin: Mouton de Gruyter. 73–101.
- Filppula, Markku. 2012a. Exploring Grammatical Differences between Irish and British English. In *New Perspectives on Irish English*. Eds. Bettina Migge, Bettina and Máire Ní Chiosáin. Amsterdam: Benjamins, 85–99.
- Filppula, Markku. 2012b. 'Irish English'. In *The Mouton World Atlas of Variation in English*. Eds. Bernd Kortmann and Kerstin Lunkenheimer. Berlin: De Gruyter Mouton. 30–47.
- Henry, P.L. 1957. *An Anglo-Irish Dialect of North Roscommon*. Dublin: Department of English, University College Dublin.
- Kallen, Jeffrey L. and John M. Kirk. 2008. *ICE-Ireland: A User's Guide*. Belfast: Cló Ollscoil na Banríona.
- Kirk, John M. 1986. *Aspects of the Grammar in a Corpus of Dramatic Texts in Scots*. Unpublished PhD Thesis, University of Sheffield.
- Kirk, John M., Jeffrey L. Kallen and Markku Filppula, 2008. 'The Progressive in British and Irish Standard English'. Paper presented at *13th International Conference on Methods in Dialectology (Methods XIII)*, Leeds, 4–8 August 2008.
- Kirk, J. M., J. L. Kallen, O. Lowry, A. Rooney. and M. Mannion. 2011. *International Corpus of English: Ireland Component*. The ICE-Ireland Corpus: Version 1.2.2. Belfast: Queen's University Belfast and Dublin: Trinity College Dublin.
- Kranich, Svenja. 2010. *The Progressive in Modern English: A Corpus-based Study of Grammaticalisation and Related Changes*. Amsterdam: Rodopi.
- Mair, Christian and Marianne Hundt. 1995. 'Why is the Progressive Becoming More Frequent in English?' *Zeitschrift für Anglistik und Amerikanistik*. 43: 111–122.
- Miller, Jim. 2004. 'Scottish English: Morphology and Syntax'. In *A Handbook of Varieties of English, Volume 2: Morphology and Syntax*. Eds. Kortmann, Bernd, Kate Burridge, Rajend Mesthrie, Edgar W. Schneider, and Clive Upton. Berlin: Mouton de Gruyter. 47–72.
- Römer, Ute. 2005. *Progressives, Patterns, Pedagogy: A Corpus-driven Approach to English Progressive Forms, functions, Contexts and Didactics*. Amsterdam: John Benjamins.
- Smith, Nicholas. 2005. *A Corpus-based Investigation of Recent Change in the Use of the Progressive in British English*. Unpublished PhD Thesis, Lancaster University.

### “The people on the island ~~sta ste~~ steal all the fish“: What we can learn from deletions in authentic learner texts?

Rolf Kreyer

University of Marburg, Germany

About two decades ago, corpus linguists began to become increasingly interested in the study of learner language with corpus-linguistic methods. This has led to the compilation of a fairly large number of learner corpora with learners from different L1s, such as ICLE (Dagneaux et al. 1998), LINDSEI (Granger 1997) or CALE (Callies & Zaytseva 2011). What these (and other) learner corpora have in common is that they represent learner language of a high intermediate to advanced level, usually from learners in university settings; learner language at a truly intermediate level has remained underrepresented (at least for German learners of English). The Marburg corpus of Intermediate Learner English (MILE), a corpus currently being compiled at the University of Marburg, wants to fill this gap. It aims to create a collection of written learner English from grades 9 to 12 of the German *Gymnasium*. The corpus is designed as a longitudinal database and will document the progress of learners in their final years of school. So far, a pilot version of the corpus exists, which contains 2.5 years of authentic material, i.e. 841 written exams, totalling roughly 340,000 words. The digitalised texts in this corpus try to represent the hand-written versions as closely as possible, e.g. by marking line breaks (lb), deletions (d) or additions (a), as shown in example (1).

- (1) The inaugural speech <d>of</d> <ad>, written by</ad> Barack Obama, <lb></lb> shows us that America is at war and in <lb></lb> the midst of crisis.

The advantage of this kind of in-text markup lies in the fact that alterations made during the writing process can be identified by the computer and, thus, can be made an object of linguistic enquiry and analysis. Such alterations are not viewed as mere mistakes or performance errors but are regarded as an additional window onto the development of L2 competence. Consider the examples below:

- (2) He <d>will</d> wants, that his pupil do things on their own and <d>right</d> write poetry.  
(3) The murderer is just more violent then the author, but <d>both</d> <d>bouth</d> both have the same aim.  
(4) [...] the solar system shines more brightly today, <d>which</d> <a>what</a> results a warmer Earth.

Example (2), for instance, makes apparent that the learner has problems with interference from German and with homonyms, a fact that does not show in the final version of the sentence. Similarly, in example (3) it is only through the deletions that we become aware that the student struggles with orthography even though s/he decides for the correct version in the end. Example (4), in contrast, shows that the learner knows more about sentential relative clauses than the final incorrect version tells. The paper will analyse more than 500 instances of deletion in the MILE-corpus, providing new and sometimes surprising insights into the interlanguage of learners at an intermediate level of proficiency.

## References

- Callies, M. & E. Zaytseva. 2011. The Corpus of Academic Learner English (CALE): A new resource for the study of lexico-grammatical variation in advanced learner varieties. In Hedeland, H., Th. Schmidt & K. Wörner (eds.), *Multilingual Resources and Multilingual Applications*, 51-56. Hamburg: University of Hamburg.
- Dagneaux, E., S. Denness & S. Granger. 1998. The International Corpus of Learner English. Handbook and CDROM. *System: An International Journal of Educational Technology and Applied Linguistics* 26. 163-174.
- Granger, S. 1997. *The Louvain International Database of Spoken English Interlanguage (LINDSEI) Project*. Centre for English Corpus Linguistics.

## Complementing stance verbs in Early Modern English: The role of zero complement clauses

Daniela Landert

University of Zurich, Switzerland

Epistemic and evidential stance – i.e. the degree of certainty and source of information – can be expressed in various ways. One of them is through complement clauses headed by lexical verbs, such as *think*, *believe*, *hear*, and *seem* (Biber et al. 1999: 972–974). These verbs allow speakers or authors to modify the propositional content of their statements by expressing how they acquired the information and/or how reliable they judge it to be. For Present-day English, complement clauses have been found to be the second-most frequent way of expressing stance after (semi-)modal verbs (Biber et al. 1999: 979), and within the category of complements, verbs are most frequently used as heads (Biber et al. 1999: 984). This makes verb + complement clause an important pattern of expressing stance. Individual stance verbs have been studied from a diachronic perspective in great detail, such as *think* (see, for instance, Aijmer 1997; Bromhead 2009; Palander-Collin 1999) and *seem* (see, for instance, Aijmer 2009). However, the overall picture of how stance is expressed in earlier periods of English is still sketchy. One area that deserves further attention is the study of zero complement clauses. So far, quantitative approaches to stance that cover the early modern period have focused on complementation patterns that can be identified more easily, such as *that* and *to* complement clauses (Biber 2004; Gray, Biber and Hiltunen 2011). This line of research has identified register variation and diachronic changes in stance marking, but the question remains how the inclusion of zero complement clauses might lead to new insights, especially since several stance verbs appear to have a strong preference for zero complements – up to 87.8% in a study of Late Modern English correspondence (Fitzmaurice 2003:123). The present study makes a contribution at filling this gap. Using the Penn-Helsinki Parsed Corpus of Early Modern English, I investigate the complementation patterns of verbs that express epistemic and evidential stance, focussing in particular on the role of zero complement clauses. Apart from the overall quantitative distribution of complementation patterns, two aspects will be investigated in detail. On the one hand, I will look at the variation of complementation patterns across registers in order to identify contexts in which stance verbs are frequently followed by zero complement clauses. On the other hand, I will study whether differences in the degree of certainty and the source of information are systematically related to different complementation patterns. These findings will lead to a more comprehensive understanding of the expression of stance in Early Modern English, while also providing important methodological insights for further studies of stance marking.

## References

- Aijmer, Karin. 1997. *I think* – an English modal particle. In Toril Swan & Olaf Jansen Westvik (eds.), *Modality in Germanic Languages: Historical and Comparative Perspectives*, 1–47. Berlin: Mouton de Gruyter.
- Aijmer, Karin. 2009. *Seem* and evidentiality. *Functions of Language* 16(1). 63–88.
- Biber, Douglas. 2004. Historical patterns for the grammatical marking of stance: A cross-register comparison. *Journal of Historical Pragmatics* 5(1). 107–136.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Bromhead, Helen. 2009. *The Reign of Truth and Faith. Epistemic Expressions in 16th and 17th Century English*. (Topics in English Linguistics 62). Berlin: Mouton de Gruyter.
- Fitzmaurice, Susan. 2003. The grammar of stance in early eighteenth-century English epistolary language. In Pepi Leistyna & Charles F. Meyer (eds.), *Corpus Analysis. Language Structure and Language Use*, 107–131. Amsterdam/New York: Rodopi.
- Gray, Bethany, Douglas Biber & Turo Hiltunen. 2011. The expression of stance in early (1665–1712) publications of the Philosophical Transactions and other contemporary medical prose: innovations in a pioneering discourse. In Irma Taavitsainen & Päivi Pahta (eds.), *Medical Writing in Early Modern English*, 221–257. (Studies in English Language). Cambridge: Cambridge University Press.
- Palander-Collin, Minna. 1999. *Grammaticalization and Social Embedding. I think and methinks in Middle and Early Modern English*. Helsinki: Société Néophilologique.

Introductory *it* patterns in native-speaker and non-native-speaker student writing: A syntactic analysis

Tove Larrson

Uppsala University, Sweden

While the objective presentation of findings remains an integral part of academic writing, successfully arguing for claims in an engaging yet authoritative way is also of great importance. The aim of the present study is to investigate how one linguistic device which is commonly used for this purpose, namely the introductory *it* pattern, is used in academic writing by native-speaker and non-native-speaker students. This pattern consists of an anticipatory pronoun *it* in subject position, a predicate and an extraposed clausal subject (cf. Quirk et al., 1985:1391), as in *It is noteworthy that the government only gave him two options*. Patterns with an extraposed clausal subject have been reported to be especially frequent in academic discourse (e.g. Biber et al., 1999:723), which suggests that the introductory *it* pattern is an important pattern for university students to learn. Studies have shown that the pattern is problematic for learners to master (e.g. Hewings & Hewings, 2002); however, little is known about what types of introductory *it* patterns present the greatest challenge. This paper will fill this gap by discussing which types of introductory *it* patterns are particularly difficult for learners, with specific focus on how the use of the pattern varies with academic proficiency. This is achieved through a contrastive, frequency-based study of how the pattern is used in native-speaker and non-native-speaker student writing. The study takes as its starting point the seven different syntactic types of introductory *it* patterns outlined by Quirk et al. (1985:1392) (repeated as 1-7 below).

- (1) SVC: It is a pleasure *to teach her*.
- (2) SVA: It was on the news *that income tax is to be lowered*.
- (3) SV: It doesn't matter *what you do*.
- (4) SVO: It surprised me *to hear him say that*.
- (5) SVOC: It makes her happy *to see others enjoying themselves*.
- (6) SV<sub>pass</sub>: It is said *that she slipped arsenic into his tea*.
- (7) SV<sub>pass</sub>C: It was considered impossible *for anyone to escape*.

The learner material is culled from the Advanced Learner English Corpus (ALEC), which is a newly compiled 1-million-word corpus of texts written in English by Swedish university students. Unlike previous learner corpora, ALEC allows for investigation of a potential correlation between, on one hand, the use of the pattern and, on the other hand, factors such as student level and the grade that the texts were awarded, which will shed light on the potential influence of academic proficiency. The learner data is compared to native-speaker data from the British Academic Written English Corpus (BAWE). Preliminary results indicate that the groups differ in their use of the pattern and that even the most highly proficient learners tend to underuse the passive patterns. The findings of the complete study will not only contribute to a deeper understanding of the use of the introductory *it* pattern in academic discourse, but also prove useful to pedagogy in English for Academic Purposes.

## References

- Advanced Learner English Corpus (ALEC). Corpus compiled by the author in 2013.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.
- British Academic Written English Corpus (BAWE). Corpus compiled by the Universities of Warwick, Reading and Oxford Brookes in 2004–2007. <http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/>
- Hewings, M., & Hewings, A. (2002). "It is interesting to note that...": A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes*, 21 (4), 367–383.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

The rise of *cos* and the downfall of *for*: Exchanging reason connectives in recent English?

Geoffrey Leech

Lancaster University, United Kingdom

Arguably the two most notable developments in the use of English reason connectives over the past 100 years have been the 'terminal' decline in frequency of the reason connective *for* and, oppositely, the consistent and fairly steep rise in frequency of the reason connective *(be)cause*. (The term 'connective' here refers to what Huddleston and Pullum (2002) call a preposition with a clausal complement, and what more traditional grammars have called subordinating conjunctions.) The key comparison in this paper, therefore, is between the usage of *for* and *(be)cause* as illustrated in the following sentences:

- (1) I have said that I will offer no further suggestions, *for* advice of mine is suspect (LOB: K13)  
[DECLINING FREQUENCY]
- (2) They'd be second-hand shoes, *because* they've been worn for three days (BNC spoken: KBL)  
[INCREASING FREQUENCY]

The question addressed by this paper is: How far are these two frequency trends interlinked? Has the English-speaking world been exchanging one connective for the other? *(Be)cause* and *for* are remarkably similar in their syntactic and discursal status. Since its entry into the language in ME, *(be)cause* has followed a well-understood process of grammaticalization, which in the past century has accelerated especially in the spoken language. Phonetically *because* is increasingly reduced to a single syllable with a reduced vowel (/kəz/ - often informally spelt 'cos in present-day BrE and 'cause in present-day AmE transcriptions). Its increase in frequency has also been strikingly consistent in the past 100 years, with the spoken language taking the lead. In contrast, *for* (as conjunction) has declined virtually to extinction in the spoken language, as well as undergoing drastic reduction in the written language. Yet, in other respects, *for* shows an uncanny similarity to *(be)cause*. Both connectives, unlike the other common reason connectives *since* and *as*, occur almost entirely in the post-matrix clause position. (As far as *(be)cause* is concerned, the word 'almost' can be deleted in the preceding sentence.) Thus if the positions of the clauses in (1) and (2) above are reversed, the resulting sentences are either unacceptable or highly infrequent:

- (1a) \**For* advice of mine is suspect, I have said that I will offer no further suggestions.  
(2a) ?*(Be)cause* they've been worn for three days, they'd be secondhand shoes.

Discursively, there is a tendency for the reason clause to become more detached from the matrix clause, beginning a new sentence or information-unit with *For...* or with *Be(cause)...* Despite these close similarities, the 'exchange' hypothesis that *for* is being phased out and replaced by *(be)cause* is a gross simplification. The two connectives are stylistically poles apart, and semantically far from interchangeable. In accounting for these changes, attention has to be given other explanations, involving not only grammaticalization, but colloquialization (Hundt and Mair 1998), Americanization (Leech et al. 2009: 253-263) and pragmaticalization (Evers-Kermeul et al. 2011). The chief corpus resources for this diachronic study are the BNC, the Brown Family of Corpora (1901-2006/7), COHA and COCA.

## References

- Evers-Kermeul, J., L. Degand, B. Fagard, and L. Mortier. 2011. Historical and comparative perspectives on subjectification: A corpus-based analysis of Dutch and French causal connectives, *Linguistics* 49: 2, 445-480.
- Huddleston, R. and G. Pullum. 2002. *Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hundt, M. and C. Mair. 1998. 'Agile' and 'uptight' genres: the corpus-based approach to historical change in progress. *IJCL* 4, 221-242.
- Leech, G., M. Hundt, C. Mair and N. Smith. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.



**He's always telling some kind of lie: Subjective progressives in the history of American English.**

Magnus Levin

Linnaeus University, Sweden

Subjective progressives of the type *You young people are always interrupting* have been covered both in synchronic and diachronic studies (e.g., Römer 2005; Smitherberg 2005; Kranich 2010). However, since these studies are based on small corpora, their results are partly inconclusive and there is thus a need to use modern mega-corpora to explore these low-frequency items. This study therefore investigates the use of progressives following *always/constantly/forever* in Mark Davies' 400-million-word Corpus of Historical American English (COHA). COHA makes it possible to carry out large-scale diachronic investigations of rare phenomena. This study covers 3,000 tokens from the 1810s, 1850s, 1900s, 1950s and 2000s from COHA from fiction, newspapers, magazines and non-fiction. The aim is to explore the factors affecting synchronic and diachronic variation with these structures and whether instances expressing (negative) subjective attitudes are changing and what factors correlate with these attitudes. The findings shed light on, for instance, colloquialization, the concept of subjectification and the influence of gender on language change. Subjective progressives typically involve the expression of negative subjective speaker attitudes and involve co-occurring *always*-type adverbials. Traugott & Dasher (2002) argue that meanings of grammatical constructions typically become increasingly subjective. This idea is supported by Kranich's (2010) data for *always* (etc.) + progressive from ARCHER, which indicate that the connection to negative connotations is mainly a 20th-century phenomenon. In contrast to these suggestions, the present investigation indicates no consistent shifts in subjectivity over time in the sense of increased expression of negative speaker attitudes. Instead, the proportion of negative connotations remains fairly stable over time. Overall, the construction more than doubles in frequency over 200 years, but this is solely due to the increase of *always* + progressive; *constantly* and *forever* + progressive either remain stable or decrease in frequency. This suggests that there is ongoing specialization where an increasingly large proportion of all subjective progressives include the adverbial *always*, rather than subjective meanings becoming generalized to be used with more adverbials. *Always/constantly/forever* + progressives are the most frequent in fiction, and this is also the genre with the highest proportions of negative subjective attitudes. This is probably due to fiction to a large part consisting of dialogue, since the progressive has been found to be most frequent in informal, speech-based genres (Leech *et al.* 2009: 122–127). The progressives in the present study often co-occur with colloquial features such as first- and second-person pronoun subjects and contracted verb forms. In the fiction sub-corpus, women authors consistently use these progressives more often than men do. Thus, women are leading the way in the increase of the progressive as previously found by Arnaud (1998) and Smitherberg (2005), and trend which is in line with Labov's (2001: 292–3) claim that women use more of the innovative forms in language change from below.

**References**

- Arnaud, René. 1998. The development of the progressive in 19th century English: A quantitative study. *Language Variation and Change*, 10, 123–152.
- Kranich, Svenja. 2010. *The progressive in Modern English. A corpus-based study of grammaticalization and related changes*. Amsterdam: Rodopi.
- Labov, William. 2001. *Principles of linguistic change. Vol II: Social factors*. Malden: Blackwell.
- Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English. A grammatical study*. Cambridge: Cambridge University Press.
- Römer, Ute. 2005. *Progressives, patterns, pedagogy: a corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam: Benjamins.
- Smitherberg, Erik. 2005. *The progressive in 19th-century English. A process of integration*. Amsterdam: Rodopi.
- Traugott, Elizabeth Closs & Dasher, Richard B. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

On the adverbialization of *may* + *be*'happen' combinations

María José López-Couso, Belén Ménde-Naya  
University of Santiago de Compostela, Spain

Pragmatic markers have become an extremely popular area of research over the last few years both from a synchronic and a diachronic perspective. This interest is in part related to the boom of grammaticalization studies, which have provided a suitable framework to adequately describe the emergence of such forms (e.g. Brinton 2008). A common developmental path for pragmatic markers is found in biclausal complementation structures in which an originally complement-taking-predicate clause is downgraded to a parenthetical. This is the case of widely studied epistemic parentheticals like *I think*, *I guess*, and *I gather*, which show a first person pronoun subject (see, e.g., Thompson & Mulac 1991; Boye & Harder 2007; Kearns 2007; Brinton 2008: chapter 10). By contrast, impersonal parenthetical clauses with a third person singular subject (e.g. *it seems*) have not attracted so much scholarly attention. Interestingly, some of these third person parentheticals have gone a step further in their development by losing the morphosyntactic features which characterized them as clauses, thus eventually becoming adverbs. An example is Middle English *methinks* (see López-Couso 1996; Palander-Collin 1996; Wischer 2000). Further examples of adverbialization are found in the development of epistemic adverbs such as *maybe* and *mayhapp(en)* which, according to the OED (s.vv.), arise from the reanalysis of a clause featuring the modal *may* followed by the verb *be* or a verb meaning 'happen'. The aim of this paper is to trace the origin and development of *maybe*, *mayhapp(en)* and related forms (e.g. *may-fall*, *may chance*, *may-fortune*, *may-tide*; OED s.v. *may* v<sup>1</sup>), which have so far eluded systematic investigation (but cf. Boye & Harder 2007 and Beijering 2010 for languages other than English). In order to provide a comprehensive view of these developments, in addition to the standard historical dictionaries, data are drawn from various corpora. Given that *maybe* and other stance adverbs seem to be particularly common in conversation in Contemporary English (see Biber et al. 1999: 562, 869), evidence from the multi-genre historical corpora *Helsinki Corpus* and *ARCHER* is complemented by that from corpora which contain texts with a high degree of speechlikeness, among them the *Corpus of Early English Correspondence*, the *Corpus of English Dialogues*, and the *Old Bailey Corpus*. The analysis of all the occurrences of third person singular *may* + *be*'happen' in the selected corpora enables us to identify the different stages in the adverbialization process of *may*-combinations. In our presentation, we show how they come over time to show criterial features of grammaticalization, like decategorialization, fusion, semantic bleaching, and pragmatic strengthening.

## References

- Beijering, Karin. 2010. "The grammaticalization of Mainland Scandinavian MAYBE." *Bergen Language and Linguistics Studies* 1.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Boye, Kasper, and Peter Harder. 2007. "Complement-taking Predicates: Usage and Linguistic Structure." *Studies in Language* 31 (3): 569–606.
- Brinton, Laurel J. 2008. *The Comment Clause in English: Syntactic Origins and Pragmatic Development*. Cambridge: Cambridge University Press.
- Kearns, Kate. 2007. "Epistemic Verbs and Zero Complementizer." *English Language and Linguistics* 11 (3): 475–505.
- López-Couso, María José. 1996. "On the History of *methinks*: From Impersonal Construction to Fossilized Expression." *Folia Linguistica Historica* 17: 153–169.
- OED = *Oxford English Dictionary Online*. Oxford University Press. <<http://www.oed.com>>
- Palander-Collin, Minna. 1999. *Grammaticalization and Social Embedding: I THINK and METHINKS in Middle and Early Modern English*. Helsinki: Société Néophilologique.
- Thompson, Sandra A., and Anthony Mulac. 1991. "A Quantitative Perspective on the Grammaticization of Epistemic Parentheticals in English." In *Approaches to Grammaticalization*, vol. 2, ed. by Elizabeth Closs Traugott, and Bernd Heine, 313–329. Amsterdam/Philadelphia, PA: John Benjamins.
- Wischer, Ilse. 2000. "Grammaticalization versus Lexicalization: 'Methinks' There Is Some Confusion." In *Pathways of Change: Grammaticalization in English*, ed. by Olga Fischer, Anette Rosenbach, and Dieter Stein, 355–370. Amsterdam/Philadelphia, PA: John Benjamins.

## Whither the Monitor Corpus?

Bill Louw, Manija Milojkovic

University of Zimbabwe, Zimbabwe / University of Belgrade, Serbia

Were 'Monitor Corpora' always an endangered species? There was no initial doubt as to their usefulness or the need to have them. Time has moved on, and most corpora are now dead letters. But AVIATOR, still in the hands of Antoinette Renouf and her team, continues to live and breathe. It is only one of many phenomena that show that we are increasingly *trusting the text* as John Sinclair suggested. And there are now other phenomena in play that have made monitor corpora look increasingly dispensable. This paper will examine the extent to which *intuitive opacity* is almost daily beginning to render any purported 'human' monitoring untrustworthy at best or risible at worst. How can an author guarantee that his/her sincerity will not be rumbled by automations that are based upon 22 years of developments in semantic prosody? And matters are looking worse and worse for monitor corpora as the cognitive stands falsified by corpus-derived subtext - the most frequent lexis used in or around any chosen grammatical string (the most frequent lexical variables are termed 'quasi-propositional variables' [QPVs], if reasoned from Russell (1972: 24), because the variables in natural language are not the same as those in logic). COBUILD may need to be revived to write a Dictionary of Subtext as other linguistic theories fall like dominoes. For example, this is the first stanza of a poem by Larkin:

*The trees are coming into leaf  
Like something almost being said;  
The recent buds relax and spread,  
Their greenness is a kind of grief.*

There is nothing in line 1 that predicts the human element which abounds in the poem, starting from line 2. However, in the BNC, the search line 'the \*s are \*ing into' yielded the following QPVs:

1 steal the shuttle we've prepared for them. All the pieces are falling into place.' ' Oh, ' Britta 2 probably is. I can see what he is thinking: the pieces are falling into place. He doesn't know  
3 letely mental. So on gaining the top, where the walkers are tucking into their second pork pie, 4 it's January 1940 and it's all much worse. The Russians are marching into Finland.' The Finns  
5 rs this question instantly:' There's no way the kids are going into racing.' The good thing is 6 risk is from the weather and the conditions the roads are getting into. Appalling conditions ar  
7 hen we learn at the start of chapter 4 that the Israelites are going into battle with the might  
8 ndow and tell it's 2C outside. How? Because the crocuses are coming into bloom. Crocuses are pl  
9 ting involves yet another skin change, when the caterpillars are retreating into their chrysalis

In the first wildcarded slot, the human element (walkers, Russians, kids and Israelites) predominates over nature (crocuses and caterpillars). The 'pieces' in the first two lines are delexical and refer to human understanding. The first seven contexts describe various degrees of conflict – even in line 6 the context is Bosnia in the ninety-nineties: 'Troops have been there some time and they're familiar with locals on the ground. Really the risk is from the weather and the conditions the roads are getting into. Appalling conditions are likely to prevent food aid reaching the remotest parts.' This paper will demonstrate how applications such as corpus stylistics can have perfectly reliable findings by regarding vocabulary merely as proxies for the arguments behind fictional worlds. The computer can no longer be easily gainsaid. And as for the monitors: *Quis custodiet custodes ipsos?* [Who will guard the guardians themselves?]

## References

Russell B. (1972). In Pears, D. F. (Ed.) *Bertrand Russell: A Collection of Critical Essays*. Anchor Books: New York.

## Indefinite and bare nominal gerunds in PD English: towards a functional account

Charlotte Maekelberghe, Liesbet Heyvaert  
K.U. Leuven, Belgium

This paper discusses two types of nominal gerunds in Present-day English, viz. the indefinite nominal gerund (iNG) (1), and the bare nominal gerund (bNG) (2):

- (1) With a tightening of her stomach Caroline watched the girl approach. (COCA)
- (2) (...) we find sections on (...) reading of scripture and preaching of the word. (BNC)

Bare NGs and (especially) indefinite NGs have been largely neglected in the literature, yet their occurrence is interesting in several respects:

- (i) NGs are generally regarded as mass nouns (Brinton 1998; Langacker 1991), which do not normally take indefinite articles and opt for zero determination to express indefinite reference (Langacker 2004);
- (ii) When they are mentioned in the literature, indefinite NGs have tended to be associated with fully lexicalized (typically count) nouns (e.g. *a building/meeting/painting*; Jespersen 1914-1929, vol. 4; Wik 1973);
- (iii) Diachronic research on NGs suggests that iNGs and bNGs have undergone quite some changes since Middle English, iNGs having gained in frequency while bNGs have become less frequent and seem to have been partly replaced by verbal gerunds as in *He regrets [losing his keys]* (Fonteyn, De Smet and Heyvaert 2013).

In this paper we present the results of a comparative analysis of 200 iNGs and 200 bNGs randomly extracted from the BNC and COCA corpus. We map out their referential status, distinguishing between referential and non-referential (3) uses, and, within the referential category, between specific (1), opaque or 'virtual' (Lyons 1999; Langacker 2004) (4), and generic (2) instances:

- (3) Prayer is (...) a sharing of experiences and practices. (BNC)
- (4) [This] may result in a raising of the self esteem of teachers. (COCA)

Results show that there is a clear division of labour between iNGs and bNGs in PDEnglish:

- (i) iNGs are the default option to express specific reference, whereas bNGs have specialized in generic reference;
- (ii) A comparative referential analysis of NGs with a set of 200 regular noun phrases shows that the overall referential behaviour of iNGs and bNGs ties in with that of count indefinite NPs and bare mass NPs, respectively;
- (iii) Both iNGs and bNGs are frequent in so-called 'opaque' or 'virtual' contexts (i.e. following verbs like *want*, *believe*, *hope*; in negated, future or modal contexts etc.).

Our detailed analysis of the referential status of iNGs and bNGs reveals (1) that NGs, rather than involving a clear-cut distinction between mass noun-like, non-lexicalized NGs and fully lexicalized count nouns in *-ing*, form a cline in which bNGs seem to represent the mass noun end (the zero determiner lacking 'delimitation', Langacker 2004: 104) and iNGs resemble count nouns in their ability to delineate situations; (2) we argue that the occurrence of iNGs is motivated by the (explicit or implicit) delimitation of the (temporal) instance that is referred to (as discussed for regular mass NPs in e.g. Allen 1966; Langacker 1991; Swan 2005); (3) we map out the various types of delimitation found with iNGs and address the frequent use of both iNGs and bNGs in opaque contexts.

## References

- Allen, Robert L. (1966). *The verb system of Present-day American English*. The Hague: Mouton
- Brinton, Laurel (1998). Aspectuality and countability: a cross-categorial analogy. *English Language and Linguistics* 2(1): 37-63
- Fonteyn, Lauren, Hendrik De Smet and Liesbet Heyvaert (2013). Gerunds in Late Middle and Late Modern English. A case of shifting discourse-functional paradigms? Paper presented at the SLE conference in Split, September 2013
- Jespersen, Otto (1914-1927). *A Modern English Grammar, vol VII*. Oxford: Allen & Unwin
- Langacker, Ronald (1991). *Foundations of Cognitive Grammar 2: Descriptive Application*. Stanford: Stanford UP
- Langacker, Ronald (2004). Remarks on nominal grounding. *Functions of Language* 11(1): 77-113
- Wik, Berit (1973). *English Nominalizations in -ing: Synchronic and Diachronic Aspects*. Uppsala : Almqvist och Wiksell.

### Fluctuating statistics, volatile genres, and structural change: modality in the extended Brown family of corpora

Christian Mair

University of Freiburg, Germany

The original “Brown quartet” of matching diachronic corpora (Brown, LOB, Frown, FLOB) made possible the corpus-based study of ongoing change in written American and British English in real time, in a thirty-year time window in the second half of the 20<sup>th</sup> century. Among the more striking changes noted were rather drastic decreases in the frequency of some core modals (e.g. *must*, *shall*) and notable increases in the rate of use of some semi-modals (e.g. *have to*, *need to*) (see Leech 2003, Leech et al. 2009: 71-117). To account for the observed shifts in the corpora, two major factors were identified: structural change (for example through grammaticalisation) and discourse-level changes in genre conventions and stylistic norms. In view of the complexity of the constructional networks involved, it was difficult to disentangle the role of these factors in individual cases, and given the short time window available for observation, random statistical fluctuation could not be ruled out completely, either. The present paper uses data from BLOB (“Before LOB”, the 1930s LOB clone), BBrown (the 1930s Brown clone now nearing completion) and ARCHER 3.2 (also nearing completion) to contextualise the findings obtained in the original Brown quartet against the background of the wider history of the English modal system from the late 19<sup>th</sup> to the early 21<sup>st</sup> centuries. In particular, I will demonstrate that the extended Brown family, with its 60-year window of observation, makes it easier to assess the relative force of discourse and structural factors in the changes observed. As for the structurally motivated changes, the 60-year time window is generally sufficient to determine whether a particular change is (1) picking up speed, (2) undergoing a steady phase of rapid propagation, or (3) slowing down. In other words, we can locate an observed phase of a particular change on the appropriate range of the characteristic s-shaped trajectory from inception to completion.

#### References

- Leech, Geoffrey. 2003. “Modality on the move: The English modal auxiliaries 1961 – 1992.” In Roberta Facchinetti, Manfred Krug, and Frank Palmer, eds. *Modality in contemporary English*. Berlin: Mouton de Gruyter. 223-240.
- Leech, Geoffrey, et al. 2009. *Change in contemporary English: A grammatical study*. Cambridge: CUP.

### Directness in Ugandan English social letters – investigating the influence of socio-cultural context on sentence structure

Christiane Meierkord

Ruhr-University Bochum, Germany

In daily language use, grammatical choices (conscious or not) are frequently constrained by extralinguistic factors. One of these is politeness, and the relation between politeness and grammar has been widely discussed (e.g. Kirk 2013 in the context of the ICE project). This presentation discusses the interfaces between culture, politeness, and sentence structure as visible in Ugandan English social letters. In the social letters genre of the Ugandan (UG) component of ICE, we noticed a particularly low use of interrogative sentence structures in comparison to various other English-as-a-second-language (ESL) ICE corpora as well as to the Ugandan business letters (Nelson and Hongtao 2012). The ICE-UG data contains 59 interrogatives, whilst ICE-EA has 153, ICE-PHI 121, and ICE-SRI 276. A closer investigation indicated that the low amount of interrogatives correlates with a high amount of explicit and direct speech acts. Particularly request often go along with a usage of the speech act verb *request*, as in (1).

- (2) Dearest mother, I intend to shift my belongings from the hostel to my new residence this weekend, I request that you send me Denis to help me with the preparations and packing.

In the ICE-UG letters, *request* occurs seven times as an explicit speech act verb to request favours of the addressee. It does not occur in the business letters. In the ICE-EA letters, it occurs twice. In ICE- SRI it does not occur at all. Using explicit speech act verbs for the above purposes is not considered inappropriately direct by Ugandans, and, in line with previous observations (e.g. Anchimbe and Janney 2011), the Western association made between politeness and indirect speech acts does not seem to apply. Further speech act verbs occurring in the letters are *promise, pray, inform, advise* (as in (2)), *suggest, beg, wish, require, greet, thank, and encourage*.

- (3) Neighbour I advise you to take your children to playgrounds every weekend because they are just ruining everyone's life in my family.

My presentation explores to which extent the key to explaining the high amount of declarative sentence structures and direct / explicit speech acts may be found in the socio-cultural context of the data. Ugandans write social letters in English for very restricted purposes, mainly to inform relatives about events that require their (financial) assistance. To describe the relation between speech acts and sentence structure, directive speech acts in the ESL ICE corpora will be scrutinized in more detail to identify direct, indirect, and explicit speech acts, to explain in which context and for which purpose they are employed, and how their use correlates with interrogative and declarative sentence structure. The findings will serve as a basis for discussing the extent to which corpus linguistics can benefit from creating an interface with macro-sociolinguistics by taking into account the social context within which texts are composed as well as the socio-cultural norms of politeness that apply in the variety.

#### References

- Anchimbe, Eric A. and Janney, Richard W. (2011). Postcolonial pragmatics: an introduction. *Journal of Pragmatics*, 43(6). 1451-1459.
- Kirk, John M. (2013). Beyond the structural levels of language: an introduction to the SPICE-Ireland Corpus and its uses. In: Cruickshank, Janet and McColl Millar, Robert (eds.). *After the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster triennial meeting, Aberdeen 2012*. Aberdeen: Forum for Research on the Languages of Scotland and Ireland. 207- 232.
- Nelson, Gerald and Hongtao, Ron (2012). Particle verbs in African Englishes: nativization and innovation. In: Hundt, Marianne and Gut, Ulrike (eds.). *Mapping Unity and Diversity World-Wide: Corpus-Based Studies of New Englishes*. Amsterdam: Benjamins. 197-213.

## Segmentation in spoken language

Ilka Mindt

University of Paderborn, Germany

The research presented in this paper investigates how native speakers of English express grammatical function, and thus meaning, by means of spoken language. The focus will be on phonetic and intonational factors inside the verb phrase. The units analysed are represented by the pattern 'VERB to VERB'. The data used for this study are taken from the Michigan Corpus of Academic Spoken English (MICASE) and the Diachronic Corpus of Present-Day Spoken English (DCPSE).

(3) *he's no good he's only fit to be a scholar or a priest. uh so, they turn him in, and the authorities **come to question** him and he tells them, what he knows and they say okay we'll go search the mansion.*

(LEL140SU074)

(4) *Mei Lanfang who was influenced by what he saw in Shanghai once said when women first started going to see plays they were naturally rather unsophisticated or waihang, and what they **came to see** was spectacle, renao.*

(COL140MX114)

In example (1) from MICASE, the verb phrase *come to question* expresses a purposive meaning in which somebody came in order to question somebody. From a grammatical perspective, two main verbs are present. Example (2) contains a similar verb phrase, *came to see*. However, in contrast to (1), *came to see* in (2) represents an example of a metaphorical use of *come to* followed by a verb. This use of *come to* is termed catenative verb by Quirk et. al. (1985: 146) and by Mindt (2000: 289).

The main question of this research is whether speakers chunk the units exemplified in (1) and (2) differently. The hypothesis is that in (1), the unit [verb] + [TO] + [verb] shows a particular set of phonetic and intonational parameters that then gives rise to a grammatical analysis as two main verbs. In (2), the unit would be chunked [verb TO] + [verb] which shows different phonetic and intonational parameters that go hand in hand with an alternative grammatical analysis corresponding to a different function and meaning. The phonetic and intonational parameters that are investigated here are vowel and syllable duration in the first verb and the duration of the item "to", as well as its vowel quality. In addition, the intonation pattern of the whole unit 'VERB to VERB' is analysed. It will be demonstrated how different phonetic and intonational parameters can give clues to a segmentation of the spoken language which reflects grammatical patterns and meaning differences (see also Adolphs/Carter 2013 on chunking).

## References

- Adolphs, Svenja/Carter, R. (2013). *Spoken Corpus Linguistics. From Monomodal to Multimodal*. London: Routledge.
- Mindt, Dieter (2000). *An Empirical Grammar of the English Verb System*. Berlin: Cornelsen.
- Quirk, Randolph., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*, London: Longman.

### Changing patterns of sustainability discourse: A corpus-based analysis of corporate self-representation strategies

Alessandra Molino  
University of Turin, Italy

According to Dryzek (2013), *sustainability* discourse is one of the competing discourses that characterise the debate over the environment, the others being *environmental problem-solving*, *limits and survival* discourse and *green radicalism*. The characterising traits of sustainability discourse are the avoidance of radicalism and the attempt to “dissolve the conflicts between environmental and economic values” (Dryzek, 2013: 16) in order to produce small and reformist changes. As a consequence, sustainability discourse enables speakers to construct economic growth and environmental protection not as conflicting but as compatible goals. The aim of this study is to explore how businesses appropriate sustainability discourse to legitimize their practices turning environmental problems into profitable business opportunities, often without really challenging the *status quo* of their operations (Laine 2010). The focus is on self-representation strategies in disclosure statements about non-financial performance, which will be investigated using corpus-based techniques, concentrating on clusters related to self-mention devices in subject position, namely first person pronouns, the brand name and labels such as “the Group”. The analysis is both quantitative and qualitative: information about the incidence of use of self-mention clusters will be complemented with the study of their collocational (lexical verbs) and colligational (tense and aspect) profiles, and with the examination of their discourse functions (i.e. why such clusters are employed). The corpus for analysis consists of twelve Sustainability Reports (SRs) by Vodafone, the British telecommunications company. Vodafone was chosen because it has a relatively long and consistent tradition of social and environmental reporting, which makes it eligible for a case study of how sustainability discourse has changed over time. Although twelve years (from 2000 to 2012) may not be a very long time span, a diachronic approach is justified considering the dramatic growth of sustainability disclosure in the recent years (Pilot, 2011) and the increasing attention of the general public to environmental issues and ethical business management. The research questions addressed in this study are: 1) In what way is self-representation functional to the achievement private business goals through sustainability discourse? 2) How and why have self-representation patterns changed over time? The hypothesis is that Vodafone has progressively reshaped its public discourse moving from a focus on compliance to the company’s sustainability claims to increasing concerns for materiality (i.e. how to practically integrate sustainable development and business practices) emphasising goals and objectives. A focus on good intentions rather than concrete actions may serve the purpose of de-problematizing environmental issues to present them as achievable and profitable targets. At the same time, however, the SR appears as a highly strategic genre, which enables companies to adjust discourse to the changing context of their business operations. In the case of Vodafone, the expansion in emerging markets and the consequent negative environmental impacts have probably induced the company to foreground the positive effect of its current business practices in the most recent SRs.

#### References

- Dryzek, John S. (2013). *The Politics of the Earth: Environmental Discourses* [3<sup>rd</sup> edition]. Oxford and New York: Oxford University Press.
- Laine, M (2010). Towards sustaining the status quo: Business talk of sustainability in Finnish corporate disclosures 1987–2005. *European Accounting Review* 19(2), 247-274.
- Pilot, S. (2011). Companies are embracing corporate responsibility in their annual reports. Article posted for the Guardian Professional Network, Thursday 29 September 2011 on [www.guardian.co.uk](http://www.guardian.co.uk).



## Phrasal verbs in spoken and written L2 learner English

Susan Nacey, Anne-Line Graedler  
Hedmark University College, Norway

This paper presents research on the use of phrasal verbs (PVs) in Norwegian L2 English, addressing the following questions:

- 1) Are there contrasting syntactic/semantic patterns of PV usage across the spoken and written modes?
- 2) How often do learners produce anomalous PVs?
- 3) Is there a correlation between anomalous PV usage and figurative use?
- 4) Is there a correlation between anomalous PV usage and L1 transfer?

PV usage is generally acknowledged as “one of the most problematic areas for learners of English” (Jenkins 2009: 52). First, PVs are subject to syntactic restrictions that may not be readily apparent to learners. Second, semantics poses a challenge, because PVs are often highly polysemous with both literal and (often several) figurative meanings. Third, negative L1 transfer may also play a role, especially for learners whose L1 has similar verb-particle constructions. In addition, stylistic considerations have an impact as do the ways in which PVs are traditionally taught (see e.g. Cowie 1993: 38-39; Waibel 2007: 21-32). At the same time, PVs are viewed as highly important to proficiency in English. Much of the previous research focuses on learner avoidance of PVs rather than actual usage. However, there are exceptions, such as the corpus-based studies by Hägglund (2001), Waibel (2007), and Gilquin (forthcoming). Pye’s (1996) report on PV errors in the Cambridge Corpus of Learner English, asserts that learners often make mistakes in this area. The present paper adds empirical evidence concerning the real magnitude of the challenge that PV use presents, by investigating the PV production in written and spoken language of Norwegian learners, thus comparing PV use across modes and for an added group of English language learners. The data consists of all PVs uttered by fifty L2 English students in the entire Norwegian subcorpus of LINDSEI (Gilquin et al. 2010) – approximately 13 hours of conversation equaling 83,000 words – along with all PVs in roughly 83,000 words of argumentative texts retrieved from the Norwegian component of the ICLE corpus (Granger et al. 2009). Informants for both corpora were Norwegian college students characterized as higher-intermediate to advanced learners of English. PVs are here viewed as distinct from prepositional verbs and free combinations, but include phrasal-prepositional constructions (cf. Quirk et al. 1985: 1152 and 1167). All PVs in the material were categorized according to structural type: Verb+Particle, Verb+Object+Particle, and Verb+Particle+Object (following Gilquin, forthcoming). The PVs in each pattern were then analyzed along a nominal scale for metaphoricity (yes/no/don’t know) (see Nacey 2013; Steen et al. 2010). Such classification allows for comparison not just of overall PV usage in the two corpora, but also for preferences for particular syntactic and/or semantic patterns across modes. An additional focus concerns ‘anomalous’ PVs - identified through lack of codification (either of the entire PV or of the particular contextual meaning) in standard English dictionaries - with a view towards establishing whether learner challenges increase as the contextual meaning shifts away from a core, concrete meaning to a more peripheral, metaphorical meaning. Finally, anomalous PVs were also investigated for possible L1 transfer of the verb, the particle, or the entire construction.

## References

- Cowie, A. P. (1993). Getting to grips with phrasal verbs. *English Today*, 9(4), 38-41.
- Gilquin, G. (forthcoming). The use of phrasal verbs by French-speaking EFL learners. A constructional and collstructional corpus-based approach. *Corpus Linguistics and Linguistic Theory*.
- Gilquin, G., Cock, S. D., & Granger, S. (Eds.). (2010). *LINDSEI: Louvain International Database of Spoken English Interlanguage*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2009). *International Corpus of Learner English, Version 2*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Hägglund, M. (2001). Do Swedish advanced learners use spoken language when they write in English? *Moderna Språk*, 95(1), 2-8.
- Jenkins, J. (2009). *World Englishes: a resource book for students*. London: Routledge.
- Nacey, S. (2013). *Metaphors in learner English*. Amsterdam: John Benjamins.
- Pye, G. (1996). Don't give up, look it up! Defining phrasal verbs for the learner of English. *EURALEX '96 Proceedings*.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: from MIP to MIPVU*. Amsterdam: John Benjamins.
- Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students*. PhD, Albert-Ludwigs-Universität, Freiburg i. Br.

### Evaluative Lexis in the News Reports of the 2011 London Riots: A Corpus-Based Discourse Analysis

Maria Cristina Nisco, Marco Venuti  
Università di Napoli Federico, Italy

Moving from the assumptions that evaluation is a significant device in news discourse to interpret the world adopting specific stances (Thompson and Hunston 2000, Bednarek 2006), and that it is regarded as an action performed in discourse (Hunston 2010), this paper explores the ways in which language is used to represent and evaluate the events connected to the 2011 London riots. The focus will be on the set of contextually determined meanings and resources employed to assign a certain status to the actions and actors involved in the riots. Therefore, special emphasis will be given to the evaluative portrayal of some of the participants and to the construction of their *textual personas* as enacted by six British newspapers representing different political orientations (left leaning and right leaning) within the quality and popular press: *The Guardian*, *The Telegraph*, *The Times*, the *Daily Mail*, the *Daily Mirror* and *The Sun*. A corpus of about 1.700 articles published between August 1<sup>st</sup> and December 31<sup>st</sup> 2011 has been analyzed with the aim to shed light on how newspapers perform their dual function to construe reality and construct a persuasive argument, while directly addressing their readership. An approach combining corpus analysis and discourse analysis (Baker et al. 2013, Partington et al. 2013) has been adopted. After collecting a reference corpus consisting of articles published by the same newspapers in the same time span and comprising all the articles of a daily edition, our first step was the identification of keywords (Scott and Tribble 2006, Scott 2008) for each newspaper. Keywords were then classified according to four categories (participants, their pre-modifiers, processes and their goals) in order to identify the main actors and their role in the media representation of the events, investigating lexical choices in terms of varying degrees of connotation. The analysis of the keywords has shown that there are marked differences but also similarities among the six newspapers. They all report on the violent and criminal nature of the events and on the trials that followed. However, tabloids focus mainly on the sensational aspects of individual stories and on the representation of rioters as 'young thugs'. Within the broadsheets the right leaning ones share some of the features identified in the popular press, while *The Guardian* is the only newspaper offering a wider and more comprehensive account of the events, their causes and their outcomes.

#### References

- Baker, P., Gabrielatos, C. and McEnery, T. (2013) *Discourse Analysis and Media Attitudes: The representation of Islam in the British press*. Cambridge: Cambridge University Press.
- Bednarek, M. (2006) *Evaluation in Media Discourse. Analysis of a Newspaper Corpus*. London/New York: Continuum.
- Hunston, S. (2010) *Corpus Approaches to Evaluation. Phraseology and Evaluative Language*. London: Routledge.
- Hunston, S. and Thompson, G. (eds) (2000) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford 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.
- Scott, M. (2008) *WordSmith Tools 5.0*. Oxford: Oxford University Press.
- Scott, M. and Tribble, C. (2006) *Textual Patterns: keyword and corpus analysis in language education*. Amsterdam: Benjamins.

### Sociolinguistic predictors of preposition stranding and fronting in 18th & 19th century Spoken English

Magnus Nissel

University of Giessen, Germany

This paper investigates diachronic change in the frequency of stranded (1)(3) and fronted (2)(3) prepositions in *wh*-questions and *wh*-relative clauses and the strength of the speaker's gender, social class, and role in the courtroom (lawyer, victim, defendant, witness, etc.) as predictors.

- (1) Where was you going to?
- (2) On what day was this?
- (3) In what position was you in?

While preposition placement is by no means understudied, most quantitative research has focused on Present-Day English (Gries 2002, Hoffmann 2011). Research on stranding and fronting in Late Modern English was conducted on the basis of comparatively small collections of letters and written texts from the *Helsinki Corpus* and *ARCHER* (e.g. Yáñez-Bouza 2008). This study uses a 11.8 million word subset of the Old Bailey Corpus (OBC, Huber et. al. 2012) that consists of utterances for which both the gender and the social class of the speakers is known. The OBC contains richly-annotated court transcripts from London's Central Criminal Court. It is based on the digitized *Proceedings of the Old Bailey* (<http://www.oldbaileyonline.org>) and spans the period from 1720 to 1913. The transcribed speech passages in the corpus are a reasonably close representation of the spoken English for the period. This study investigates 10 prepositions (*of, in, to, for, with, on, at, from, by, about*) that occur both in stranded and fronted contexts. All questions and relative clauses that contain at least one *wh*-pronoun and preposition are extracted from the corpus, resulting in more than 50,000 tokens. This is done with a Python script that also retrieves the available sociobiographical and pragmatic annotation for each utterance and gathers additional information – such as the position of all *wh*-pronouns and prepositions in a phrase – to facilitate the manual classification. False positives are removed by hand. A preliminary analysis based on a random sample of *wh*-questions shows a significant decline in preposition stranding over time. This paper will analyze the effect of gender, social class, and courtroom role in both *wh*-questions and *wh*-relative clauses. It will also consider the interaction of these sociobiographical and pragmatic variables with linguistic variables such as the type of verb or the length of the phrase.

### References

- Gries, Stefan. 2002. "Preposition stranding in English: predicting speakers' behaviour." In *Proceedings of the Western Conference on Linguistics*. Vol. 12, ed. Samiian, Vida, 230-241. California State University
- Hoffmann, Thomas. 2011. *Preposition Placement in English: A Usage-based Approach*. CUP.
- Huber, Magnus; Nissel, Magnus; Maiwald, Patrick; Widlitzki, Bianca. 2012. *The Old Bailey Corpus. Spoken English in the 18th and 19th centuries*. [www.uni-giessen.de/oldbaileycorpus](http://www.uni-giessen.de/oldbaileycorpus)
- Yáñez-Bouza, Nuria. 2008. "Preposition stranding in the eighteenth-century: Something to talk about." In *Grammars, Grammarians, and Grammar-Writing in Eighteenth-Century England*, ed. Tiekens-Boon van Ostade, Ingrid, 251-277. Mouton de Gruyter

**A comparison of data commentary in chemical engineering research papers and master theses – exploring intersemiosis of written and visual material for applied purposes.**

Lene Nordrum  
Lund University, Sweden

In the field of English for Specific Purposes, data commentary, the written comment on visual material has been recognized as a particularly complex area that presents problems even for fairly advanced university students (Blåsjö 2011; Wharton 2012). Yet, few studies focus on it. In approaching data commentary from a pedagogical perspective, two problems emerge 1) practices tend to vary depending on disciplinary field and on the type of visual presented (Stoller and Robinson, 2013), and 2) teachers in the various science fields find it challenging to teach data commentary, possibly because they are too entrenched in disciplinary practices (Blåsjö, 2011). To uncover areas of particular difficulty for ESP students, there is thus a need for investigation of disciplinary variance of how the written and visual modes are integrated in disciplinary texts as well as of studies comparing expert and student writing (Gilquin et al, 2007). In a first step towards addressing these needs, this study investigates differences in data commentaries in master theses written by Swedish advanced learners of English and published research articles written by native and non-native disciplinary experts. All data commentaries are extracted from the result and/or result & discussion section of papers from chemical engineering and compose a small discipline-specific *learner* (master theses) and *expert* (published research articles) corpus, where the definition of learner is motivated by a distinction between *apprentice* and *expert* writers. The data commentaries are annotated for rhetorical moves (Swales, 1990) according to the Biber-Connor-Upton approach (Biber et al., 2007) and by means of the UAM corpus tool, developed by Michael O'Donnell. Preliminary results indicate that expert and novice writers differ in their choices both at the level of discourse, in terms of selection of rhetorical moves, and at the level of lexicogrammar, in terms of the phraseology associated with specific moves. An intriguing finding concerns differences in the lexicogrammatical realization of the rhetorical move addressing the most important result, or trend, displayed in a visual. Expert writers in chemical engineering commonly use a lexical nominalization for this move, whereas students overwhelmingly rely on *that*-clauses. I argue that the experts' use of lexical nominalization is similar to the well-known rhetorical function of lexical nominalization to condense Given information (Halliday and Martin, 1993) in science discourse, but here the Given information is first presented in a visual. This finding is of importance for ESP courses addressing the intersemiosis (O'Halloran, 2005) of visual and written material in the multi-modal presentation of results in chemical engineering, and is an interesting topic for further studies of data commentary and multimodality across disciplines.

## References

- Biber, D., U. Connor, J. Jones & T. Upton. 2007. *Discourse on the Move*. Amsterdam/Philadelphia: John Benjamins.
- Blåsjö, M. 2011. From percentage to prediction: University students meeting a parallel language of visuals and numerals. *Ibérica* 22, 123-140.
- Gilquin, G., S. Granger, M. Paquot. 2007. Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes* 6, 319-335.
- Halliday, M.A.K. and J. Martin. 1993. *Writing Science. Literacy and Discursive Power*. London: Falmer Press.
- O'Halloran, K.L. 2005. *Mathematical Discourse: Language, Symbolism and Visual Images*. London: Continuum.
- Stoller, F-L. & M.S. Robinson. 2013. Chemistry journal articles: An interdisciplinary approach to move analysis with pedagogical aims. *English for Specific Purposes* 32, 45-57.
- Swales, J.M. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge university press.
- Wharton, S. 2012. Epistemological and interpersonal stance in a data description task: Findings from a discipline-specific learner corpus. *English for Specific Purposes* 31, 261-270.

## Negative intensification in the spoken language of British adults and teenagers. A preliminary corpus-based study

Ignacio Miguel Palacios Martínez

Intensification, a general device used by speakers to convey their message more clearly and to strengthen their position to it (Bolinger 1972), has been discussed widely in the literature. However, the specific use of intensification on negatives (i.e. *it didn't do any harm at all, I hadn't the faintest idea who he was*) has received much less attention. There are some brief descriptions in general reference grammars of English (Quirk et al. 1985: 785, Downing and Locke 2002: 22, Huddleston, Pullum et al 2002: 828), and a few pages on the issue can also be found in monographic studies on negation (Jespersen 1917, Tottie 1991, Bernini and Ramat 1996, van der Wouden 1997, Mazzon 2004, Horn 1989, 2011). The aim of this paper is to identify and study more thoroughly the intensifying resources used by British adults and teenagers in the strengthening of negative constructions. A second focus of interest will be the extent to which differences exist here in the language of adults and teenagers, given that the latter are, broadly speaking, more prone to intensification: teens show a generally higher level of spontaneity in their expression and are fond of lending more force to their message to make it more credible. In order to explore these issues, I will analyse data extracted from comparable samples of several British English corpora: DCPSE (Diachronic Corpus of Present-Day Spoken English) and BNC (British National Corpus) for adult language, and COLT (The Bergen Corpus of London Teenage Language) and LIC (Linguistic Innovators Corpus) for teen talk. Findings indicate that the strengthening of negatives in English can be achieved in different ways: a) by means of a number of expressions with negative import, such as *in the slightest, in the least, at all, even* (i.e. *that doesn't surprise me in the slightest*); b) the repetition of the adverb *never* or the combination *never ever* (i.e. *They were never never any hassle, I've never ever heard Jim's voice before*); c) particular uses of *never*, specially when negating something in the past (i.e. *Vernon never called for me yesterday*); d) cases of multiple negation or negative concord intended to heighten a negative meaning (i.e. *I didn't get no fucking matches, did I?*); and e) negative polarity collocations and idiomatic expressions (i.e. *I didn't have a clue what she was talking about, I haven't got a piss boy*). Contrary to expectations, no significant differences in the frequency of negative intensification in the two sociolects were found. However, it was observed that while teenagers prefer the use of idiomatic or semi-idiomatic expressions together with locutions such as *no way* (i.e. *No way am I hanging out with these people again*), adults opt more often for negatively oriented polarity sensitive items (NPIs).

### References

- Bernini, Giuliano and Paolo Ramat. 1996. *Negative sentences in the languages of Europe: A typological perspective*. Berlin and New York: Mouton de Gruyter.
- Bolinger, Dwight. 1972. *Degree words*. The Hague and Paris: Mouton de Gruyter.
- Downing, Angela and Philip Locke. 2002. *A university course in English grammar*. London: Routledge.
- Horn, Lawrence. 1989. *A natural history of negation*. London and Chicago: University of Chicago Press.
- Horn, Lawrence. (ed.). 2011. *The expression of negation*. Berlin and New York: Mouton de Gruyter.
- Huddleston, Rodney, Geoffrey Pullum et al. 2002. *The Cambridge grammar of English*. Cambridge: Cambridge University Press.
- Jespersen, Otto. 1917. *Negation in English and other languages*. De Kgl. Danske Videnskabernes Selskab. Historisk-Filologiske Meddelelser I.5. Copenhagen. Reprinted in 1962 in *Selected writings of Otto Jespersen*. London: Allen and Unwin.
- Mazzon, Gabriela. 2004. *A history of English negation*. London: Pearson-Longman.
- Quirk, Radolf, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London and New York: Longman.
- Tottie, Gunnel. 1991. *Negation in English speech and writing*. San Diego/London: Academic Press.
- Van der Wouden, Ton. 1997. *Negative contexts. Collocation, polarity and multiple negation*. London and New York: Routledge.

## Involved or Informational? Gradability of Adjectives in Direct Conversation and Academic Writing

Yinqing Qiu

University of Hong Kong, Hong Kong

Gradability of adjectives has been extensively discussed in not a few grammar books and journal articles (e.g., Quirk et al., 1985; Collins Cobuild, 1990; Biber et al., 1999; Paradis, 1999, 2001; Huddleston & Pullum, 2002; Kennedy, 1999, 2007). While these works have significantly deepened our understanding of the gradable features of adjectives, most of them failed to consider register differences. The underlying assumption is that the linguistic phenomenon in English can be described in a general manner (Biber, 2012). However, global generalizations are not accurate and marked linguistic differences exist across registers or sublanguages (Kittredge, 1982). Using the British Component of the International Corpus of English (ICE-GB) (Nelson et al., 2002) as data, this study investigates register differences of the use of adjectives in terms of their gradable features with a purpose to test whether gradability of adjectives can be used to distinguish registers of different channels, i.e., the spoken and written channel. Accordingly, the research questions are as follows: (1) What are the distributional features of various types of gradable and non-gradable adjectives in the spoken and written registers? What are their similarities and differences? (2) How does contextual modulation (i.e., the phenomenon of one adjective switching between different types of gradable features in different contexts, e.g., between gradable and non-gradable adjectives) work in the spoken and written registers? What are the underlying reasons behind? The data chosen for the current study are the adjectives in two registers of ICE-GB – Direct Conversation and Academic Writing for reasons that these two registers represent the two ends of a spoken-written continuum based on the dimension of *Involved vs. Informational Production*, which is the most powerful predictor of register differences with a predictive power of 84.3% (Biber, 1988: 127). Due to the large amount of adjectives in these two registers, only 10% of them are randomly sampled for investigation, which amounts to around 1500 tokens. The categorization of these adjectives into various gradable and non-gradable groups rigidly follows Paradis's (2001) two criteria and the notion of 'boundness' as the standard for classification: 1) the type of **degree modifiers** with which the adjective may combine; 2) the type of **opposition** involved in the conceptualization of the adjective. The subsequent data analysis confirms the existence of register variation in the use of adjectives in terms of their gradable features and yields some interesting findings which have not been mentioned by existing descriptions, such as the tendency for limit, extreme, and even non-gradable adjectives to get a scalar reading in Direct Conversation and the opposite tendency for Academic Writing. The underlying reasons behind can be attributed to the different communicative purposes of these two spoken and written registers (*involved vs. informational production*) and the different semantic meanings conveyed by various types of gradable and non-gradable adjectives, which validates the hypothesis that the gradable features of adjectives can serve as a differentia for discriminating registers.

## References

- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge, UK: Cambridge University Press.
- Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, 8(1): 9-37.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London, UK: Longman.
- Collins Cobuild. (1990). *Collins Cobuild English Grammar*. London, UK: Harper Collins Publishers.
- Huddleston, R., & Pullum, G.K. (2002). *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press.
- Kennedy, C. (1999). *Projecting the Adjective: the Syntax and Semantics of Gradability and Comparison*. New York: Garland Pub.
- Kennedy, C. (2007). Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguist Philos*, 30(1): 1-45.
- Kittredge, R. (1982). Variation and homogeneity of sublanguages. In R. Kittredge & J. Lehrberger (Eds.), *Sublanguage: Studies of Language in Restricted Semantic Domains* (pp. 107-137). New York: Walter de Gruyter.
- Nelson, G., Wallis, S., & Aarts, B. (2002). *Exploring Natural Language: Working with the British component of the International Corpus of English*. Amsterdam, NL: Benjamins.
- Paradis, C. (1997). *Degree Modifiers of Adjectives in Spoken British English*. Lund, SE: Lund University Press.
- Paradis, C. (2001). Adjectives and boundedness. *Cognitive Linguistics*, 12(1): 47-65.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London, UK: Longman.

## Beyond Collocations – How Local Grammar affects the Choice of Translation Equivalents

Renate Reichardt

University of Birmingham, United Kingdom

The lexical and syntactic patterning observed in language use in the form of collocations and colligations is often described as “routine” (Stubbs 1993: 2), “stereotyped” (Clear 1993: 272) or “primed” (Hoey 2005: 8) use of language. This indicates that language use is to a large extent based on conventions amongst its users. The talk explores the conventions amongst translators with regard to the interplay of the local grammar of a word and meaning interpretation, whereby the ‘meaning’ of a word in contrastive studies is established through translation equivalents (cf. Teubert 2001: 144). The English verb CONSIDER and its near-synonyms BELIEVE, FEEL and THINK are explored together with their German translation equivalents. The underlying assumption of the approach is that the meaning of a word is determined by its unique syntactic and / or semantic surroundings, expressed by Wittgenstein (in Firth 1968: 179) as “the meaning of a word lies in its use”. Valency theory proves to be a suitable methodology as it is capable of distinguishing between different levels of language analysis. Its flexibility regarding the valency complement categorisation types allows to overcome the differences on the surface level between English and German and to define categories that are equally suitable for both languages. For example, the specific question of whether the meaning of the verb CONSIDER is different or the same in German when it occurs in a divalent structure with a subject and an object complement, as in the sentence “We have considered all the points in the resolution”, than when it occurs in a trivalent structure with a subject, object and an adjectival complement, as in the sentence “We consider the reforms necessary”, will be addressed. The corpus approach includes qualitative and quantitative analysis of the data. Following Sinclair (1991) and Groom (2007) the patterns and the translation equivalents were identified manually based on a sufficiently large number of concordance lines. Frequency analysis was used to verify relevant valency sentence patterns (cf. Ágel 1988) and key translation equivalents. The findings highlight the problems of investigating the lexicogrammatical interplay in a contrastive context and demonstrate that the use of words is constrained by their local grammar, whereby ‘grammar’ has to be understood syntactically and lexically. Thus, meaning is not only defined lexically, i.e. through phrases and collocations, but also grammatically, i.e. through colligations represented as valency complements in the investigation. As a consequence, language competence requires both syntactic and lexical knowledge.

## References

- Ágel, V. (1988) *Überlegungen zur Theorie und Methode der historisch-synchronen Valenzsyntax and Valenzlexikographie*, Tübingen: Max Niemeyer
- Clear, J. (1993) ‘From Firth Principles – Computational Tools for the Study of Collocation’ in Baker, M., Francis, G., Tognini-Bonelli, E. (eds) *Text and Technology – In Honour of John Sinclair*, pp 271-292, Amsterdam: John Benjamins
- Firth, J. R. (1968) *Selected Papers of J. R. Firth 1952-59*, edited by Palmer, F. R., London: Longmans
- Groom, N. (2007) *Phraseology and epistemology in humanities writing: a corpus-driven study*, PhD thesis submitted to The University of Birmingham
- Hoey, M. (2005) *Lexical Priming – A new theory of words and language*, Abingdon: Routledge
- Sinclair, J.M. (1991) *Corpus, Concordance, Collocation*, Oxford: Oxford University Press
- Stubbs, M. (1993) ‘British Traditions in Text Analysis – From Firth to Sinclair’ in Baker, M., Francis, G., Tognini Bonelli, E. (eds) *Text and Technology – In Honour of John Sinclair*, pp 1-34, Amsterdam: John Benjamins
- Teubert, W. (2001) ‘Corpus Linguistics and Lexicography’ in *International Journal of Corpus Linguistics* Vol. 6, pp 125-153

**Hapax Legomena: Patterns and Functions in Text**

Antoinette Renouf

Birmingham City University, United Kingdom

In recent years, we have examined aspects of neology from a modern diachronic corpus-based perspective. We have sought to model generalisable facts about neology, based on evolving patterns of occurrence. One model, for example, demonstrates the life-cycle of a word in newstext (Renouf, 2013), another, the pattern of cumulative neologising activity accompanying 'alpha' neologisms (Renouf, forthcoming). Our focus in those cases was on neologisms which capture the public and media imagination. Lately, however, we have turned our attention to the 'silent majority' of neologisms which have not recurred in our corpus (so far). The 'hapax legomenon' is a word occurring once only; over half the word types in a large corpus of text are of this kind. A hapax formation can both remain a neologism and become 'conventionalised' (Kerremans, 2012) or 'institutionalized' (Fischer, 1998) in the language. Such words have been traditionally disregarded in lexical inventories, particularly those intended for large-scale database search in commercial settings, since they are judged by default to be errors or of no linguistic consequence. Typographical errors aside, understanding the nature and role of hapax legomena in text is of value for several reasons. As an object of study, hapaxes are the best window to the productivity of the language (Baayen and Renouf 1996). Such formations are not randomly but strategically created, for service within a specific context. Though they do not contribute significantly to the lexicon in bulk, hapax formations enrich it in variety and flexibility of expression. Crucially, at both word and sub-word level, hapaxes combine to form 'classes' or 'sets', semantic and other, which bring coherence to text across time, as well as influencing its evolution. As Boussidan (2013) says, "Words that appear once... have something to say when they are part of a substantial set". The study will select from a database of all words recorded as occurring for the first time in an unbroken flow of 'Independent' and 'Guardian' news text from 1984 to end 2011. The primary analytical apparatus will be the WebCorp Linguist's Search Engine (Kehoe and Gee, 2009). Our aim in this paper is to conduct a pilot study of this subset of neologisms passing through news text, identifying their patterns and functions, as suggested above. In doing so, we are essentially observing the productivity of the range of forms to which words freshly and uniquely attach to create new hapaxes: affixes, combining forms, grammaticalising words and words themselves. We also assess the creativity of hapaxes. The study of English hapax legomena sits at the intersection of pure and applied areas of linguistic study, software development and lexical statistics. The new knowledge gained is expected to throw new light on the evolving lexicon which is of use to language teachers and researchers, NLP specialists, dictionary makers and the language professions.

**References**

- Boussidan, Armelle. (2013). 'Dynamics of Semantic Change. Detecting, analyzing and modelling semantic change in a corpus of short diachrony'. Thèse de Doctorat de Sciences Cognitives, Université Lumière-Lyon 2.
- Fischer, Roswitha. 1998. *Lexical Change in Present-day English. A Corpus-based Study of the Motivation, Institutionalization, and Productivity of Creative Neologisms*. Tübingen: Narr.
- Kehoe, Andrew and Matt Gee. (2009), 'Weaving Web data into a diachronic corpus patchwork', in: Antoinette Renouf and Andrew Kehoe (eds.) 255-279
- Kerremans, Daphné. 2012. *A web of new words: On the conventionalisation of English neologisms*. Unpublished PhD thesis, Ludwig-Maxilians-University Munich.
- Renouf, Antoinette. 2013. A Finer Definition of Neology in English: the life-cycle of a word, in Hasselgård Hilde, Signe Oksefjell Ebeling and Jarle Ebeling (eds.). *Corpus Perspectives on Patterns of Lexis*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Renouf, Antoinette. (forthcoming). A finer definition of neology in English: From word to register. In Davidse, Kirsten, and Lieven Vandelotte (eds.). *Advances in Corpus Compilation and Corpus Applications*. Amsterdam: Rodopi.
- Renouf, Antoinette & Andrew Kehoe (eds.). (2009), *Corpus Linguistics: Refinements and Reassessments*. Amsterdam: Rodopi.



## Investigating expletives in varieties of English

Patricia Ronan

Université de Lausanne, Switzerland

The use of expletives has been a popular research topic for decades, and various studies have been carried out on properties of, partly select, expletives in specific varieties of English (McEnery and Xiao 2004, Murphy 2010, Wierzbicka 2002). Further, research has been done on diachronic or international typologies of expletives (McEnery 2006, Ljung 2010). It has been found that expressives are normally based on taboo items within a society, which, according to Andersson and Trudgill (1991: 15), are mostly derived from religious expressions (*God, Jesus, Christ*), animal comparisons (*bitch, pig*), or expressions concerned with body-related matters such as sex and excretion (*fuck, shit*). Overall, swearwords are considered to be mainly an example of youth language, but it has been shown that they can also function as in-group markers which strengthen group solidarity (Aitchinson 2006: 23). Particularly in this context these items are not only used as expletive expressions, but also as intensifiers, as Wierzbicka (2002) has shown for the use of the expression *bloody*. The use of swearwords is not, however, stable across time and place, and the preferred swear-items are not only subject to change over time, but they also vary in different cultures (Ljung 2010). This being the case, it seems likely that not only societies from different linguistic backgrounds will use diverse types of swearing, but also that different English speaking societies across the globe will differ in their use of expletives. So far, little comparative, corpus based research has been carried out on this topic, however. The proposed paper will investigate to what extent variation can be found in swearwords in different varieties of English, and to what extent this variation is likely to be influenced by differences in societal taboos. To this end we will compare the use of a select set of swearwords from the religious and body-related domains. These are collected with the use of the corpus tool AntConc (Anthony 2005) in different ICE corpora representing Inner and Outer Circle varieties of English (ICE-GB, ICE-IRE, ICE-CAN, ICE-IND, ICE-SIN). We will aim to determine to what extent quantitative and qualitative differences can be found in use of religious and body-related expletives in these ICE corpora and the findings will allow us to consider the influence of cultural and societal parameters on the use of swearwords in the varieties of English under scrutiny.

## References

- Andersson, L. and P. Trudgill. 1991. *Bad Language*. London: Blackwell.
- Anthony, L. 2005. 'AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit'. *Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning*, 7-13.
- Farr, F. and B. Murphy. 2009. 'Religious references in contemporary Irish-English: "for the love of God almighty....I'm a holy terror for turf".' *Intercultural Pragmatics*, 6(4), 535-560.
- Ljung, M. 2010. *Swearing. A Cross-Cultural Linguistic Study*. London: Palgrave Macmillan.
- McEnery, A. 2006. *Swearing in English: bad language, purity and power from 1586 to the present*. London: Routledge.
- McEnery, A. and Z. Xiao. 2004. 'Swearing in modern British English.' *Language and Literature* 13, 235-68.
- Murphy, B. 2009. "'She's a fucking ticket": The pragmatics of FUCK in Irish English – an age and gender perspective.' *Corpora* 4.1, 85-106.
- Wierzbicka, A. 2002. 'Australian cultural scripts – *bloody* revisited.' *Journal of Pragmatics* 34, 1167-1209.

### Marked themes in a contrastive and learner perspective

Sylvi Rørvik

Hedmark University College, Norway

Contrastive studies of fiction texts in English and Norwegian have shown that Norwegian has greater tolerance for marked themes than English does, despite similar syntactic resources in the two languages (Hasselgård 1997, 2004, 2005). However, Rørvik (2013: 51-52) found that argumentative newspaper texts in English and Norwegian do *not* differ in their proportions of marked and unmarked themes, thereby indicating that text type may influence the frequency of marked themes in the two languages. This same investigation found that Norwegian advanced learners of English underuse marked themes compared to native speakers (ibid: 170). The present paper has two main aims: to further examine the use of marked themes in L1 argumentative newspaper texts in English and Norwegian, and to identify possible causes for the underuse of marked themes in L2 learner English. The following research questions are addressed:

1. Do L1 writers of English and Norwegian differ in the types of marked theme they use?
2. Is the underuse of marked themes in texts written by advanced learners caused by underuse of one specific type of marked theme? If so, can this be linked to (possible) cross-linguistic differences?

The method employed is based on the Integrated Contrastive Model (Gilquin 2000/2001), involving an initial contrastive analysis of the L1 material. This comparison serves as the basis for predictions about the learners' behavior. Next, the learner material is compared to the L1 English material in order to check whether the predictions hold true. To answer the first research question, the paper investigates all instances of marked themes in a corpus consisting of 100 English and 100 Norwegian argumentative newspaper texts, with the aim of identifying which constructions have been used in thematic position. To answer the second research question, the native-speaker texts are compared to 100 argumentative texts written by Norwegian advanced learners of English. This learner material comes from the Norwegian component of the International Corpus of Learner English (NICLE). In declarative sentences a marked theme involves a non-subject as theme (for other sentence types, see Halliday 2004: 78). Previous studies have identified overuse of initial adverbials in L2 texts compared to texts written by native speakers of English (see e.g. Shaw 2004: 76; Hasselgård 2009a: 95; Hasselgård 2009b: 126; Chen 2010: 78), and because such initial adverbials are classified as marked themes, the working hypothesis of this study is that the overall underuse of marked themes in the NICLE material is due to underuse of other types of marked theme, such as objects.

### References

- Chen, Xiao. 2010. *Discourse-Grammatical Features in L2 Speech: A Corpus-Based Contrastive Study of Chinese Advanced Learners and Natives Speakers of English*. Doctoral dissertation, The City University of Hong Kong.
- Gilquin, Gaëtanelle. 2000/2001. "The Integrated Contrastive Model. Spicing up your data." *Languages in Contrast* 3:1, 95-123.
- Halliday, Michael A. K. 2004. *An Introduction to Functional Grammar*. 3<sup>rd</sup> edition, revised by C. M. I. M. Matthiessen. London: Arnold.
- Hasselgård, Hilde. 1997. "Sentence openings in English and Norwegian." In Ljung, Magnus (ed.), *Corpus-based studies in English. Papers from the 17th International Conference on English Language Research on Computerized Corpora*. Amsterdam: Rodopi, 3-20.
- Hasselgård, Hilde. 2004. "Thematic choice in English and Norwegian." *Functions of Language* 11:2, 187-212.
- Hasselgård, Hilde. 2005. "Theme in Norwegian." In Berge, Kjell Lars and Eva Maagerø (eds.), *Semiotics from the North. Nordic approaches to systemic functional linguistics*. Oslo: Novus Press, 35-47.
- Hasselgård, Hilde. 2009a. "Temporal and Spatial Structuring in English and Norwegian Student Essays." In Bowen, Rhonwen, Mats Möbärg and Sölve Ohlander (eds.), *Corpora and Discourse – and Stuff. Papers in Honour of Karin Aijmer*. Gothenburg: Acta Universitatis Gothoburgensis, 93-104.
- Hasselgård, Hilde. 2009b. "Thematic choice and expressions of stance in English argumentative texts by Norwegian learners." In Aijmer, Karin (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, 121-139.
- Rørvik, Sylvi. 2013. *Texture in learner language*. Doctoral dissertation, University of Oslo.
- Shaw, Philip. 2004. "Sentence openings in academic economics articles in English and Danish." *Nordic Journal of English Studies* 3:2, 67-84.

## Colloquialization in TIME magazine: The case of inserts

Christoph Rühlemann

Philipps University Marburg, Germany

Recent analyses of written text types have discovered significant increases of conversational and/or colloquial language use, a trend referred to as conversationalization (Fairclough 1995) and colloquialization (e.g., Leech et al. 2009). Examples of features deemed typical of informal spoken language that have made inroads into writing include verbal and negative contractions (e.g., Leech 2002), personal pronouns (e.g., Baker 2009), questions, the progressive as well as zero relative clauses (e.g., Leech et al. 2009). This paper falls into two parts. In the first part I present new evidence of colloquialization. The evidence comes from the TIME Magazine Corpus (Davies 2007) containing all content from the online archive of TIME magazine. Spanning nine decades (1920s-2000s) the data in the corpus invite analysis of diachronic change in written American English throughout the 20<sup>th</sup> century and the very early 21<sup>st</sup> century. The focus is on high-frequency types of what Biber et al. (1999) term 'inserts', marginal items many of which are not 'proper' words which nonetheless "play an important role in communication" (Biber et al. 1999: 56), for example, as discourse markers (e.g., *well, oh*, etc.), backchannels (*yeah, uh-huh*, etc.), and hesitators (*uh, um*, etc.) and which are "especially frequent in spoken texts" (ibid.). The trends observable in TIME are clear-cut: overall, inserts significantly increase in TIME in correlation with the decades, many of them accelerating in the later stages (cf. Westin 2002), thus contributing to "a kind of spontaneous directness which (though often contrived) is clearly supposed to inject into journalistic discourse some of the immediacy of oral communication" (Leech et al. 2009: 239). In the second part I discuss explanatory hypotheses, starting with democratization (Leech 2003, Millar 2009), reduction of male bias (Baker 2009), popularization (Biber 2003), and what could be called dramatization through the use of quotation (Leech et al. 2009). I also propose a new explanatory hypothesis complementing the existing ones: the increasing influence on public writing of "normative conceptions of attitudes and activities" appropriate for women (West & Zimmermann 1987: 127); I refer to this influence as feminization. To support this hypothesis I cite both linguistic and socio-political evidence. The former type of evidence includes recent findings on the rise of gender-neutral *they* at the expense of 'generic *he*' (Leech et al. 2009), as well as research suggesting that inserts (as most other conversational style markers) are significantly more typical of women's than of men's conversational speech style (Rühlemann 2010). Socio-political evidence is derived from demographic statistics documenting the ongoing shift in women's place from the private to the public domain/sphere.

## References

- Baker, P. 2009. 'The BE06 Corpus of British English and recent language change', *IJCL* 14/3: 312-337.
- Biber, D. 2003. 'Compressed noun-phrase structures in newspaper discourse: The competing demands of popularization vs. economy', in J. Aitchison and D. M. Lewis (eds.) *New media language*, London: Routledge, pp.169-181.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson
- Davies, M. 2007. *TIME Magazine Corpus (100 million words, 1920s-2000s)*. Available at: <http://corpus.byu.edu/time>
- Fairclough, N. 1995. *Media discourse*, London: Arnold.
- Leech, G. 2003. 'Modals on the move: The English modal auxiliaries 1961-1991', in R. Fachinetti, F. R. Palmer and M. G. Krug (eds.) *Modality in contemporary English*, Berlin/NY: Mouton de Gruyter, pp. 223-240.
- Leech, G., M. Hundt, C. Mair, and N. Smith. 2009. *Change in contemporary English. A grammatical study*. Cambridge: Cambridge University Press.
- Millar, N. 2009. 'Modal verbs in TIME. Frequency changes 1923-2006', *IJCL* 14/2:191-220
- Rühlemann, C. (2010). 'Conversational grammar–feminine grammar? A sociopragmatic corpus study', *Journal of English Linguistics* 38/1: 56-87
- West, Candace & Don H. Zimmermann. 1987. 'Doing gender', *Gender & Society* 1(2): 125-151.
- Westin, I. 2002. *Language change in English newspaper editorials*. Amsterdam/New York: Rodopi

## Referential null subjects from Old to Early Modern English

Kristian Rusten  
University of Bergen, Norway

The occurrence of referential null subjects in numerous early English texts has been noted by philologists and historical linguists alike. This occurrence – remarkable since Present-day English demands overtly expressed subjects in most finite environments – still represents an understudied area of early English syntax. While recent quantitative work has begun charting the occurrence of such subjects at the Old English stage of the language (cf. Rusten 2010, 2013; Walkden 2012, 2013; Coppess & Pires 2013), scant quantitative attention has thus far been paid to their long-term diachrony. Even so, van Gelderen's (2000: 137, 147) influential account argues that referential null subjects are not restricted to Old English, but that 'pro drop continues to occur' in Middle and Early Modern English and that the phenomenon only 'dies out in the 17<sup>th</sup> century'. This claim finds some support in the traditional literature. For instance, Mustanoja (1960: 138) states that null subjects of all persons and numbers are 'quite frequent' in Old and Middle English, and that practically all 'types of non-expression of the subject-pronoun' demonstrated in Old English are also found in Middle English. Additionally, Ohlander (1981: 37) argues that certain types of subject omission are actually 'more common' in Middle than in Old English. Thus, previous researchers have implied that referential null subjects constitute a relatively frequently occurring syntactic phenomenon in early English. This paper presents a corpus-based quantitative study which seeks to test the validity of such claims. Drawing on the YCOE (Taylor et al. 2003), the YCOEP (Pintzuk & Plug 2001), the PPCME2 (Kroch & Taylor 2000) and the PPEME (Kroch et al. 2004) – four corpora of historical English totalling c. 4.5 million words and covering c. 800 years of language development – it will be shown that referential null subjects are actually exceedingly rare at all stages of English. With the exception of certain early texts belonging to specific genres (notably poetry and medical handbooks), frequencies for such subjects are negligible in all periods. While statistical tests show that the difference between the periods is statistically significant – thus potentially indicating a process of meaningful diachronic change – effect size is consistently low. Hence, it will be suggested that the grammatical status of referential null subjects has not undergone notable change in the span of c. 800 years covered by the analysed texts, and that it is thus doubtful whether such subjects can be considered a productive grammatical feature at any attested stage of English. It will tentatively be argued that the pro-drop phenomenon thus must have disappeared from the grammar of English at a stage prior to the composition and/or transmission of the earliest extant texts. This constitutes a clear contrast to the situation in other early Germanic languages, such as Old High German (Axel 2007), Old Icelandic (Walkden 2012) and Old Norwegian (Kinn 2013), where referential null subjects are much more widely used.

## References

- Axel, Katrin. 2007. *Studies on Old High German syntax: Left sentence periphery, verb placement and verb-second*. Amsterdam/Philadelphia: John Benjamins.
- Coppess, Emily & Acrisio Pires. 2013. The residue of syntactic change: Partial pro-drop in Old English. Talk given at the *LSA Summer Institute – Ann Arbor Diachronic Syntax Workshop*.
- Gelderen, Elly van. 2000. *A history of English reflexive pronouns. Person, self and interpretability*. Amsterdam: John Benjamins.
- Kinn, Kari. 2013. Null arguments in Old Norwegian in an early Germanic perspective. Talk given at the *21st International Conference on Historical Linguistics (ICHL)*.
- Kroch, Anthony, Beatrice Santorini & Ariel Diertani. 2004. *Penn-Helsinki parsed corpus of early Modern English*. <http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-2/index.html>.
- Kroch, Anthony & Ann Taylor. 2000. *Penn-Helsinki parsed corpus of Middle English*. 2nd ed. <http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-3/index.html>.
- Mustanoja, Tauno F. 1960. *A Middle English syntax*. Part 1. Helsinki: Société Néophilologique.
- Ohlander, Urban. 1981. Notes on the non-expression of the subject-pronoun in Middle English. *Studia Neophilologica* 53: 37–49.
- Pintzuk, Susan & Leendert Plug. 2001. *The York-Helsinki Parsed Corpus of Old English Poetry*. <http://www-users.york.ac.uk/~lang18/pcorpus.html>.
- Rusten, Kristian A. 2010. *A study of empty referential pronominal subjects in Old English prose*. MA thesis, University of Bergen.

## Noun phrase complexity across varieties of English: focus on syntactic function and text type

Marco Schilk, Steffen Schaub

University of Hildesheim, Germany / Philipps University Marburg, Germany

The noun phrase (NP) is one of the areas of interest for the study of regional variation in English and has been at the heart of several World English studies. Two frequently investigated features in this respect are adaptations made to the article system (cf. Platt, Weber & Ho 1984; Sand 2004, *forthc.*; Wahid 2013), as well as irregular noun pluralization (cf. Platt, Weber & Ho 1984; Sand *forthc.*; Hall, Schmidtke & Vickers 2013). Typical explanations for such variety-specific uses include substrate influence, grammatical simplification and universal tendencies. However, apart from this focus on isolated features, the structure of the NP remains largely unexplored in World English studies. In particular, there is little descriptive and comparative work on NP complexity. NP complexity, that is, the presence or absence as well as the amount and depth of pre- and postmodification, has been demonstrated to vary considerably depending on factors such as subjectivity (cf. Aarts 1971) or text type (cf. de Haan 1993, Jucker 1992). Given the potential for complexity inherent in the English NP, and considering the even greater range of typological diversity coming from the various substrate languages, it seems reasonable to hypothesize that NP complexity should also vary across varieties of English. The present paper explores the possibility of NP complexity being a potential characteristic feature of varieties of English. The study is based on a sample of 8000 NPs extracted from five different subcorpora of the International Corpus of English (ICE-Canada, ICE-HK, ICE-India, ICE-Jamaica, and ICE-Sin). The data is stratified according to text type, based on four different text type categories (academic written, conversation, social letters, and public speeches). NPs are extracted from the raw data and manually annotated for complexity, syntactic function, text type and variety. The study treats NP complexity as a categorical response variable, differentiating between (1) *head only*, (2) *premodification + head*, (3) *head + postmodification* and (4) *premodification + head + postmodification*. Possible explanatory variables are *variety*, *text type* and *syntactic function*. The response as well the potential explanatory variables are categorical, therefore the analysis is based on a logistic regression approach. Our findings indicate that, although all explanatory variables have significant influence on the response, variety is a relatively weak predictor of structural variation, which is in line with earlier studies (which, however, focused on the verb phrase (cf. e.g. Bernaisch et al. *fc.*)). More specifically, there is substantial support for the growing conviction that text type has a far greater influence on surface form than variety-specific preferences. Additionally, syntactic function appears to be a strong predictor of NP complexity. Across all text types and varieties, NPs with subject function are more likely to be simple, although there is a general simple-complex cline between the conceptually spoken text types and academic written texts. Moreover, NP complexity displays a greater amount of regional variation in some of the text types (e.g. academic writing) than in others (e.g. social letters). Finally, results show that Canadian and Singapore English are largely homogeneous in their propensity for NP complexity across the whole spectrum of text types. Hong Kong English and Jamaican English also behave relatively homogeneously (albeit less so than Canadian and Singapore English), while Indian English differs most strongly from the other varieties. Overall, the results confirm the importance of text type and syntactic function as predictors of NP complexity. Furthermore, variety is shown to be a relatively weak predictor when text type differentiation

## References

- Aarts, F. G. A. M. 1971. On the distribution of noun-phrase types in English clause-structure. *Lingua* 26: 281–293.
- Bernaisch, Tobias, Stefan Th. Gries & Joybrato Mukherjee. *forthcoming*. The dative alternation in South Asian English(es): modelling predictors and predicting prototypes. *English World-Wide* 35(1).
- de Haan, Pieter. 1993. Noun phrase structure as an indication of text variety. In Andreas H. Jucker (ed.): *The noun phrase in English. Its structure and variability*. Heidelberg: Winter. 85–106.
- Hall, Christopher J.; Daniel Schmidtke & Jamie Vickers. 2013. Countability in world Englishes. *World Englishes* 32(1): 1–22.
- Jucker, Andreas H. 1992. *Social stylistics. Syntactic variation in British newspapers*. Berlin, New York: Mouton de Gruyter.
- Platt, John, Heidi Weber & Mian Lian Ho. 1984. *The New Englishes*. London: Routledge and Kegan Paul.
- Sand, Andrea. 2004. Shared morpho-syntactic features in contact varieties of English: article use. *World Englishes* 23(2): 281–298.
- Sand, Andrea. *forthcoming*. *Angloversals? Shared morpho-syntactic features in contact varieties of English*. Amsterdam: Benjamins.
- Wahid, Ridwan. 2013. Definite article usage across varieties of English. *World Englishes* 32(1): 23–41.

### **Animacy versus complexity? A corpus-based longitudinal investigation of the competition between possessive *whose* and *of which***

Edgar W. Schneider

University of Regensburg, Germany

The choice between the possessive relativizers *whose* and *of which* represents free variation in some contexts (*the house whose roof leaked* / *the house the roof of which leaked*) but not in others (*the man whose car exploded* but not *\*the man the car of which ...*). Little attention has been paid to the constraints which govern this competition. The present paper investigates these constraints from late Middle English, when this paradigmatic relationship emerged, to the present day, when we find variable preferences and constraints in different varieties. The diachronic trajectory builds upon data from the Helsinki corpus for Middle English and Early Modern English and from Archer for the following centuries, backed by some data from the OED quotations database; recent changes and the principles governing the present-day distribution of both forms in the global lead varieties of English are described using the "Brown quartett" and the British and American ICE corpora; and data from ICEIndia and ICE-Philippines provide a glimpse of how New Englishes have handled the complexity involved. Quantitative distributions and representative examples are provided throughout. A particularly vibrant and complex developmental trajectory established the basic parameters governing the choice between both forms up to the Early Modern English period. Relativizing *whose*, derived from its earlier interrogative function, was reasonably well established by late Middle English. The form *of which* first occurred in accidental chance cooccurrences of the preposition *of* and the pronoun *which* and then in the cognitively simpler partitive relationship before becoming a possessive relativizer. It can be shown how it adopted the defining characteristics of a possessive relativizer, expressing definiteness, possession and co-reference, only gradually (with these functions often marked explicitly in addition to *of which* during the transition period). Obviously it was motivated by a well-known overarching process of change of that period, that of distinguishing animate from non-animate referents in pronominalization, with animate antecedents of *of which* decreasing and ultimately disappearing. Interestingly enough, throughout the Early Modern English period in that function it was faced with the competition of another option, *whereof*, which, however, ultimately disappeared. Syntactically, *of which* originally appeared in the same pre-nominal position as *whose*, but frequency changes throughout the period clearly show that it inherited its current preferred post-nominal position from *whereof*. Recent changes and distributions in today's varieties suggest an ongoing development in which animacy effects, repeatedly observed in noun genitive constructions in recent research, compete with complexity constraints. This conflict has not yet been fully resolved, as two observable trends indicate:

(1) while *of which* has been restricted to inanimate referents, *whose* is preferred with animate referents but allows inanimate ones as well (at varying proportions); (2) the semantic complexity of possessive relativization and the structural awkwardness of postnominal *of which* appear to have contributed to an overall decrease of both forms, again to varying degrees between varieties.

## Mass Nouns in English: Automatic Detection and Regional Variation

Gerold Schneider, Daniel Schreier  
University of Zurich, Switzerland

For this study we extract mass nouns from large automatically parsed English corpora in a data-driven fashion, and we compare their regional variation. In the first step, we detect mass nouns semi-automatically by comparing and summing over pre- and postmodifying contexts using collocation statistics in the British National Corpus BNC (Aston & Burnard 1998). In order to extract those we use T-score (e.g. Evert 2009) and O/E collocation statistics. The automatic extraction includes comparing and combining features such as

1. typical determiners ('some milk'),
2. mass quantifiers ('a glass of milk'),
3. high frequency of zero determiners ('I like  $\emptyset$  milk'), 4. Rareness of plural use ('? She drank several milks').

In a second step, we measure the use of the mass nouns thus extracted across varieties in the International Corpus of English ICE (Nelson 2002), in particular how often certain varieties use mass nouns in the plural, as in sentence (1)

(1) I can send you the *informations* (ICE---HKW1B---008)

The use of this "innovation in pluralisation" (Mesthrie & Bhatt 2008:53) has been described for a range of outer circle varieties, including Indian English (Sharma 2013:525), Hong Kong English (Wong 2013:552), Jamaican English (Sand 2013:212). We include these outer circle ICE corpora (India, Hong Kong, Jamaica) as well as ICE-Philippines and ICE-Singapore in our investigation, comparing to the inner circle varieties ICE-GB, ICE-Ireland, ICE-Canada and ICE-NZ. We discuss that many mass nouns have word senses in which plural forms are possible (e.g. *controls*, *woods*), that the 'type of' conversion is highly productive (e.g.

*wines* = types or bottles of wine), and that also inner circle varieties show some unexpected uses of mass nouns plurals, like sentence (2).

(2) This one's on and the telescope on the right is the William Herschel four point two metre and that's the telescope used to obtain these *informations* (ICE-GB S2A-058)

To conclude the automatic extraction of mass nouns, we conduct an evaluation delivering a list of over 300 validated mass nouns from the BNC. Our evaluation shows that with a combination of the above features we reach between 90% precision for the top 50 and almost 70% for the top 400 candidates. Our large list of validated mass nouns allows us to confirm the prototypical cases of outer circle variety uses described previously (*staffs*, *advices*, *furnitures*, *behaviours*, *equipments*, ...), but also to go beyond it, and obtain a higher coverage, as far as the relatively sparse data from the 1 million word ICE corpora permits. Nevertheless, there are clear quantitative differences, showing that the readiness to use mass noun plural forms is elevated in outer circle varieties, that there are differences between various outer circle varieties and different genres, and that the use of mass nouns belongs to the class of subtle lexico-grammatical differences (Schneider 2004) which are in flux in all English varieties and which, according to Pawley and Syder (1983) can mark the distinction between an advanced and a native speaker of English, and that such subtle differences profit from quantitative investigation.

## References

- Aston, Guy and Lou Burnard (1998). *The BNC Handbook. Exploring the British National Corpus with SARA*. Edinburgh University Press, Edinburgh.
- Evert, Stefan (2009). "Corpora and collocations". In Anke Lüdeling and Merja Kytö, (eds.) *Corpus Linguistics. An International Handbook*, article 58. Berlin: Mouton de Gruyter.
- Kortmann, Bernd and Kerstin Lunkenheimer, eds. (2013). *The Electronic World Atlas of Varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Mesthrie, Rajend and Rakesh M. Bhatt (2008). *World Englishes: the study of new linguistic varieties*, Cambridge: Cambridge University Press.
- Nelson, Gerald, Sean Wallis, and Bas Aarts (2002). *Exploring Natural Language: Working with the British Component of the International Corpus of English. Varieties of English Around the World: G29*. John Benjamins, Amsterdam.
- Pawley, Andrew and Frances Hodgetts Syder (1983). "Two puzzles for linguistic theory: Native-like selection and native-like fluency". In J. C. Richards and R.W. Schmidt, editors, *Language and Communication*. Longman, London.
- Sand, Andrea (2013). "Jamaican English". In Kortmann & Lunkenheimer (2013).
- Schneider, Edgar W. (2004). "How to trace structural nativization: particle verbs in world Englishes". *World Englishes* 23(2): 227–249.
- Sharma, Devyani (2013). "Indian English". In Kortmann & Lunkenheimer (2013).
- Wong May L.---Y. (2013). "Hong Kong English". In Kortmann & Lunkenheimer (2013).

## Diachronic developments of the concessive markers *notwithstanding*, *in spite of* and *despite* in written American English

Ole Schützler

University of Bamberg, Germany

Concessive adverbials have been discussed in terms of their origin and semantics (e.g. König 1985) and the ways in which they can be encoded linguistically (e.g. Quirk et al. 1985: 1097-1103). However, relatively little is known about the actual use and the relative frequencies of certain concessive markers in varieties of English, their diachronic development, and the frequency of particular construction types in which different markers are used. Using the *Corpus of Historical American English* (COHA; Davies 2010), this paper mainly inspects the diachronic development of *notwithstanding* from the early 19th century to the present day, partly building upon work by Rissanen (2002) and Berlage (2009). Results show that *notwithstanding* has decreased considerably in frequency, reaching a relatively stable low level in the 1930s. This affects in parallel fashion the use of *notwithstanding* as a preposition (1), as a conjunction followed by a finite clause with or without *that* (2), the specific constructions *notwithstanding the fact that* (3) and *notwithstanding this* (4), as well as the use of *notwithstanding* as a main clause conjunct (5) (examples from COHA):

- (1) **Notwithstanding** the excessive rates now imposed, we do continue importing [...].
- (2) **Notwithstanding** (that) Ellis had taken no dinner, he had little appetite [...].
- (3) [...] and the cars still suffered great delay at a late hour last night **notwithstanding** the fact that the storm was over.
- (4) Most clergymen fail to comply with that requirement, and yet, **notwithstanding** this, you contribute to their support, and go to hear them preach, [...].
- (5) [...], since the total sum available therefrom was only \$107,000. **Notwithstanding**, the land grant was a big help in selling the company's first mortgage bonds [...].

The only construction type to have resisted the general trend of declining frequency involves *notwithstanding* as a postmodifier (6), which, as will be shown, has even increased in frequency:

- (6) But this affection **notwithstanding**, a curious respect for her existed.

The paper argues that the exceptional behaviour of this subtype can be accounted for in terms of construction grammar (cf. Hoffmann & Trousdale 2011). The construction NP + *notwithstanding* not only defies the general decline of *notwithstanding*, but additionally the postmodified noun phrases are characterised by decreasing complexity, diachronically. The realisation of this particular construction type becomes more restricted over time, which can be interpreted as a reduction in schematicity. In addition to *notwithstanding*, the concessive markers *in spite of* and *despite* will be discussed. General developments in frequency suggest that there is competition between these two markers during the 19th century, with both of them increasing in frequency until the 1910s. From the early 20th century onwards, it is *despite* which becomes dominant. Once again, it can be shown that in addition to overt changes in the frequencies of the two markers, constructional change plays a role, too – in this case the competition between noun phrase complements (e.g. *despite / in spite of his ambitions*) and gerundial clauses (e.g. *despite / in spite of being ambitious*).

### References

- Berlage, Eva. 2009. Prepositions and postpositions. In Günter Rohdenburg & Julia Schlüter (eds.) *One Language, Two Grammars? Differences between British and American English*. Cambridge: Cambridge University Press. 130-148.
- Davies, Mark. 2010. *The Corpus of Historical American English: 400 million words, 1810-2009*. Available online at <http://corpus.byu.edu/coha/>.
- Hoffmann, Thomas & Graeme Trousdale. 2011. Variation, Change and Constructions in English. *Cognitive Linguistics* 22 (1): 1-23.
- König, Ekkehard. 1985. Where do concessives come from? On the development of concessive connectives. In Jacek Fisiak (ed.) *Historical Semantics. Historical Word-formation*. Berlin: Mouton de Gruyter. 263-282.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Arnold.
- Rissanen, Matti. 2002. 'Despite or notwithstanding? On the development of concessive prepositions in English.' In Andreas Fischer, Gunnel Tottie and Hans Martin Lehmann (eds.) *Text Types and Corpora. Studies in Honour of Udo Fries*. Tübingen: Gunter Narr Verlag. 191-203.



## Colloquialization and a century in the life of vague quantifiers

Nicholas Smith, Amy Wang

Leicester University, United Kingdom / Manchester Metropolitan University, United Kingdom

This study explores the connections between two phenomena of growing interest to researchers, namely, vague language and colloquialization. Vague language (e.g. *thing*, *something like*) is considered especially frequent in spoken discourse (Channell 1994, Carter and McCarthy 1997, Biber et al. 1999). Colloquialization is the tendency of written (printed) genres of recent English to become more speech-like in their style (Hundt and Mair 1999, Leech et al. 2009). Our particular focus is on the use of vague quantifiers (VQs), such as *a lot (of)*, *a bit (of)* and *many*, across a century of published writing in British English. Our overarching research question is: To what extent, and in what ways, have VQs (individually and collectively) changed in frequency and use during this time? One might predict the answer to be a straightforward convergence of the two tendencies given above, i.e. that VQs have grown in use in writing as a direct result of increasing colloquialization. However, individual VQs in fact vary considerably in their degree of colloquialness, as reflected by their distribution across corpus registers and their collocational preferences, for example. Historically, moreover, VQs differ in the extent to which they have been prescribed, proscribed or omitted in usage guides over the last century. Thus additional research questions are warranted:

1. To what extent can we position the different VQs on a scale of colloquial/non-colloquial? Has there been any discernible change in the functions of the quantifiers? To what degree can this change account for colloquialization?
2. Is there a correlation between degree of colloquialness of a given quantifier and extent of infiltration of that quantifier into printed written texts?
3. Have the pragmatic functions and collocations of individual vague quantifiers changed over time?

Our paper seeks to provide an important methodological contribution to discussion on colloquialization and recent language change. First, in order to “calibrate” the degree of colloquialness of individual VQs in standard English, we use two methods: i) gauge their frequency in different registers of reference corpora (BNC and SEU/ICEGB), and ii) consult commentaries on the VQs in 20th-century usage guides (e.g. the three editions of Fowler’s *Modern English Usage*) and grammars. To address the second and third research questions, we use the British components of the Brown family of written corpora, spanning 1901 to 2006, to determine frequency changes, and probe individual quantifiers both quantitatively and qualitatively (using case studies) in terms of their pragmatic functions and collocations. Results so far demonstrate interesting variation across the individual VQs and mixed support for the colloquialization hypothesis. For example, the determiner *a lot of* increases dramatically from the 1960s onwards, whereas the pronominal (and even more colloquial) use *a lot* peaks in the 1960s. A decline of *plenty (of)* suggests it is not as colloquial a quantifier as some sources (e.g. Quirk et al. 1985, Biber et al. 1999: 276) suggest, and moreover its collocational profile is rather specialized. The paucal quantifier *a bit (of)* grows significantly between the 1960s and 1990s, boosted by its increasing use as a device of mitigation, which is a common strategy in ordinary conversation.

## References

- Channell, J. 1994. *Vague Language*. Oxford: Oxford University Press.
- Hinrichs, L. and B. Szmrecsanyi. 2007. “Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora”. *English Language and Linguistics* 11(3): 437–474.
- Hundt, M. and C. Mair. 1999. “Agile” and “uptight” genres: the corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2): 221–242.
- Leech, G., M. Hundt, C. Mair and N. Smith. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.
- Wang, A. 2007. *Vagueness in Interaction: A Cross-Cultural Examination of Vagueness in Taiwanese and British Courtroom Discourse*. Unpublished PhD thesis, Lancaster University.

## On the ontological status of the 'Great Vowel Shift'

Gjertrud Flermoen Stenbrenden  
University of Oslo, Norway

The 'Great Vowel Shift' is the label given to a series of changes to the long monophthongs of Middle English (ME), whereby the high vowels diphthongised, and the other vowels were raised. Jespersen (1909) and Luick (1899, 1912, 1914-40) were the first to coin this name, no doubt as a mental shorthand to cope with a complex series of changes. However, as often happens when phenomena receive capitalised labels, the Shift was soon reified and treated as a *noumenon*, a unitary event with a definite beginning and end, which required its own special explanation. Jespersen and Luick famously disagreed whether it was the high or the high-mid vowels which changed first and set the Shift in motion, giving rise to 'drag-chain' and 'push-chain' models of the Shift respectively. More ink has been spent on defending and upholding these ideas of unitariness and chain-shift models than on investigating the evidence for the Shift, or the chronology and loci of the changes. Part of the purpose of this paper is therefore to present a large number of innovative ME spellings for the long monophthongs, with a view to determining these questions. The material is extracted from three corpora of ME localised texts (*LAEME*, *SMED*, *LALME*). The first stages of the Shift are now generally set at about 1400-1450 AD, and it is thought to have reached completion in the mid-eighteenth century. Thus, the Shift spanned 250 years, even by conservative estimates. Given the dating of the first stages, the Shift is also treated as something different from long-vowel changes which took place in late Old English (OE) and early ME. The earlier changes comprise the unrounding of the reflex of OE  $\bar{y}$ , the monophthongisation and unrounding of OE  $\bar{e}o$ , the fronting and raising of OE  $\bar{o}$  in Northern ME, and the backing, rounding and raising of OE  $\bar{a}$  in Southern ME. However, recent research using material from *LAEME*, *SMED*, and *LALME* has shown that the early stages of the Shift must have begun in the mid-thirteenth century (Stenbrenden 2010), supporting the 'early vowel-shift hypothesis' (Stockwell 1960, 1972, 1978, 1985, 2006; Stockwell and Minkova 1988a, 1988b). That is, the initial stages of the 'GVS' overlap temporally with the presumed earlier changes. Hence, if there was any such thing as the 'Great Vowel Shift', it must include the late OE and early ME changes as well, since they were going on at the same time. These facts have far-reaching consequences for our interpretation of the changes in question. Is it really possible for changes which span 700-850 years to be part of one unitary, coherent Shift? And whatever the initial changes were, once they had happened, did the rest of the vowels have to change as well, as the 'push-chain' and 'drag-chain' models presuppose? Are there more realistic and phonetically-based ways of accounting for such recurring changes in the Germanic languages? The second purpose of this paper is to attempt to answer these questions.

## References

- Jespersen, O. 1909 [4<sup>th</sup> ed. 1928]. *A Modern English Grammar, Part I*. Heidelberg: Carl Winters Universitätsbuchhandlung.
- LALME*: see McIntosh, Samuels, Benskin *et al.* 1986
- Luick, K. 1899. 'Über die Diphthongierung van me.  $\bar{u}$ ,  $\bar{i}$  und verwandte deutsche Erscheinungen'. *Archiv für das Studium der neueren Sprachen und Litteraturen* 53/103: pp. 267-277.
- Luick, K. 1912. 'Über die neuenglische Vokalverschiebung'. *Englische Studien* 45: pp. 432-437.
- Luick, K. 1914-1940. *Historische Grammatik der englischen Sprache, Vol. I, Parts 1 & 2*. Oxford: Basil Blackwell.
- SMED*: see Kristensson 1967, 1987, 1995, 2001, 2002
- Stenbrenden, G.F. 2010. *The Chronology and Regional Spread of Long-Vowel Changes in English, c. 1150-1500*. Ph.D. dissertation, University of Oslo.
- Stockwell, R.P. 1960. 'The Middle English "Long Close" and "Long Open" Mid Vowels'. In: *University of Texas Studies in Literature and Language* 2; pp. 529-538.
- Stockwell, R.P. 1972. 'Problems in the Interpretation of the Great English Vowel Shift'. In: Smith, M.E. (ed.) 1972. *Studies in Linguistics in Honor of George L. Trager*. The Hague: Mouton; pp. 344-362.
- Stockwell, R.P. 1978. 'Perseverance in the English vowel shift'. In: Fisiak, J. (ed.) 1978. *Recent Developments in Historical Phonology*. The Hague: Mouton; pp. 337-348.
- Stockwell, R.P. 1985. 'Assessment of alternative explanations of the Middle English phenomenon of high vowel lowering when lengthened in the open syllable'. In: Eaton, R. *et al.* (eds.). *Papers from the 4<sup>th</sup> International Conference on English Historical Linguistics*. Amsterdam, Philadelphia: John Benjamins ;pp. 303-318.
- Stockwell, R.P. 2006. 'The Status of Late Middle English <e> Spellings as Early Evidence of the English Vowel Shift'. In: Schaefer, U. (ed.). 2006. *The Beginnings of Standardization: Language and Culture in Fourteenth-Century England*. Frankfurt am Main: Peter Lang; pp. 175-180.
- Stockwell, R.P. & D. Minkova. 1988a. 'The English Vowel Shift: problems of coherence and explanation'. In: Kastovsky and Bauer 1988; pp. 355-394.

### Medical case reports in Late Modern English

Irma Taavitsainen, Anu Lehto  
University of Helsinki, Finland

The medical case report is a narrative of a single case of disease or injury and it is one of the genres that have continuity throughout the history of English medical writing from the late medieval period to the present. However, its functions and linguistic realisations vary in different periods: in modern medical writing case reports often focus on unusual cases, whereas the genre had a central position in medical teaching in the early periods (see Taavitsainen 2011). The eighteenth century represents a transition from the earlier thought-styles to more modern approaches to medicine with the first statistical assessments towards the end of the period. The study traces how these developments in medical thinking are reflected in eighteenth-century case reports in a large, unexplored database; case reports of the period have been earlier studied only in the *Edinburgh Medical Journal* (Atkinson 1992). Our aim is to chart the uses of medical case reports in the corpus of Late Modern English Medical Texts 1700–1800 (LMEMT, forthcoming) under work by the Scientific Thought-styles team LMEMT contains some two million words and reflects an inclusive view of medicine with a full scale of texts from academic treatises to writings targeted at heterogeneous lay people. We shall first place the genre of case studies in context with different layers of medical writing by charting the occurrences of case reports in various longer texts and fields of medical writing. Case studies are frequent in LMEMT and they are found in treatises on specific topics, surgical texts, institutional texts and in periodicals. In the earlier periods case reports were often embedded in other genres, and there is a great deal of variation. The paper analyses the various functions of case reports. Instruction is the most common purpose, but new functions emerge, as the case reports introduce new methods of treatment and cures, and patients begin to write down their own experiences as patients. We shall further analyse case reports paying special attention to the degree of conventionalisation and the perspective through which the narrative is told (Simpson 1993). Involvement and personal affect features are of special interest: the eighteenth-century case reports vary in their degree of authorial involvement and affect from highly personal accounts to more detached styles of writing. These differences can also be related to genres, e.g. the more involved style of writing is favoured in texts for lay audiences, as authors often write in the first person (*I perceived...*) and use more personal pronouns in general. The detached style of writing becomes more common towards the end of the eighteenth century along with the changes in medical thinking, and authors use the passive voice more (*The frictions were successfully continued...*) and stock noun phrases become more frequent (*The dysenteric symptoms disappeared...*). We shall employ the Key Word method and assess n-grams (WordSmith) to discover even more subtle differences between the texts of different genres for various target groups.

### References

- Atkinson, Dwight. 1992. The evolution of medical research writing 1735–1985: The case of the *Edinburgh Medical Journal*. *Applied Linguistics*, 13.4., 337–374.
- Taavitsainen, Irma. 2011. Medical Case Reports and Scientific Thought-styles. *Revista de Linguas para Fines Específicos* (special issue, ed. by Alonso Almeida, Francisco), 75–98.
- Taavitsainen, Irma and Päivi Pahta. 2000. Conventions of Professional Writing: The Modern Medical Case Report in a Historical Perspective. *Journal of English Linguistics*, 28 (1), 60–76.
- Simpson, P. 1993. *Language, Ideology and Point of View*. London: Routledge and Kegan Paul.

### 'Fancy meeting up for a drink?': The role of metonymy in a corpus of British text messages

Caroline Tagg, Jeanette Littlemore  
University of Birmingham, United Kingdom

This talk reports on an investigation into metonymy in a corpus of British text messages. Metonymy is a figure of speech by which we talk about one thing through reference to another thing associated with it – often a whole will stand for a part, or a part for a whole. The example in the title of this talk exhibits both processes: 'drink' in this context refers specifically to a particular kind of drink (an alcoholic one); while 'meeting up for a drink' stands for a more complex set of cultural activities: socialising, having a chat, catching up. With few exceptions (Deignan et al, 2013), the limitation of metonymy research is that it has not been based on naturally-occurring data, and nor has it taken into account potential genre differences. The present investigation brings together these three elements to explore the role that metonymy plays in one genre, that of text messaging, through corpus data. As the discussion of 'fancy meeting up for a drink' suggests, the social and cultural context in which the text messages were produced is also key to identifying metonymy and explaining its role. Thus, the research questions that this talk addresses are:

1. What does corpus analysis reveal about the extent to which metonymy may be shaped by genre and by social and cultural factors?
2. What does the study of metonymy in naturally-occurring data reveal about text messaging as a form of communication in this cultural context?

In the talk, we start by outlining arguments made in metonymy research (e.g. Radden and Kövecses, 1999; Warren, 2006), including the assumption that metonymy mainly plays a referential role and that it tends not to be extended creatively in the way that metaphor can be. I then go on to describe the data used in this study, CorTxt, a corpus of over 11,000 text messages collected in the first decade of the twenty-first century in Britain, before outlining the methods used in the investigation. We carried out an in-depth analysis of 1000 text messages, identifying and categorising metonymy according to distinctions made in the literature and, importantly, adapting or rejecting categories to fit the data to ensure that our approach was to some extent data-driven. The results show that around 20% of the text messages include examples of metonymy, and that much of this is explained by the recurrence of frequently-occurring and highly conventional metonyms, chiefly variations of 'give us a ring' (effect standing in for cause), 'fancy a drink?' (category for member of a category / sub-event for whole event) and 'hello gorgeous' (trait for person). But the corpus also throws up novel and creative examples which challenge previous assumptions about metonymy; and which characterise text messaging as a complex genre combining both convention and creativity. Finally, the investigation also raises interesting methodological questions about the problems associated with identifying metonymy in attested data; and the extent to which quantitative corpus analysis into metonymy is possible given its varied and unpredictable surface realisations.

### References

- Deignan, A., J. Littlemore and E. Semino (2013) *Figurative Language, Genre and Register*. Cambridge: Cambridge University Press.
- Radden, G. and Kövecses, Z. (1999) 'Towards a theory of metonymy' in Panther, K.-U. and G. Radden (eds) *Metonymy in Language and Thought*. Amsterdam: John Benjamins, pp. 17-59.
- Warren, B. (2006). *Referential Metonymy*, Scripta Minora, 2003-4, Royal Society of Letters at Lund, Sweden.

## Native Combining Forms In English And Their Productivity – A Diachronic Corpus-based Study

Leonie Wiemeyer

University of Bremen, Germany

Combining forms (CFs) are bound lexical elements which are abundant in the English language. Examples of well-established CFs include *eco-*, *geo-*, *-holic*, and *-athon*. In the past, only so-called neo-classical CFs played a role in English word-formation. In the 20<sup>th</sup> century, however, a new class of CFS derived from English words emerged. Native combining forms often occur in jocular formations, predominantly on the Internet. They are used to name new cultural phenomena, for example those brought about by technological progress, and given rise to a considerable amount of neologisms in the past two decades, many of which have become established. Thus far, they have not been given much attention in the linguistic literature, but the large number of elements listed in dictionaries testify that combining forms are not just a fad. This paper presents the results of a corpus-based investigation of the productivity of native combining forms in English. In a diachronic approach, the productivity of 21 CFs was determined for the period from 1950 until 2009 using data from the Corpus of Historical American English (COHA). The aim of the study was to determine whether the native combining forms investigated are currently productive and how this productivity has changed over time. A further aim was to identify factors influencing productivity and to establish a connection between an element's productivity and the "currentness" of the cultural phenomenon it denotes. Each element's productivity was determined by calculating type and token frequencies for each decade using the COHA data and by counting the neologisms per decade. This diachronic approach was chosen because the number of neologisms per decade is an indicator of whether a combining form was and continues to be used in word-formation. Phonological and semantic transparency, the source words and possible restrictions on the bases were considered in order to infer factors which influence productivity. Additionally, semantic and pragmatic traits were investigated to determine the connection between an element's productivity and its meaning as relating to contemporary cultural phenomena. The corpus-based study on the productivity of native combining forms revealed that, generally, CFs can be considered productive in word-formation. All of the elements displayed productivity at some point in the period from 1950 until 2009. However, the investigation also showed that CFs vary in their productivity due to various factors, e.g. semantic transparency. The study found that many combining forms relate to current technology. They are often used in eye-catching terms in the media and online. The productivity of a CF depends on its meaning especially in those cases where it relates to current cultural phenomena such as scientific or technological advances. For example, in the 1990s there was a significant rise in combinations referring to Internet-related phenomena. The productivity of some CFs decreases as soon as their referent goes out of fashion. However, as the corpus study has shown, there are also combining forms that enter the language as fashionable and creative coinages but become established and retain a moderate or even high productivity.

## References

- Baayen, Harald & Lieber, Rochelle (1991). Productivity and English Derivation: A Corpus-Based Study. *Linguistics* 29(5), 801-844.
- Bauer, Laurie (1998). Is there a class of neoclassical compounds in English and is it productive? In: *Linguistics* 36, 403-422.
- Bauer, Laurie (2006a). Compounds and Minor Word-formation Types. In Aarts, Bas & McMahon, April M. S. (Eds.), *The Handbook of English Linguistics* (pp. 483-506). Malden, MA: Blackwell.
- Fischer, Roswitha (1998). *Lexical Change in Present-Day English. A Corpus-Based Study of the Motivation, Institutionalization, and Productivity of Creative Neologisms*. Tübingen: Gunter Narr Verlag.
- Fradin, Bernard (2000). Combining forms, blends and related phenomena. In Doleschal, Ursula & Thornton, A.M. (Eds.), *Extragrammatical and Marginal Morphology* (11-55). Munich: LINCUM.
- Kastovsky, Dieter (2009). Astronaut, astrology, astrophysics: About Combining Forms, Classical Compounds and Affixoids. In McConchie, R. W., A. Honkapohja & J. Tyrkkö (Eds.), *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX) (1-13)*. Somerville, MA: Cascadia Proceedings Project. Retrieved from <http://www.lingref.com/cpp/hel-lex/2008/paper2161.pdf>.
- Lehrer, Adrienne (1998). Scapes, Holics and Thons: The Semantics of English Combining Forms. *American Speech* 73(1), 3-28.
- Lehrer, Adrienne (2003). Understanding trendy neologisms. *Rivista di Linguistica* 15(2), 371-384. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.171.676&rep=rep1&type=pdf>.
- Lehrer, Adrienne (2007). Blendalicious. In Munat, Judith (Ed.), *Lexical Creativity, Texts and Contexts*. Amsterdam: John Benjamins.

## From phrase to clause(-like): on present participle and verbal noun in Middle Scots

Eva Zehentner

University of Vienna, Austria

This paper investigates the development of present participles and verbal nouns in the history of English. More specifically, the paper addresses the question of whether there is a correlation, or even a causal relationship, between the two main phenomena which can be identified in their development (cf. de Smet 2010; Kisbye 1972; Swan 2003): First, the phenomenon of non-finite phrases developing into clause-like constructions. Prime examples of such 'clausalisation' are the gradual dissociation of the present participle from its adjectival origins (Swan 2003), as well as the acquisition of more and more verbal features by the originally purely nominal verbal nouns in *-ing* from the early Middle English period onwards (cf. de Smet 2010; Fanego 2004; Jack 1988; Tajima 1985). The development is clearly visible in the differences between the following instances of the forms in question (adjectival vs. adverbial modification, *of*-object vs. direct object, etc.):

(1)

a) and án **scinende** culfre . scaet of þam fyre 'and a shining dove darted out of the fire' (*Ælfric Live of Saint Basil*, 10<sup>th</sup> ct.)

b) quhen he is **wakand** and **trauailand** 'when he is awake and working' (*Ratis raving*, 1450)

(2)

a) At this horrible **murthering** of trew Christians 'at this horrible murdering of true Christians' (*Diary Robert Birrel, 1532-1605*)

b) in **telling** ther jvdgments frielie 'in voluntarily telling their judgements' (*Official correspondence, John First Earl of Middleton to Sir George Mackenzie Lord Tarbat, 1662*).

Second, the paper is concerned with the formal as well as functional merger of present participles and verbal nouns, a development which is assumed to have started around 1200 in the South of England (Lass 1992: 146). This coalescence then resulted in the Present Day English situation where rather than two distinct categories with two forms (*-nd*-suffix vs. *-ing*-suffix), there is now a 'gradient' category with one single form (*-ing*) (cf. Quirk & Greenbaum 1985: 1290). Methodologically, the paper is based on a quantitative study of over 15,000 tokens of non-finites in the *Helsinki Corpus of Older Scots* (HCOS). The decision to focus on this descendant of Old English was driven by the fact that the formal collapse of *-ing*-forms and present participles was delayed until the early 15<sup>th</sup> ct. there (Gardela 2011; Dons & Moessner 1999), but both forms nevertheless developed clausal behaviour from early on. While the Scots evidence thus seems to support a considerable causal impact of the clausalisation of the forms on their merger, it will be shown that their further development was in turn heavily influenced by their merging - confirming to a certain degree Demske's (2002: 88) claim that "[i]t is obviously due to this change [i.e. the merger] that *ing*-nominals are prevented from [the] gradual shift in terms of nominalization" seen in the history of German.

## References

- Demske, Ulrike. 2002. Nominalization and argument structure in Early New High German. *ZAS Papers in Linguistics* 27, 67-90.
- De Smet, Hendrik. 2010. English *-ing*-clauses and their problems: the structure of grammatical categories. *Linguistics* 48(6), 1153-1193.
- Dons, Ute; Moessner, Lilo. 1999. The present participle in Middle Scots. *Scottish Language* 18, 17-33.
- Fanego, Teresa. 2004. On reanalysis and actualization in syntactic change: The rise and development of English verbal gerunds. *Diachronica* 21, 5-55.
- Gardela, Wojciech. 2011. Spelling variants of the present participle in a selection of Northern English and Scots texts of the late 14th and the 15th centuries. In Gregersen, Frans (ed.). *Language variation - European perspectives III: selected papers from the 5th International Conference on Language Variation in Europe (ICLaVE 5), Copenhagen, June 2009*. Amsterdam: Benjamins, 201-214.
- Jack, George. 1988. The origins of the English gerund. *NOWELE* 12, 15-75.
- Kisbye, T. 1971. *An historical outline of English syntax. Part I*. Aarhus: Akademisk Boghandel.
- Lass, Roger. 1992. Phonology and Morphology. In Norman Blake (ed.). *The Cambridge history of the English language. Vol. 2. 1066-1476*. Cambridge: CUP, 23-155.
- Quirk, Randolph; Greenbaum, Sidney. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Swan, Toril. 2003. Present participles in the history of English and Norwegian. *Neuphilologische Mitteilungen* 104, 179-195.
- Tajima, Matsuji. 1985. *The syntactic development of the gerund in Middle English*. Tokyo: Nan'un-do.

### The motivating factors for the long passives in academic and non-academic writing

Cathy Guiping Zhang

*The Chinese University of Hong Kong, China*

Over the past decades, we have been discouraged from using the passive voice by various manuals and guides for academic writing (e.g. Strunk & White, 2009; *The Publication Manual of American Psychological Association*, 2010). Meanwhile, however, there has been well-documented preference for the passive by academic writing (e.g. Biber, Johansson, Leech, Conrad & Finegan, 1999; Biber & Gray, 2010; Quirk, Greenbaum, Leech & Svartvik, 1985; Svartvik, 1966; Wanner, 2009). To resolve the possible confusion caused by this, I conducted a corpus-based comparative study of the motivating factors for long passives in academic and non-academic writing by comparing the animacy, heaviness/length, and givenness of the subjects and agents of the long passives in the two registers. This study investigates the extent to which the long passives in academic and non-academic writing are motivated by End Weight, End Focus and Animated First principles. I used the academic and non-academic writing sub-corpora of ICE-GB, the British component of the International Corpus of English to obtain the 370 long passives in academic writing and the 308 in non-academic writing for the study. For the animacy analysis, I adopted the human sub-categorization of the General Animacy Scale developed by Yamamoto (1999). The length of the subjects and agents in the two registers was compared using SPSS 19.0. The givenness of the subjects and agents was compared based on their referential distance, which measures the number of clauses separating the subjects and agents from the last occurrence of their referents in the preceding text (Givón, 1983; 1994). The results show that animate subjects are rarely used in academic writing (6.22%), that the animacy of the subjects is lower than that of the agents (31.62%), and that the animacy of the subjects and agents of the long passives in academic writing are lower than those in non-academic writing. This indicates that the concern for putting the more animated entities at the beginning of the sentences is not a motivating factor for using long passives in academic writing and that academic writing is more impersonal than non-academic writing. It is also found that most of the agents of the long passives in both registers are longer than the subjects and that the subjects and agents in academic writing are respectively longer than the subjects and agents in non-academic writing. This shows that end weight is one motivating factor for the long passives in academic writing and the longer subjects and agents in academic writing can be an indication of the syntactic elaboration/complexity of academic writing mentioned by many previous researchers (Hyland, 2002; Hyland & Tse, 2005; Keen, 2004; Li & Ge, 2009; Wright, 2008). The givenness comparison demonstrates that most of the subjects of the long passives in both registers carry more given information than the agents, but overall, this trend is more obvious in academic writing. Therefore, academic writing is also a motivating factor for the choice of long passives in academic writing.

# Work-in-progress reports



## Old 'truths', new corpora: conjunct clauses revisited

Kristin Bech  
University of Oslo, Norway

It has often been claimed, to the point of becoming an axiom, that Old English conjunct main clauses, i.e. clauses starting with the coordinating conjunction *and* or *but*, and with an overt subject, typically have verb-final word order (see e.g. Mitchell 1985; van Kemenade 1987; Traugott 1992; Pintzuk 1995, Baker 2012), as in (1).

- (1) *7 On his broke he Gode fela behæsa behet*  
and in his illness he God.DAT many promises made  
'and in his illness he made many promises to God'  
(cochronE,ChronE\_[Plummer]:1093.3.3116)

Even though Bech (2001a, 2001b) showed that the frequency of conjunct main clauses is in fact very low, the opposite conception still seems to prevail. In this work-in-progress report, I therefore return to the issue and present new data from the *York-Toronto-Helsinki Parsed Corpus of Old English Prose* (Taylor et al. 2003), which corroborate the early studies in showing that conjunct clauses are not typically verb-final (9.5%). If verb-late clauses are included as well, the percentage rises to 21.6%, which is still low. I will demonstrate how I went about eliciting the verb-final and verb-late clauses from the corpus, as the latter in particular are not easily teased out. Furthermore, a crucial point in the early studies was that although conjunct clauses are not typically verb-final, verb-final clauses are typically conjunct clauses (57%). I will briefly suggest some reasons for this asymmetry that have to do with the intersection between the function of sentence types and the function of word order. The function of conjunct clauses is typically to elaborate on and modify the initial main clause (Traugott 1992), or to express consequence, result, contrast or addition (Quirk et al. 1985). The different word orders of the early stages of English, when word order was freer than it is today, can also be said to have discourse functions, since the word order of a clause to a large extent is dependent on the previous context. For example, verb-final order is typically used when there is focus on the verb, while the other clause elements are backgrounded. The paper outlines some of the results of this interplay between clause function and word order function.

## References

- Baker, Peter S. 2012. *Introduction to Old English*. Wiley-Blackwell.
- Bech, Kristin. 2001a. Are Old English conjunct clauses really verb-final? In L.J. Brinton (ed), *Historical linguistics 1999: selected papers from the 14th International Conference on Historical Linguistics, Vancouver, 9-13 August 1999*, 49-62. John Benjamins.
- Bech, Kristin Bech, Kristin. 2001b. Word Order Patterns in Old and Middle English: A Syntactic and Pragmatic Study. PhD dissertation, University of Bergen. <<https://bora.uib.no/handle/1956/3850>>
- van Kemenade, Ans. 1987. *Syntactic Case and Morphological Case in the History of English*. Foris Publications.
- Mitchell, Bruce. 1985. *Old English Syntax*, vols. I & II. Clarendon Press.
- Pintzuk, Susan. 1995. Variation and change in Old English clause structure. In D. Sankoff, W. Labov and A. Kroch (eds), *Language Variation and Change 7*, 229-260.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.
- Taylor, Ann, Anthony Warner, Susan Pintzuk & Frank Beths. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE). Department of Linguistics, University of York. Oxford Text Archive, first edition. <http://www-users.york.ac.uk/~lang22/YcoeHome1.htm>
- Traugott, Elizabeth Closs. 1992. Syntax. In R. M. Hogg (ed), *The Cambridge History of the English Language*, vol. I, 168-289. Cambridge University Press.

### Rhythmic well-formedness and syntactic mess: Evaluating the influence of the Principle of Rhythmic Alternation on the Big Mess Construction

Anne-Katrin Blass

University of Trier, Germany

It is a well-founded observation that languages tend to prefer rhythmic constellations in which stressed syllables alternate regularly with unstressed syllables. This almost universal preference has been termed the "Principle of Rhythmic Alternation", and its effects have been traced in numerous studies (see Schlüter 2005: 17-42). In basic terms, the effects of the principle manifest themselves in strategies to avoid or compensate stress lapses and especially stress clashes. Particularly fascinating are those strategies which help to avoid violations of the principle not by intra-phonological but by extra-phonological means, such as the choice of grammatical structures that are preferable to variant structures in prosodic terms. By the same token, this means that rhythmic preferences can be expected to co-determine grammatical variation and change (cf. Schlüter 2005). In my study, I test whether the Principle of Rhythmic Alternation provides an explanation of the non-canonical order of constituents underlying a construction that is in fact so far removed from generalisations about the prototypical English noun phrase (NP) that it has been termed the "Big Mess Construction" (Berman 1974). For typical examples of the construction see (1) to (3) below.

(1) To try and take it in was like drinking too strong a wine. (BNC: GWF 2996)

(2) Nobody dreamed of so swift a break. (BNC: ABG 573)

(3) I didn't get that good a look. (BNC: HTL 2630)

In each of the italicised NPs, a premodifying adjective phrase (AP) (*too strong*, *so swift*, *that good*) precedes the element realising the determiner function (the indefinite article *a*) rather than following it. As examples (1) to (3) suggest, the possibility for the APs to occur in this fronted position is dependent on the fact that the head adjectives are in their turn premodified by one of a limited set of degree modifiers (here *too*, *so*, *that*). For those degree items (e.g. *too*) which can introduce both a Big Mess NP (e.g. *too strong a wine*) and an attributive AP inside an NP which conforms to the canonical order of NP constituents (e.g. *a too strong wine*), I compare the instances of both structural sets in terms of their stress constellations; this comparison makes it possible to test Bolinger's (1972) (purely speculative) suggestion that the predeterminer position of the AP has the function of avoiding a stress clash which would ensue in the alternative attributive structure. A quantitative study of the different stress patterns underlying the instances of the Big Mess set and the attributive alternative set reveal, however, that rhythmic preferences can at most be claimed to have a co-determining effect on this area of grammatical variation. Furthermore, I also investigate in how far rhythmic preferences can be claimed to have an effect on the use of the Big Mess order in versions of the construction where the existence of an attributive alternative is at best questionable (e.g. *that good a look* versus *a that good look*).

#### References

Berman, Arlene (1974) Adjectives and adjective complement constructions in English. PhD thesis, Harvard University.

Bolinger, Dwight (1972) Degree words. The Hague: Mouton.

Schlüter, Julia (2005) Rhythmic Grammar: The Influence of Rhythm on Grammatical Variation and Change in English. Topics in English Linguistics 46. Berlin/ New York: Mouton de Gruyter.

## An Early Start to English Language Learning: Collocations in the Primary School Classroom

Martina Brendenbröker

University of Paderborn, Germany

A variety of studies (cf. Meunier/Granger 2008, Aijmer 2009, Römer 2009, Reppen 2010, etc.) stress the importance of transferring research results from corpus linguistics to the context of language pedagogy. Their focus is almost always on learners older than 11 who are at **secondary** school level. However, to provide a solid foundation for life-long language learning as proposed in the Common European Framework of Reference for Languages (CEFR), it is necessary to start earlier, e.g. from school-beginner age. Consequently, this paper is one of the first to focus on young learners in **primary** school classrooms. The Oxford Children's Corpus (OCC) is the basis for an investigation of lexical structures relevant for young learners of English. It has been developed by Oxford University Press and is used by their lexicography team as a reference for writing children's dictionaries (Wild et al. 2012; Banerji, N. et al. 2013). It currently comprises 126 million tokens, including material written for 5- to 14-year-old children (e.g. fiction, websites, magazines) and over 86 million tokens of writing by children themselves. Using SketchEngine (Kilgarriff, A. et al. 2004), an online corpus query tool, a subcorpus (OCC-SUB) was compiled, which is currently at ca. 26 million tokens. OCC-SUB contains writing exclusively by primary school children and is therefore pertinent to the purpose of this study.

Lemma frequency lists derived from the OCC-SUB will be contrasted with BNC frequency lists. Five nouns and five lexical verbs from the top ranks of both corpora have been chosen for a study of their collocational behaviour. The aim is to develop a foundational set of collocations by applying different statistical measures. An example may illustrate the need for linguistic analyses carried out in a specially-designed children's corpus, since authentic language usage by young children can differ considerably from that of adults: The verb *SEE* is commonly used by adults with inanimate subject nouns such as *year*, *century* and *world*, whereas the OCC-subjects are animate and consist mostly of personal pronouns. A possible explanation for this finding is that the basic meanings of *SEE*, namely *perception by sight*, *watch* and *understand* presuppose not only animate but in particular human subjects. Clearly, children acquire these meanings first. A common collocation in both corpora is *to see things* as in: 'Then I saw a green human thing coming towards me.' [OCC: BBC-E-111392]. The MI- score (29.29) and the T-score (47.48) indicate that it is both a strong and a certain collocation. Therefore, it should be included in the basic vocabulary taught in EFL textbooks at primary school level. The resulting set of collocations will be used as a starting point for an investigation of language- pedagogical consequences. By applying reliable vocabulary selection criteria, such as learnability, availability, familiarity, coverage and regularity (cf. Nation 2001), the chunk collection will be evaluated with regard to the specific needs of young learners, including implications for different vocabulary presentation and new forms of motivating and playful exercises which are suitable for children at that early age.

### References

- Aijmer, K. (2009). *Corpora and Language Teaching*. Amsterdam: Benjamins.
- Banerji, N., Gupta, V., Kilgarriff, A., Tugwell, D. (2013). 'Oxford Children's Corpus: A Corpus of Children's Writing, Reading, and Education'. In: *Corpus Linguistics 2013 Abstract book* <<http://ucrel.lancs.ac.uk/cl2013/doc/CL2013-ABSTRACT-BOOK.pdf>> (accessed 9.9.13).
- European Commission. (2003) 'Action Plan on Language Learning and Linguistic Diversity'. <[http://europa.eu/legislation\\_summaries/education\\_training\\_youth/lifelong\\_learning/c11068\\_en.htm](http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11068_en.htm)> (accessed 9.9.13).
- Kilgarriff, A., Rychly, P., Smrz, P., Tugwell, D. (2004). 'The Sketch Engine'. In: Williams, G., Vessier S. (Eds.) *Proceedings of the Eleventh Euralex Congress*, Lorient: UBS.
- Meunier, F.; Granger, S. (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: Benjamins.
- Nation, P. (2001). *Learning Vocabulary in another Language*. Cambridge: Cambridge University Press. Reppen, R. (2010). *Using Corpora in the Language Classroom*. Cambridge: Cambridge University Press.
- Römer, U. (2009). 'Corpus Research and Practice: What Help do Teachers Need and what can we Offer?' In: Aijmer, K. *Corpora and Language Teaching*. Amsterdam: Benjamins.
- Wild, K., Kilgarriff, A., Tugwell, D. (2012). 'The Oxford Children's Corpus: Using a Children's Corpus in Lexicography'. In: *International Journal of Lexicography*, ecs017v1-ecs017. Oxford: Oxford University Press.

## A diachronic study of the stative possessive in spoken British English

Joanne Close

University of Chester, United Kingdom

This paper investigates variation and change in the stative possessive in contemporary spoken British English using the Diachronic Corpus of Present-Day Spoken English (DCPSE). DCPSE comprises approximately one million words of transcribed spoken data from the London Lund Corpus (LLC; 1958-77) and the British Component of the International Corpus of English (ICE-GB; 1990-2). The grammatical annotation and the time span of approximately 30 years between the sub-corpora make DCPSE an ideal platform for studying recent grammatical changes. In fact, the corpus was built for this purpose. Many studies in the field of recent (or current) change assume that shifts in the frequency of lexical items or grammatical constructions are evidence for linguistic change (e.g. Mair and Leech 2006, Leech et al. 2009). Other studies apply a sociolinguistic methodology whereby the alternation between variants of a variable is studied in order to find evidence of change (e.g. Aarts and Close 2010, Bowie and Aarts 2012, Aarts et al. 2013). This paper falls into the latter category: in order to find evidence of change in the stative possessive, frequencies of the two variants, HAVE and HAVE got as exemplified in (1), are compared in the sub-corpora which make up DCPSE.

- 1) a. We think we've got the main point about that. (DCPSE:DI-A03/ICE-GB:S1A-003 #0039:1:A)  
b. And I have a machine. (DCPSE:DI-A03/ICE-GB:S1A-003 #0029:1:B)

This comparison of the frequencies of two variants addresses the research question outlined in (i), below.

- i. is the stative possessive undergoing change in spoken British English?  
ii. what (if any) are the grammatical factors governing the variability of the stative possessives HAVE and HAVE got?

In order to address (ii), all instances of HAVE and HAVE got occurring in variable contexts were extracted from the corpus and then coded according to a number of grammatical factors including contraction, negation, nature of the object (concrete/abstract), and type of subject (person, number, and pronoun/noun phrase). This follows Tagliamonte (2003) who found in her apparent time study of the stative possessive that, although HAVE and HAVE got used to be in competition, HAVE got is now the leading variant (see also Quinn 2009). Preliminary findings suggest that this is also the case in DCPSE where HAVE got is more frequent in ICE-GB (1990s) than in LLC (1950s-70s). Study of the grammatical factors also indicates that type of subject is likely to provide some interesting results.

## References

- Aarts, B., Close, J. & Wallis, S. (2013). Choices over time: methodological issues in investigating current change. In Aarts, B., J. Close, G. Leech & S. Wallis (Eds.), *The verb phrase in English: Investigating recent language change with corpora* (pp. 14 - 45). Cambridge: Cambridge University Press.
- Bowie, J. & Aarts, B. (2012). Change in the English infinitival perfect construction. In Nevalainen, T. and E. Closs Traugott (Eds.), *The Oxford handbook of the history of English* (pp. 200-210). Oxford / New York: Oxford University Press.
- Close, J. & Aarts, B. (2010). Current change in the modal system of English. In Lenker, U., J. Huber & R. Mailhammer (Eds.), *English historical linguistics 2008. Selected papers from the fifteenth international conference on English historical linguistics (ICEHL 15), Munich, 24—30 August 2008. Volume I: The history of English verbal and nominal constructions* (pp. 165-82). Amsterdam: John Benjamins.
- Leech, G., Hundt, M., Mair, C. & Smith, N. (2009). *Change in Contemporary English: A grammatical study*. Cambridge: Cambridge University Press.
- Mair, C. & Leech, G. (2006). Current changes in English syntax. In Aarts, B. & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 318-42). Oxford: Blackwell.
- Quinn, H. (2009). Downward reanalysis and the rise of stative HAVE got. In Crisma, P. & G. Longobardi (Eds.), *Historical syntax and linguistic theory* (pp. 212-30). Oxford: Oxford University Press.
- Tagliamonte, S. A. (2003). *Every place has a different toll: Determinants of grammatical variation in cross-variety perspective*. In Rohdenburg, G. & B. Mondorf (Eds.), *Determinants of grammatical variation in English* (pp. 531-54). Berlin: Mouton de Gruyter.

### Translator's voices and the rendering of English –ing clauses in Norwegian

Hildegunn Dirdal  
University Oslo, Norway

It has come to be recognized that translators do not just recreate the author's voice, but add their own voices to the translated text (Alvstad 2013). In a previous study (Author, forthcoming), I investigated whether the use of clause building and reduction could be seen to contribute to translators' individual voices, and found great individual variation both in terms of overall numbers and in the types of structures employed. The study focused mainly on changes to coordinate structures and prepositional phrases, which exist in both English and Norwegian. In this study, I look at *-ing* clauses, which have no formal equivalent in Norwegian, in order to answer the following questions:

- do we find individual preferences in the rendering of English *-ing* clauses in Norwegian?
- is there less/more variation in the rendering of *-ing* clauses than in the translation of structures that exist in both languages?
- is individual variation linked to different degrees of explicitation?

Baker (1996) demonstrates the advantage of using corpora to compare translators' styles, but acknowledges the drawback inherent in comparisons of translations of different originals. Corpora used in translation studies may contain only translations or both translated and original texts, but they do not tend to contain several contemporary translations of the same text into the same language, simply because such parallel translation activity very seldom takes place. However, only with such a corpus is it possible to keep constant the factors of source language, author's style, time period, etc. and be more confident that the differences found are actually due to the translators' individual styles. The English–Norwegian multiple-translation corpus is unique in this respect. It was created by Stig Johansson and Linn Øverås for the study of individual variation between translators, and consists of one fictional and one non-fictional text and ten different translations of each. For the present study, all the *-ing* clauses in the fictional text have been identified, and the ten translations have been compared. Preliminary results indicate that translators have individual preferences in the rendering of *-ing* clauses, but that there is no more variation than in the translation of structures that exist in both languages. The amount and type of variation is dependent on the relation between the *-ing* clause and the main clause and the possibilities that exist in Norwegian for expressing the same clause relation. In the cases where more and less explicit renderings of the clause relation are possible, some translators prefer explicit renderings more often than others.

### References

- Alvstad, C. 2013. 'Voices in Translation'. In Yves Gambier and Luc van Doorslaer (eds.) *Handbook of Translation Studies*, Volume 4. Amsterdam: John Benjamins.
- Author. Forthcoming. 'Individual variation between translators in the use of clause building and clause reduction'. *Oslo Studies in Language* 6.
- Baker, M. 1996. 'Towards a Methodology for Investigating the Style of a Literary Translator'. *Target* 12:2, 241-266.

## The placement of adverbial connectors of contrast in English and French editorials: a corpus-based study

Maité Dupont

Universite catholique of Louvain, Belgium

Languages in the world differ in terms of word order flexibility. English has often been claimed to have a very rigid word order (Becher 2011, Odlin 1989), notably as compared to French (Trévisé 1986). Another difference between English and French word order patterns pertains to adverb placement: while French favours *verb-adverb-object* structures, in English adverbs typically occur before the verb (Osborne 2008). This study sets out to test these claims for one specific type of adverb, i.e. adverbial connectors expressing contrast. Our main objective is to determine whether the two languages differ in terms of word order flexibility in the use of adverbial connectors of contrast and to identify the syntactic slots that are preferred overall in each language. A second objective is to assess the degree of idiosyncrasy of individual connectors within each language system as regards positioning. The first step consisted in drawing up a comprehensive list of English and French adverbial connectors of contrast. The lists of connectors collected from various reference books were supplemented using Dyvik's (1998) mirror analysis. The resulting inventory consists of 27 French and 24 English adverbial connectors expressing contrast (e.g. *however*, *in contrast*; *cependant*, *par contre*). Each connector was searched in a large comparable corpus of English and French editorials extracted from the Mult-Ed corpus (<http://www.uclouvain.be/en-cecl-multed.html>). The automatic extraction was followed by a careful process of manual disambiguation whereby all the occurrences in which these items were not used as adverbial connectors of contrast were discarded. For the syntactic analysis, the occurrences of each connector were classified into five categories: initial 1 (at the very beginning of the clause, [1]); initial 2 (before the subject, but after another optional sentence constituent, [2]); medial 1 (between the subject and the verb, [3]); medial 2 (within the verb phrase, between the verb and the complement, or within the verb complement, [4]); and final (at the very end of the clause, [5]).

- [1 ] **Néanmoins**, les grands groupes structurent le paysage par leurs retombées. Now,
- [2 ] **however**, the world's surplus capacity is disappearing.
- [3 ] Mark Mardell, **on the other hand**, has white-suit cred written all over him. Il n'a
- [4 ] **cependant** pas ouvert de piste pour l'avenir. That's not yet enough to meet demand,
- [5 ] **though**.

The results show that all five positions are occupied in both languages. Thus, no language appears to make available more syntactic slots for the use of connectors than the other. However, English connectors are much more evenly distributed across the five positions than French connectors, which tend to cluster in the initial 1 and medial 2 categories. French typically places connectors medially, within the core *subject-verb-object* structure of the clause. English, by contrast, places the majority of its connectors at sentence periphery (in the initial and final positions). The analysis of individual connectors in each language brings out a high degree of idiosyncrasy, with each connector being primed (Hoey 2013) for a particular (set of) position(s) in the sentence. In my presentation I will give a detailed description of the results and discuss their implications for the teaching of connectors.

### References

- Becher V. 2011. When and why do translators add connectives?. *Target* 23(1), 26-47.
- Dyvik H. 1998. A Translational Basis for Semantics. In Johansson S. & Oksefjell S. (eds.) *Corpora and cross-linguistic research: Theory, method and case studies*. Amsterdam & Atlanta: Rodopi, 51-86.
- Hoey M. 2013. Lexical Priming. In Chapelle C. A. (ed.) *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell Publishig Ltd.
- Odlin T. 1989. *Language Transfer: Cross-linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Osborne J. 2008. Adverb placement in post-intermediate learner English: a contrastive study of learner corpora. In Gilquin G., Papp S. & Díez-Bedmar M.B. (eds.) *Linking up contrastive and learner corpus research*. Amsterdam & New-York: Rodopi, 127-146.
- Trévisé A. 1986. Is it transferable, topicalization?. In Kellerman E. & Sharwood Smith M. (eds.) *Crosslinguistic Influence in Second Language Acquisition*. New-York: Pargamon, 186-206.

### Lexical Bundles in British and West African English. A Corpus-Based Analysis of British, Ghanaian and Nigerian News Texts.

Sabrina Edler

University of Muenster, Germany

The focus of this PhD project is on the contrastive analysis of multi-word combinations in one inner circle variety, i.e. British English (BrE), and two outer circle varieties, i.e. Ghanaian English (GhE) and Nigerian English (NigE). The need for phraseological studies in the field of these so-called New Englishes is emphasized among others by Nesselhauf (2009, p. 2)<sup>i</sup>, who points out that "studies of lexical or lexico-grammatical co-selection phenomena in New Englishes [...] are rare". Focusing on but not limiting itself to this research gap, this PhD project aims at providing answers to the following research questions:

- (I.) Which lexical bundle (LB) structures are preferred in the registers of British, Ghanaian and Nigerian press reportage and editorials and which relationship can be observed between the grammatical structure and the discourse functions of LBs?
- (II.) Which factors account for the divergent distribution of LBs across the selected varieties and registers (second language acquisition, register conventions and/or the development of GhE and NigE as distinctive national standards)?
- (III.) Do the results mirror the development of GhE and NigE towards endonormative stabilization as suggested by Schneider (2007)<sup>ii</sup> in his *Dynamic Model*? Based on the findings, is it justifiable to postulate the existence of distinctive labels such as "Ghanaian English" and "Nigerian English"? And which conclusions may be drawn from the results regarding the wider applicability of the *Dynamic Model*?

The analysis is conducted using data from two registers, i.e. press reportage and editorials. For the West African varieties, the data is drawn from the online editions of five national Ghanaian and Nigerian newspapers, respectively, while the newspaper section of the British National Corpus (BNC) is used for the investigation of the British news texts. *Lexical bundles*, defined by Biber et al. (1999, p. 992)<sup>iii</sup> as "[...] combinations of words that [...] recur most commonly in a given register" are the relevant linguistic variable. For the retrieval of the bundles, the free concordancer AntConc 3.3.5w is used. A pilot study investigating the distribution of 4-word bundles across the data has shown that the three varieties prefer LBs with a nominal or prepositional core (e.g. *the end of the*, *at the end of*, respectively) to LBs forming a verb phrase (e.g. *stress the need for*). GhE and NigE, however, have been found to use a much larger stock of LBs in general and LBs representing a noun phrase or prepositional phrase, in particular. These findings are confirmed by the present study which moreover yields that BrE, GhE and NigE all display a very low type-token ratio across both press reportage and editorials. In the registers investigated, GhE and NigE have thus not been confirmed to show characteristics of nativization or even endonormative stabilization. The currently conducted functional and structural analysis of LBs, however, might yet reveal both a pragmatic and grammatical nativization of these varieties.

#### References

- Nesselhauf, N. (2009): "Co-selection phenomena across New Englishes. Parallels (and differences) to foreign learner varieties." *English World-Wide*, 30(1). 1-26.
- Schneider, E. W. (2007). *Postcolonial English. Varieties around the world*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

### Exploring information-theoretic complexity trends of morphs and constructions in English: A corpus-based approach

Katharina Ehret

University of Freiburg, Germany

This paper is an excursion into methodologically unexplored territory and analyses the contribution of morphs and constructions to complexity in English marrying corpus linguistics to information-theoretic methodologies. Specifically, I assess the degree to which individual morphs such as *-ing* or *-ed*, and functional constructions such as progressive (*be* + verb-*ing*) or perfect (*have* + verb past participle) contribute to the syntactic and morphological complexity in English texts. On an interpretational plane, the complexity of these morphs and constructions in the context of a given text is inferred from their complexity contribution to the text. In general terms, I demonstrate how information-theoretic methods can be successfully applied to naturalistic text corpora. The point of departure is the ongoing complexity debate and quest for complexity metrics. For decades linguists tacitly adhered to the dictum that all languages are equally complex (; Crystal 1987; Hockett 1958; O'Grady et al. 1997). Recent evidence to the contrary—some languages seem, in fact, to be less complex than others (Kusters 2003; McWhorter 2001)—has caused a stir and triggered a wave of research on how to define and measure linguistic complexity (Dahl 2004; Kortmann and Szmrecsanyi 2012; Miestamo et al. 2008; Sampson 2009). Against this backdrop, I present a radically objective and economical method to assess linguistic complexity in corpora. A new flavour of the Juola-style compression technique (Juola 1998)—targeted manipulation—is used to assess the contribution of specific linguistic units to (con)textual complexity in English. The idea is to measure the linguistic complexity in text samples by way of approximating their relative informativeness with compression algorithms. Thus, a set of ten features is analysed in a mixed-genre corpus sampling texts from *Alice's adventures in Wonderland*, the Gospel of Mark and a newspaper corpus. I find that higher amounts of morphological markers lead to higher amounts of morphological complexity in the corpus. Syntactic complexity, on the other hand, is reduced because the presence of morphological markers enhances the predictability of morphosyntactic patterns. Non-transparency resulting from irregularity or allomorphy increase the morphological complexity of the texts. In fact, the results tie in with well-established, quantitative complexity metrics (McWhorter 2012; Szmrecsanyi 2009; Szmrecsanyi and Kortmann 2009; Trudgill 2004) and connect with findings on morpheme order acquisition (Brown 1973; de Villiers and de Villiers 1973; see also Goldschneider and DeKeyser 2005). In short, this paper provides a fresh perspective on corpus linguistics by applying information-theoretic methods to naturalistic text samples and demonstrates that compression measurements produce linguistically meaningful results. Thus, the compression technique is a powerful tool for assessing general complexity trends in corpora. I conclude by sketching directions for further research and point out advantages and drawbacks of the method.

### References

- Brown, R.. *A First Language. The Early Stages*. Harvard University Press, Cambridge, Mass., 1973.
- Crystal, B. *The Cambridge Encyclopedia of Language*. Cambridge University Press, Cambridge, 1987.
- sten Dahl. *The growth and maintenance of linguistic complexity*. John Benjamins, Amsterdam, Philadelphia, 2004.
- de Villiers, J. and de Villiers, P. A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, 2(3):267-278, 1973.
- Goldschneider, J. and DeKeyser, R.. Explaining the "Natural order of L2 morpheme acquisition" in english: A meta-analysis of multiple determinants. *Language Learning*, 55(S1):27-77, 2005.
- Hockett, C. F. *A course in modern linguistics*. Macmillan, New York, 1958.
- Juola, P.. Measuring linguistic complexity: the morphological tier. *Journal of Quantitative Linguistics*, 5(3):206-213, 1998.
- B. Kortmann and B. Szmrecsanyi, editors. *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Lingua & Litterae. Walter de Gruyter, Berlin/Boston, 2012.
- Kusters, W. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. LOT, Utrecht, 2003.
- McWhorter, J. Complexity hotspot. In B. Kortmann and B. Szmrecsanyi, editors, *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*, 243-246. Walter de Gruyter, Berlin/Boston, 2012.
- Miestamo, M., Sinnemki, K. and Karlsson, F. editors. *Language complexity: typology, contact, change*. Benjamins, Amsterdam, Philadelphia, 2008.
- O'Grady, W., Dobrovolsky, M. and Aronof, M. *Contemporary Linguistics: An Introduction*. St. Martin's, New York, 1997.
- Sampson, G. A linguistic axiom challenged. In G. Sampson, D. Gil and P. Trudgill, editors, *Language Complexity as an Evolving Variable*, pages 1-18. Oxford University Press, Oxford, 2009.
- Szmrecsanyi, B. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of english. *Language Variation and Change*, 21(3):319-353, 2009.
- Szmrecsanyi, B. and Kortmann, B. Between simplification and complexification: non-standard varieties of english around the world. In *Language Complexity as an Evolving Variable*, pages 64-79. Oxford University Press, Oxford, 2009.
- Trudgill, P. Linguistic and social typology: The austronesian migrations and phoneme inventories. *Linguistic Typology*, 8:305-320, 2004.



### Corpora and EAP

*Maria Freddi*

*University of Pavia, Italy & University of Cambridge, United Kingdom*

The particular demands EAP teaching poses for teachers in the higher education system call for continuous reflection on course and syllabus design. In this paper I argue for a combination of paradigms in designing an EAP course, specifically the corpus-based approach, on the one hand (e.g. Aijmer 2009; Coxhead 2010; Hyland 2000, 2006; Krishnamurthy and Kosem 2007; Nesi and Gardner 2012; Römer 2010, Thompson 2006, 2007), and the discourse analytic, drawing from classical genre and text theories, on the other (e.g. Swales 1990; Johns 1997; Peck-McDonald 2004). Within this perspective, the argument is made for referring to corpus methods and findings to guide the instructors and course developers' choices without necessarily bringing corpus observation tools into class (unlike done, for example, by Charles 2007, 2014). The focus is on EAP reading pedagogy. As argued by Flowerdew 2004, 2009, 2012, there is certainly a way in which corpora can contribute to effective language teaching, and corpus data might be extremely relevant for observational purposes, i.e. to develop receptive skills. In the academic and specialised contexts, although a lot of work has been done to pinpoint recurrent features of academic genres in certain disciplines (e.g. Charles 2007, Bondi 1999), the humanities are still understudied as to the linguistic structures that need to be made salient for the novice who is learning how to read, write, etc. about/in certain disciplines (e.g. film studies, classical studies, art history) and related genres (e.g. the argumentative essay). This is reflected in EAP coursebooks, that despite being sometimes corpus-informed, do not necessarily pinpoint the disciplinary differences enough. The paper explores ways in which insights from corpus approaches to academic English combined with genre theory can be brought to bear on the design of the course syllabus and argues for a pedagogically targeted mix of the various paradigms under consideration. The method consists in concordancing a number of authentic texts from both the humanities and the hard sciences with a view to identifying patterns that might emerge from within each text taken individually first, from across texts within one genre, and from across genres. As a final step, patterns belonging to different disciplines are compared. The various comparative steps are taken to identify frequent phraseology that can help distinguish genres and point the instructor to more pedagogically relevant authentic material. To that extent, the methodology followed is corpus-driven. However, it is combined with a more corpus-based approach whereby the texts/corpora are searched for specific structures that are known to genre studies to be problematic both with students of the humanities and the hard sciences, namely the noun phrase (Parkinson 2014 but see also previous work done by Mahlberg 2009). It is expected that by combining the two approaches to materials selection a deeper understanding of the disciplinary differences will be gained, which can usefully be reused to adjust the design of the syllabus.

### Collocational Preferences of Tibetan Loanwords in a Corpus of Tibetan Buddhist English

Jessica Frye

Manchester Metropolitan University, United Kingdom

Buddhist English has been widely neglected within the field of English for Specific Purposes. The presentation will provide insight into the textual genre of Buddhist commentaries by using a corpus approach. The paper will present findings of a study compiling a small, specialized corpus of written Tibetan Buddhist commentaries (*shastras*) studied in Buddhist colleges (*shedras*) in both the East and West, offering insight into a textual genre that has been widely neglected. Providing a corpus of Buddhist texts will be a valuable contribution to monastic education in the East, where donation-run monasteries aim to enable non-ordained students to communicate in English. The creation and exploitation of small specialised corpora for materials design has been discussed previously (e.g. Gavioli 2005), yet corpus-informed materials for Buddhist English have not been developed. Increased interest in Tibetan Buddhism in the West can also benefit from an exploitation of a corpus of Buddhist texts to gain a better understanding of the accurate use of common terminology. Building on this background, the primary focus of the paper is an investigation of loanwords within the corpus. Preliminary findings of this corpus have identified an unusually high frequency of loanwords from Tibetan in Buddhist discourse (e.g. *bodhisattva*). General frequency data will be used to select high frequency loanwords and to provide a general overview of the distribution of loanwords within the corpus. Further investigation will examine their preference of co-occurrence. These insights will enhance the field of English for Specific Purposes by providing linguistic insight into Buddhist discourse.

### References

Gavioli, L. 2005. *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.

### Appraisal in discussion sections of doctoral theses in the discipline of ELT/Applied Linguistics at Warwick University

*Yifan Geng*

*University of Warwick, United Kingdom*

Investigation of interpersonal meaning in academic discourse by means of Appraisal Theory has recently drawn increasing interest (e.g. Chang & Schleppegrell, 2011; Chatterjee, 2008; Hood, 2004, 2005; Koutsantoni, 2004; Liu & Thompson, 2009; Swain, 2010; Wu, 2007). Despite the robust interest in this research field, examination of interpersonal meaning in doctoral theses, acknowledged by Swales (2004: 99) as the “most sustained and complex piece of writing” which researchers undertake, seems to receive less attention in literature. Drawing on Martin and White’s (2005) Appraisal Framework and adopting a corpus-based approach, my research in progress therefore attempts to analyze and compare the deployment of interpersonal meaning in discussion sections of doctoral theses in the discipline of ELT/Applied Linguistics by six L1 (First Language) Chinese and six L1 English writers at Warwick University, UK. My research aims to answer the following questions: 1) What Appraisal options are used to construe interpersonal meanings in the discussion sections of doctoral theses in the discipline of ELT/Applied Linguistics at CAL, Warwick University? 2) Are there any systematic differences in the choices made by L1C and L1E speakers, or not? 3) Is there a difference to be observed in tendencies for linguistic realizations of Appraisal options found, and in tendencies for combinations of some Appraisal options, across the L1C and L1E sub-corpora? 4) What are some salient rhetorical consequences of the Appraisal options or combinations of some options found? In my presentation, I will focus on presenting my research results including: 1) Quantitative results about whether there are any statistically significant differences in the use of Appraisal options across the L1C and L1E sub-corpora. 2) Preliminary qualitative results about how the Appraisal options or combinations of some options found enabled thesis writers to discuss their own research findings with respect to previous literature.

### Exploring probabilistic grammar(s) in varieties of English around the world

Jason Grafmiller, Benedikt Heller, Melanie Roethlisberger, Benedikt Szmrecsanyi  
University of Leuven, Belgium

We report on an ongoing project marrying the probabilistic grammar framework (Bresnan et al. 2007; Bod, Hay and Jannedy 2003) to research along the lines of the “English World-Wide Paradigm”, which is concerned with the sociolinguistics of, and linguistic variation across, post-colonial English-speaking communities around the world (Mesthrie 2006; Schneider 2007). Through state-of-the-art quantitative analyses of syntactic alternations across nine varieties of English represented in the International Corpus of English (St BrE, Hong Kong E, Jamaican E, Canadian E, Singapore E, Philippines E, Indian E, New Zealand E, Irish E), the project addresses the central question: How does language users’ grammatical knowledge differ across post-colonial varieties of English? In treating variation as a “core explanandum” (Adger and Trousdale 2007: 274) of linguistic theory, the project contributes to the development of usage-based theoretical linguistics by adopting a variational, large-scale comparative, and sociolinguistic perspective. Previous research within the English World-Wide Paradigm has primarily focused on the presence, absence, or variable usage frequencies of grammatical features (e.g. Kortmann and Szmrecsanyi 2004; Steger and Schneider 2012). While such studies are undoubtedly interesting, this project is more ambitious in scope. It is an investigation of linguistic knowledge that goes beyond mere description of frequencies, and links the English World-Wide Paradigm to recent advances in usage- and experience-based models of language (see also Bernaisch et al. To appear). The project explores variability in the hidden though cognitively ‘real’ probabilistic constraints that fuel variation within and across speech communities. Such constraints, e.g. the tendency to place long constituents after short constituents (Wasow and Arnold 2003), are not necessarily tied to surface material but to subtle stochastic generalizations about language usage, which- according to experimental evidence (Ford and Bresnan 2010)- language users implicitly know about. Thus, rather than simply describing probabilistic variation in corpus data, the project aims to illuminate aspects of the linguistic knowledge that language users with differing English backgrounds implicitly command. We accomplish this objective by investigating the probabilistic effects of various constraints on linguistic choice-making as a function of regional variety using advanced statistical methods, primarily mixed-effects logistic regression modelling (Pinheiro and Bates 2000) and conditional random forests analysis (Tagliamonte and Baayen 2012). Four patterns of grammatical variation (“alternations”) sensitive to a range of constraints take centre stage: the GENITIVE ALTERNATION, the DATIVE ALTERNATION, PARTICLE PLACEMENT, and (NON-)FINITE COMPLEMENTATION. Each of these alternations are examined within and across data from parallel spoken and written registers in each of the nine English varieties. Throughout this endeavour, we focus on questions related to, among other things: the extent to which different English varieties share a core grammar that is explanatory across different varieties; the ways implicit knowledge of linguistic probabilities develops in historically diverging groups of speakers; the degree to which individual probabilistic constraints are crosslectally, sociohistorically, and/or culturally malleable; and the degree to which the alternations under study exhibit cross-constructural parallelisms.

### References

- Adger, David and Graeme Trousdale. 2007. Variation in English syntax: Theoretical implications. *English Language and Linguistics* 11:261–278.
- Bernaisch, Tobias, Stefan Th. Gries, and Joybrato Mukherjee. to appear. The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide* .
- Bod, Rens, Jennifer Hay, and Stefanie Jannedy, eds. 2003. *Probabilistic Linguistics*. Cambridge, MA: MIT Press.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Baayen Harald. 2007. Predicting the dative alternation. In G. Boume, I. Kraemer, and J. Zwarts, eds., *Cognitive Foundations of Interpretation*, pages 69–94. Amsterdam: Royal Netherlands Academy of Science.
- Ford, Marilyn and Joan Bresnan. 2010. Studying syntactic variation using convergent evidence from psycholinguistics and usage. In M. Krug and J. Schluter, eds., *Research Methods in Language Variation and Change*. Cambridge: Cambridge University Press.
- Kortmann, Bernd and Benedikt Szmrecsanyi. 2004. Global synopsis: Morphological and syntactic variation in English. In B. Kortmann, E. Schneider, R. Mesthrie, and C. Upton, eds., *A Handbook of Varieties of English*, pages 1142–1202. Mouton de Gruyter.
- Mesthrie, Rajend. 2006. World Englishes and the multilingual history of English. *World Englishes* 25:381– 390.
- Pinheiro, Jose C. and Douglas M. Bates. 2000. *Mixed Effects Models in S and S-PLUS*. New York: Springer.
- Schneider, Edgar. 2007. *Postcolonial English: Varieties Around the World*. New York: Cambridge University Press.
- Steger, Maria and Edgar Schneider. 2012. Complexity as a function of iconicity: The case of complement clause constructions in new Englishes. In B. Kortmann and B. Szmrecsanyi, eds., *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin: de Gruyter.
- Tagliamonte, Sali and Harald Baayen. 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24:135–178.
- Wasow, Thomas and Jennifer Arnold. 2003. Post-verbal constituent ordering in English.

### “Phraseological teddy bears”: frequent lexical bundles in academic writing by Norwegian learners and native speakers of English

Hilde Hasselgård  
University of Oslo, Norway

Some recent studies (e.g. Chen & Baker 2010, Ädel & Erman 2012, Author & Ebeling forthcoming, Paquot 2013) have found important differences in the ways that native speakers and learners of English use lexical bundles. The present investigation explores the most frequent 4-word bundles in two corpora of novice academic English: BAWE (English L1) and VESPA (Norwegian learners of English). The aim is to find out which bundles are most frequently used by the two writer groups, and *how* these bundles are used with respect to their distribution, meanings and functions. Only the linguistics subsections of the corpora were used, and only those texts whose writers had English (BAWE) or Norwegian (VESPA) as their first language. WordSmith was used to extract 4-word bundles. I focus mainly on those bundles that appear in the highest frequency band, i.e. those that account for at least 0.01% of the corpus according to the WordSmith's Wordlist function. There are more high-frequent bundles in VESPA than in BAWE (20 vs. 15). Moreover, the most frequent bundles have higher frequencies in VESPA: the two most frequent bundles, *on the other hand* and *the use of the*, occur 38 and 33 times per 100,000 words, respectively. By comparison, the two most frequent bundles in BAWE, *it is important to* and *in the case of*, occur 19 and 17 times, respectively. Though the VESPA frequencies drop sharply after rank 2, they remain higher than those of BAWE at each rank of the top-20 most frequent 4-word bundles. Some of the most frequent bundles are content-specific, e.g. *Norwegian learners of English*, *Australian and New-Zealand English*, *as in Tager and Flusberg*. The removal of these left the VESPA list shorter than the BAWE one. Of the remaining bundles four were shared between the corpora (*on the other hand*, *the use of the*, *the meaning of the*, *in the case of*); five were unique to the VESPA top-frequency list, and eight were unique to the BAWE list. These preliminary results indicate that advanced learners of English are prolific users of lexical bundles, but they choose their bundles in a different way from native speakers. As with single lexical items, the learners have clear favourites among lexical bundles; thus we may perhaps speak of ‘phraseological teddy bears’ (cf. Hasselgren 1994). In a study related to the present one, Lie (2013) conducted an in-depth analysis of three lexical bundles that learners and native speakers use in qualitatively different ways. One of these was *on the other hand*, which is more frequently used alongside *on the one hand* by the learners, and also more commonly as a general topic-shifter than as a marker of dichotomous contrast. A hypothesis to be explored in the proposed investigation is that other bundles too may show similar differences in their patterns of use across L1 and L2 academic writing.

### References

- Ädel, A. and B. Erman. 2012. Recurrent word combinations in academic writing by native speakers and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes* 31: 81-92.
- Chen, Y.-H. and P. Baker. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology* 14 (2): 30-49.
- Hasselgren, A. 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics* 4: 237-259.
- Lie, J. 2013. “The fact that the majority seems to be...” A corpus-based investigation of non-native academic English. Unpublished MA thesis, University of Oslo.
- Paquot, M. 2013. Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics* 18:3, 391–417.
- Author & S.O. Ebeling. Forthcoming. Learners' and native speakers' use of recurrent word-combinations across disciplines.

### Automatic N-Gram Analysis (ANGA) on the Basis of Biber et al.'s (1999) Lexical Bundle Categories

Benedikt Heller

University of Leuven, Belgium

Biber et al. (1999, Ch. 13) specify a list of lexical bundle categories that are typical of the two complementary registers *conversation* and *academic prose*. For example, in conversations the authors find a high number of lexical bundles that belong to the category *personal pronoun + lexical verb phrase* (44%), e.g. the 4-gram *I don't know what*. In academic prose, on the other hand, most of the bundles (33%) fit into the category *preposition + noun phrase fragment*, such as the 4-gram *as a result of*. In total, they describe 12 categories on the basis of which they distinguish between the two registers – five that are typical of each of the two registers, plus two additional categories, *to-clause fragments* and *other expressions*, which contain bundles that appear in both registers (cf. Biber et al. 1999: 996). Up to this point, there was no easy way to analyze lexical bundles within this framework, and therefore, there is not much empirical evidence of how accurately the proposed lexical bundle categories are able to distinguish between the registers *conversation* and *academic prose*. ANGA is capable of performing the categorization of n-grams into the aforementioned categories quickly and accurately. The only prerequisite is that texts have to be tagged with the CLAWS4 tagger (Garside & Smith 1997) beforehand. The program has a graphical user interface (GUI), which makes it easy to use, and it performs an analysis of a one-million-word corpus in about 60 seconds. Its output shows a list of all n-grams in the input file including different tagging variants, and a colored bar chart that indicates the percentages of n-grams in each of the 12 categories, which also compares the findings to Biber et al.'s (1999) baseline values. ANGA further computes an aggregate value that indicates the total deviation of a text from each of the two registers, merging all of the 12 categories. This total deviation indicates a text's proximity to each of the two registers. The results that were created on the basis of ICE-GB, ICE-NZ, ICE-USA and the Switchboard corpus show that a number of written and spoken (sub)corpora can be clearly distinguished from each other on the basis of their pattern distribution in the Biber et al. (1999) framework. These values make it possible, for instance, to clearly distinguish between the written and spoken components of the ICE family; texts that are of a spoken nature (e.g. the Switchboard corpus, which includes telephone conversations) are closer to *conversation* while written texts (as in e.g. ICE-GB written) are closer to *academic prose*. In sum, the program is able to distinguish between conversation and academic prose in a very economic way, and it indicates that Biber et al.'s (1999) categories are a valid means of distinguishing between texts from these two registers.

### References

- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. & R. Quirk (1999): Longman Grammar of Spoken and Written English. Harlow: Pearson.
- Garside, R., and Smith, N. (1997): A hybrid grammatical tagger: CLAWS4, in Garside, R., Leech, G., and McEnery, A. (eds.) Corpus Annotation: Linguistic Information from Computer Text Corpora. Longman, London, pp. 102-121.

### ICAME corpora in CLARINO

*Knut Hofland*

*University of Bergen, Norway*

This presentation will demonstrate the use of ICAME corpora within the CLARINO project. The CLARINO project is the Norwegian part of the EU CLARIN project. The software used is Corpuscle, one of the two corpus management systems used in CLARINO. CLARIN (Common Language Resources and Technology Infrastructure) is an EU project where the goal is "to make language resources and technology much more accessible to all researchers working with language material, in particular in the humanities and social sciences. CLARIN is committed to establish an integrated and interoperable research infrastructure of language resources and its technology. It aims at lifting the current fragmentation, offering a stable, persistent, accessible and extendable infrastructure and therefore enabling eHumanities". After a preparatory phase of 3 years the construction phase started in 2011 for 5 years. After this there will be an exploitation phase of at least 5 years. The construction phase is financed by the participating countries. The partners in CLARIN come from 26 countries. In 2012 CLARIN became a European Research Infrastructure Consortium (ERIC). In 2012 a Norwegian consortium of universities and other institutions, led by Koenraad De Smedt from University of Bergen, got resources from the Norwegian Research Council for a 5- year national CLARIN project, called CLARINO. ICAME material is one of the resources we plan to include and this was approved by The ICAME Executive Board in 2013. In 2013, CLARINO has signed Deposition License Agreement with several of the compilers of ICAME corpora. The corpora have been converted to the Corpuscle format and metadata for the corpus have been entered. Work on more of the ICAME corpora is in progress. The CLARIN licenses can be divided into 3 categories; public, academic or restricted. Most of the ICAME corpora will be in the academic category. The users can then login via a system for federated authentication and authorization. At the moment, Corpuscle has support for eduGAIN,, a project of 41 partners across Europe. Users from one of these partners can use their own university login to access the corpora. For users outside these partners, we have to use a system of user IDs based on manual approval. The Corpuscle system contains basic search functionality, concordance, collocations, distribution (based on metadata) and word lists. It is possible to download the results for further processing locally.

### References

Corpuscle: <http://iness.uib.no/korpuskel/main-page>

CLARIN: <http://www.clarin.eu/external/>

ICAME: <http://icame.uib.no/>

eduGAIN: <http://www.geant.net/service/eduGAIN/Pages/home.aspx>

Meurer, P. (2002) "Corpuscle: A new corpus management platform for annotated corpora" In G. Andersen (ed.). 2012. *Exploring newspaper language. Using the Web to create and investigate a large corpus of modern Norwegian*. John Benjamins Publishing Company: 31-49

### How different is the English translation from the native writings of English? A multi-dimensional analysis based on the balanced comparable Corpus of Translational English

Xianyao Hu, Richard Xiao  
Lancaster University, United Kingdom

The objective of this paper is to address the somewhat debatable proposal of the “Translational Universals” as first defined by Baker in 1993, the “universal features of translation, that is features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems” (43). The basic idea may be formulated into these challenging questions: Is the language used in translations *different* from the native writings in the same language? If yes, how different? And why? Except for the easily reasonable “interference” of the Source Language or the “translationese” caused by the translator’s incompetence, are there other more profound reasons for the differences? Can translation in the strict sense be regarded as the “Third Code” as proposed by Frawley (1984)? In the past 20 years, there has been an increasing amount of interest and research into those questions. Evidence has been given by many corpus-based studies on the typical or discriminating features of the translated texts in contrast to their native writing counterparts to testify the Translational Universals hypotheses. (For example, Kenny (2001), Laviosa (2002), Granger et al (2003), Mauranen et al (2004), Hansen (2003), Olohan (2004), Mauranen (2004), Anderman and Rogers (2008), Xiao (2010), etc.) However, evidence and counter-evidence coexist; linguistic features at different levels (lexical, syntactical or textual) give paradoxical support to the same universal hypotheses (e.g. simplification, explicitation, normalization, etc.) So, do Translation Universals really exist? This paper is, of course, an attempt at answering the question. We think one of problems of previous corpus-based studies on TUs is presumably the lack of a unified analytical framework which helps to identify most, if not all, of the linguistic features that actually *can* differentiate the translated from the non-translated. It would be ideal if such a framework could also tell us about the respective weights of these linguistic parameters in terms of differentiation. If so, the conflicts between contradictory evidence at various levels can be resolved. Accordingly, we hope to show it is useful and effective to apply the Multi-Dimensional or Multi-Featured (MD/MF) analysis (Biber 1988/1991) to the current study. With the aim of finding out the distinctive linguistic features of the translated English, we created a one-million-word balance comparable *Corpus of Translational English* (COTE) which is designed as a translational counterpart for the *Freiburg–LOB Corpus of British English* (F-LOB) (Hundt et al 1998). Based on the comparable corpora of COTE and F-LOB, we will be able to carry out analyses of the multiple linguistic features as presented by the MD/MF analytical framework. We hope that this investigation will demonstrate that the MD/MF analysis is effective in discriminating the translated texts from the texts originally written in English. And as previously proposed, we wish to demonstrate that the stylometric analysis is also effective in identifying the linguistic features that enable the differentiation.

#### References

- Anderman, Gunilla M. and Margaret Rogers, (2008). *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters.
- Biber, D. (1988/1991). *Variation across speech and writing*. Cambridge University Press.
- Granger, S., Lerot, J. & Petch-Tyson, S. (2003). *Corpus-Based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi.
- Hansen, S. (2003). *The Nature of Translated Text: An Interdisciplinary Methodology for the Investigation of the Specific Properties of Translations*. Saarbrücken: DFKI/Universität des Saarlandes.
- Hundt, M., Sand, A. & Siemund, R. (1998). *Manual of Information to Accompany the Freiburg-LOB Corpus of British English*. Freiburg: University of Freiburg.
- Kenny, D. (2001). *Lexis and Creativity in Translation. A Corpus-Based Study*. Manchester: St. Jerome Publishing.
- Laviosa, S. (2002). *Corpus-Based Translation Studies. Theory, Findings, Applications*. Amsterdam: Rodopi.
- Mauranen, A. and Kujamäki, P. (Eds) 2004. *Translation Universals: Do They Exist?* Amsterdam: John Benjamins.
- Xiao, R. (2010). *Using Corpora in Contrastive and Translation Studies*. Newcastle: Cambridge Scholars Publishing.



### The modals in recent English: A closer look at *may* and *must*

Svenja Kranich

University of Mainz, Germany

While there are a number of recent studies dealing with the decline of the modals in the 20<sup>th</sup> century (e.g. Leech & Smith 2006, Collins 2009, Leech 2011, Aarts et al. *forthc.*), some puzzling aspects of this development are still unresolved. For instance, the decrease of the core modals is often assumed to be balanced by an increase of the semi-modals, which are, however, mostly limited to deontic uses (Mair 2006). This raises the question of what happens to epistemic modality. Either the core modals are not going down in their epistemic uses or they must be replaced by other expressions. Alternatively, it is also imaginable that epistemic modifications in general become less frequent, a tendency which has been observed for popular scientific texts (Kranich 2011). The present paper aims at investigating the hypothesis of the core modals being increasingly restricted to epistemic functions. This hypothesis seems plausible, since an ongoing specialization of the modals to epistemic functions (historically their latest function) would tie in well with general tendencies observed in grammaticalization research, where functional specialization is often witnessed in late stages of grammaticalization (cf. e.g. Bybee et al. 1994). I will take a closer look at the functional development of *must* and *may*. Based on genre-balanced samples of 300 occurrences per decade of each of these modals extracted from *COHA*, I will present a functional analysis of the use of *may* and *must* in the 1960s and the 2000s, differentiating between non-epistemic (dynamic, deontic) and epistemic modal meanings. The results of the analysis show that *must* and *may* decline in different ways: *Must* shows no differences in its distribution across functions in the two time-spans, regardless of which genre we look at. *May*, on the other hand, exhibits statistically significant differences in the distribution of its functions between the two decades studied, with epistemic uses becoming clearly more frequent in the 2000s. What is even more interesting is that this difference is solely due to a highly significant difference between the non-fictional texts in the two time frames, whereas the fictional sample does not yield statistically significant differences. The reason for this changing usage seems to be a decline in the use of a particular hedging pattern with non-epistemic *may*, evidenced in the following example: For instance, we **may** tentatively put: (I) a believes that  $p =$  in all possible worlds... This pattern appears to be very common in 1960s (popular) scientific writing: The author uses deontic *may* to give himself / herself permission to pursue a particular argumentative step. This use of *may* seems to have fallen almost completely out of use in the 2000s. The present findings do therefore not support the assumption of a reorganization of the modal paradigm, but rather suggest that more fine-grained changes in genre conventions are at least partly responsible for the decline of this modal.

### References

- Aarts, Bas, Sean Wallis and Jill Bowie (*forthc.*). 'Profiling the English verb phrase over time: modal patterns.' In: Irma Taavitsainen, Merja Kytö, Claudia Claridge and Jeremy Smith (eds.), *Developments in English: expanding electronic evidence*. Cambridge: Cambridge University Press.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The Evolution of Grammar. Tense, Aspect and Modality in the Languages of the World*. Chicago: The University of Chicago Press.
- COHA. Corpus of Historical American English*. 1810-2009. Compiled by Mark Davies at Brigham University. (accessible at <http://corpus.byu.edu/coha>)
- Collins, Peter (2009) *Modals and Quasi-modals in English*. Amsterdam: Rodopi.
- Kranich, Svenja (2011). 'To hedge or not to hedge: the use of epistemic modal expressions in popular science in English texts, English–German translations, and German original texts.' *Text & Talk* 31 [1], 77–99.
- Leech, Geoffrey & Nicholas Smith (2006) 'Recent grammatical change in written English 1961-1992: some preliminary findings of a comparison of American with British English.' In Antoinette Renouf and Andrew Kehoe (eds.), *The Changing Face of Corpus Linguistics*, Amsterdam: Rodopi, pp. 186-204.
- Leech, Geoffrey (2011) 'The modals ARE declining. Reply to Neil Millar's "Modal verbs in TIME: Frequency changes 1923–2006"', *International Journal of Corpus Linguistics* 14:2 (2009), 191–220.' *International Journal of Corpus Linguistics* 16:4 (2011), 547–564.
- Mair, Christian (2006) *Twentieth-Century English: History, Variation and Standardization*. Cambridge: Cambridge University Press.

### **Beyond Chi-Squared, ANOVAs and Dull Tables: Accurate Inferences and Effective Visualization with Hierarchical Modeling for Corpus and Experimental Data**

*Damien Littre*

*Université Catholique de Louvain, Belgium*

The shortcomings of chi-squared tests (Bestgen 2013) and, to a lesser extent, untransformed ANOVAs on frequencies (Jaeger 2008) are well-known in the psycholinguistics literature, as recent publications have shown that these techniques often yield inflated p-values and can therefore lead to incorrect inferences. Yet, in spite of such findings, these tests remain widespread in corpus linguistics, even as the recent rise of longitudinal corpora and mixed methods designs, which often involve repeated measures on the same participants and thus even less independence between observations, has made more pressing the need to go beyond these widespread yet problematic techniques. Using results from a study on the use of the English present progressive by intermediate-to-advanced French-speaking learners, this paper will show how hierarchical modeling (Gelman & Hill 2006) can advantageously replace less advanced methods and yield more reliable inferences for both (longitudinal) corpus data and experimental data. As will be shown with a hierarchical logistic regression model, this technique better accounts for the multilevel, non-independent structure of much linguistic data (e.g. words within texts within learners, or responses to items within learners). By allowing the effect of predictors to vary between individual learners and by recognizing that observations from the same text and/or learner share similar characteristics, hierarchical modeling helps uncover important information about individual learners, while avoiding the unit of analysis problem (Altman & Bland 1997) and thus also providing more reliable estimates of the average effects of the predictors. As the well-known difficulty in diagnosing and interpreting logistic regression models is compounded by the greater complexity of hierarchical models, this paper will also show how results can often be displayed graphically as more reader-friendly heat maps, forest plots and line graphs. While examples will feature the same hierarchical model as above, all visualizations will be easily generalizable to other models and research topics and should therefore be of interest to a larger audience than (potential) users of hierarchical models.

#### **References**

- Altman, D.G. & Bland J.M. (1997). Statistics notes: Units of analysis, *BMJ*, 314, 1874
- Bestgen, Y. (2013). Inadequacy of the chi-squared test to examine vocabulary differences between corpora. *Literary and Linguistic Computing, Advance Access*.
- Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434-446.

### Passive Voice as an Indicator of Colloquialization in the TIME Magazine Corpus

Sarah Nilsson  
Uppsala University, Sweden

Colloquialization is the process whereby written language stylistically approaches spoken-language norms through an increased acceptance of informal linguistic features. The *get*-passive, as in *he got arrested*, has been shown to be one such informal feature (Biber et al 1999: 476), and its increase in written texts is therefore taken as a marker of colloquialization (Mair 2006: 187). The *be*-passive, on the other hand, as in *he was arrested*, is a construction more typical of formal written language (Biber et al 1999: 476), and Leech (2004: 73) argues that any diachronic decrease in the frequency of *be*-passives should thus be taken as a “negative manifestation of colloquialization.” The aim of the present study is to look at these two passive constructions in a diachronic corpus of written American English to see how their distribution has changed over time, and to search for empirical evidence of the factors that govern the choice of construction. I investigate the use of the constructions in the TIME Magazine Corpus (Davies 2007). Newspaper language has been shown to be particularly prone to colloquialization (Hundt and Mair 1999: 236) due to its need to appeal to as wide a readership as possible. TIME magazine, an American news publication first printed in 1923 and still published today, has always been seen as one of the more informal and stylistically colloquial of its kind (Grunwald 1983), and the 100 million word TIME Magazine Corpus, which covers the period 1923-2006, is thus an excellent place to look for features of colloquialization. Because not all instances of *be* or *get* followed by past participles actually constitute central passives (Quirk et al 1985: 167-171), I examine large subsets of data manually in order to estimate frequency changes for true passives only. I expect a diachronic decrease in central *be*-passives and increase in central *GET*-passives, which would indicate colloquialization. I also study the qualitative factors which govern the choice of *be* versus *get* in central passives where both verbs could be used. The three factors investigated are adversativity, agentivity, and dynamicity. The *get*-passive has been associated with all of these factors (Chappell 1980; Huddleston and Pullum 2002: 1440-1443; Toyota 2008: 148-172). Adversativity describes a situation where something undesirable is happening to the passive subject, e.g. *he got killed*. Agentivity implies subject responsibility; in the case of *he got killed*, the choice of *get* supposedly conveys the added sense that he has brought his death upon himself. I code for animate subjects as an indicator of agentivity, as any “tests” for agentivity would be overly subjective. A weakening of the association with adversativity and agentivity would indicate that the *get*-passive has become more integrated into written English. Lastly, I separate stative and dynamic uses of the passives in order to ascertain whether *get* may disambiguate cases where *be*-passives might be interpreted as stative, as in *get fired* vs. *be fired*.

### References

- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Chappell, H. (1980). Is the *get*-passive adversative? *Papers in Linguistics* 13, 411 – 452.
- Davies, M. (2007–). *TIME Magazine Corpus: 100 million words, 1920s – 2000s*. Available online at <http://corpus.byu.edu/time/>.
- Grunwald, H. (1983). TIME at 60. *TIME*. Oct. 5, 1983.
- Huddleston, R. & Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hundt, M. & Mair, C. (1999). ‘Agile’ and ‘uptight’ genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4 (2), 221 – 242.
- Leech, G. (2004). Recent grammatical change in English: Data, description, theory. In K. Aijmer & B. Altenberg (eds.), *Advances in Corpus Linguistics*, 61 – 81. Amsterdam: Rodopi.
- Mair, C. (2006). *Twentieth-Century English: History, Variation and Standardization*. Cambridge: Cambridge University Press.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London & New York: Longman.
- Toyota, J. (2008). *Diachronic Change in the English Passive*. Hampshire: Palgrave MacMillan.

### The use of the items *of* and *to* in semi-prepared spoken English texts

Michael Pace-Sigge

University of Eastern Finland, Finland

This paper looks at the usage patterns found for the highly frequent items *of* and *to* in different sets of corpora. The paper disputes claims that these two items can merely be referred to as *prepositions* or that they carry little meaning in themselves. Instead, looking at historical roots (cf. Borström: 1965, Hook: 1975, O'Dowd: 1998) and earlier corpus linguistic research (Sinclair *et al.*: [1970] 2004), this paper aims to demonstrate that both *of* and *to* have clear semantic and pragmatic functions. This paper then, looks at how the items are typically found in corpora of transcripts of longer public speeches. By way of exemplification, the focus is on the transcripts of *BBC Reith Lectures*, *London School of Economics* (LSE) public lectures, and public lectures on key works of art at *National Museums Liverpool* (NML). Drawing on material first presented by Sinclair (1991), Gries (2003) and Hoffmann (2005) it will be shown that *of* and *to* are far less easy to classify than is generally assumed. This paper will show that *of* occurs in a far more stable and fixed mode than *to* (this mirrors earlier findings by Pace-Sigge, 2009). The two words are constituent parts of longer *lexical items*: this appears to reflect both the roots and the communicative functions they fulfil. The occurrence frequency for *of* and *to* are very similar in works of fiction, and in transcripts of academic lectures. In spoken British English, *to* is significantly more frequent; in the singlepresenter podcasts, however, *of* is notably more frequent than *to*. This makes public lectures unlike other spoken forms, yet written academic texts also show a clearly higher proportional use of the item *of* compared to the item *to*. This paper makes clear that it is crucial to recognise that *of* and *to* are not only highly frequent items by themselves; they are also highly prominent in the most frequent bigrams, trigrams and even longer clusters in prepared and semi-prepared spoken utterances. It is crucial too, to be aware of the fact (as Sinclair: 1990, and Stubbs: 1996, have pointed out) that certain word forms predominantly appear in one single construction. This needs to be taken into account in our understanding of language. The way *of* and *to* occur in podcasts have been found to vary, most crucially when single-speaker sessions are compared to multiple-speaker sessions and, as a whole can be seen as different from lectures given in an academic setting. The unconscious specific usage pattern of these two items as recorded in a particular subset of spoken language, mirrors findings by Biber (2000) and Hoffmann (2005) and can be seen to support for the idea that speakers are primed for preference and nesting patterns, as described in the *lexical priming theory* (Hoey: 2005).

### References

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan. [1999] 2007. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education.
- Biber, Douglas. 2000. Investigating Language Use through Corpus-Based Analyses of Association Patterns. In: Borström, Sverker. 1963. *The increasing frequency of the preposition about during the Modern English Period*. Stockholm: Almqvist & Wiksell.
- Borström, Sverker. 1965. *Studies on the use of the preposition of in the 15<sup>th</sup> century correspondence*. Stockholm: Almqvist & Wiksell.
- Gries, Stefan Th. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1, pp. 1–27.
- Hoey, Michael. 2005. *Lexical Priming. A new theory of words and language*. London: Routledge.
- Hoffmann, Sebastian. 2005. *Grammaticalization and English Complex Prepositions*. London: Routledge.
- Hook, J.N. 1975. *History of the English Language*. New York: The Ronald Press Company.
- O'Dowd, Elizabeth M. 1998. *Prepositions and Particles in English. A Discourse-functional Account*. New York/Oxford: Oxford University Press.
- O'Keefe, Michael Mc Carthy and Ronald Carter. 2007. *From Corpus to Classroom*. Cambridge: CUP. Pace-Sigge, Michael. 2009. Why TO is a weird word. Available from [www.academia.edu/4107981/Why\\_TO\\_is\\_a\\_weird\\_word](http://www.academia.edu/4107981/Why_TO_is_a_weird_word) (last accessed 10/11/2013)
- Pace-Sigge, Michael. 2013. *Lexical Priming in Spoken English Usage*. Houndmills, Basingstoke: Palgrave Macmillan.
- Sinclair, John M., Susan Jones and Robert Daley. [1970] 2004. *English Collocation Studies: The OSTI Report*. London: Continuum.
- Sinclair, John M. 1990. *Collins Cobuild English Grammar*. London/New York: Collins ELT.
- Sinclair, John M. 1991. *Corpus Concordance Collocation*. Oxford: OUP.
- Stubbs, Michael. 1996. *Text and corpus analysis. Computer-assisted analysis of language and culture*. Oxford: Basil Blackwell.

### Corpus Analysis of Gobbledygook

Gill Philip

University of Macerata, Italy

In July 2013, the British Civil Service issued a new style guide for its online documents, *Government Digital Service Content Principles*. In itself, this is nothing remarkable: all forms of publishing require style guides to ensure consistency across authors, text types and topics. What is remarkable, however, is the peculiarly prescriptive tone of the document, combining Grice's (1975) maxims with observations on undesirable language use that is reminiscent of Orwell's 1945 essay 'Politics and the English Language'. Readers of the document are urged not only to "use plain English" (§1.5), but also reminded that "... government 'buzzwords' and jargon [...] are too general and vague and can lead to misinterpretation or empty, meaningless text. *We can do without these words*" (ibid., my emphasis). In this contribution I discuss how the 37 'offending' expressions listed in the style guide are used in online policy documents, and attempt to connect this with (i) the choice of expressions listed, (ii) the definitions provided next to many of them, suggesting what their 'proper' meaning is, i.e. how they *should* be used, and (iii) if those explicitly listed as metaphors "to be avoided" are indeed metaphorical. Corpus linguists are well aware that it is the phrasal environment of a word that fixes its meaning; and they are also familiar with the notion that the frequent use of particular phraseologies can lead to delexicalisation, i.e. a 'loss of identity' of individual words relative to the form and function of the phrase of which they form part. So it should not come as a surprise that the words on the black list often appear in phrasal sequences as well as in compositional (or "open-choice", Sinclair 1991) language; and it is in such phrasal uses that their meaning does indeed seem to vanish into "empty, meaningless text". What needs to be investigated, however, is which factors cause meaning to slip from our grasp – whether it is found in all phrasal uses, or if it is possible to identify more precise indicators as to the *kinds* of phrase (e.g. syntactic role, composition, length, variability) that are associated with a loss of meaning, at least as far as this text genre is concerned. The data studied comprises the policy documents published in the year 2013, available from <https://www.gov.uk/government/policies>. WordSmith Tools (Scott 2008) was used to compile word lists from the documents' titles as well as from the complete corpus, as a means to verify the presence and frequency of the style guide's list of 'offenders' (the frequencies vary considerably from term to term, and some only occur in inflected forms). Once their presence was verified, each term was subjected to 'traditional', KWIC-concordance analysis in order to identify their meaning(s) and use(s) in the UK government policy documents; these were in turn compared with patterns in general language use using the BNC as a reference corpus. Findings to date indicate that civil service gobbledygook is a phraseological phenomenon that (i) is register- and genre-specific, and (ii) delexical rather than metaphorical.

### References

- Grice, H.P. 1975. Logic and Conversation. In Cole & Morgan (eds) *Syntax and Semantics vol. 3*, pp. 41-53. New York: Academic Press.
- Orwell, G. 1945. Politics and the English Language.
- Scott, M. 2008. WordSmith Tools version 5. Liverpool: Lexical Analysis Software.
- Sinclair, J.M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Government Digital Service Content Principles*. Available: <https://www.gov.uk/design-principles/style-guide> (accessed 30 July 2013).

### Particle placement in Late Modern English and Twentieth-century English

Paula Rodríguez-Puente  
University of Cantabria, Spain

English phrasal verbs can be intransitive, as in (1), or transitive, as in (2).

- (1) She **gave in**.
- (2) She **turned [on]** the lights **[on]**.

When transitive, the particle can be placed at either side of the DO, except when the latter is a pronoun, which usually precedes the particle (see (3) below), or when there exists a fixed order of constituents and no alternation is possible (see (4) and (5) below).

- (3) She **turned** them **on**.
- (4) She **gave up** hope [**\*up**].
- (5) He **invited** [**\*over**] his friends **over**.

Several studies have pointed out that particle alternation in PDE is not arbitrary and that a number of phonological, morphological, semantic and discourse-functional variables may influence the choice of the speaker (see, among others, Dehé 2002; Gries 2003; Lohse et al. 2004; Cappelle 2005, 2006), such as, for example, the stress pattern of the verb phrase, the phonetic shape of the verb, the NP type of the DO, the determiner, the length and complexity of the DO, the meaning of the verb phrase, the news value of the DO and/or the last mention of the referent of the DO. In his comprehensive analysis of particle placement, Gries (2003) has tested the majority of the variables traditionally said to influence particle distribution both individually (monofactorial analysis) and in connection to one another (multivariate analysis). His analysis was later on applied to children's language to assess whether the position of the particle might vary with the same features as the language of adults (see Diessel & Tomasello 2005; Gries 2011). Haddican & Johnson (2012), in turn, focus on particle placement across English dialects. Elenbaas (2013) examines the possible motivations behind the choice of one or another order during the Middle English and Early Modern English periods. However, an approach to particle placement in Late Modern English and Twentieth-century English has not been attempted yet. This paper traces the features which may affect the position of the particle between 1650 and 1990 to compare them to those identified for other stages of the language. To this purpose, examples of phrasal verbs were extracted from the British component of ARCHER 3.1 (*A Representative Corpus of Historical English Registers*), in which a total of 2,420 examples of transitive combinations were found instantiating both the VOP (Verb-Object-Particle) and VPO (Verb-Particle-Object) orders.

### References

- Cappelle, Bert. 2005. *Particle patterns in English: A comprehensive coverage*. PhD dissertation, Katholieke Universiteit Leuven. Faculteit Letteren. Departement Linguïstiek.
- Cappelle, Bert. 2006. Particle placement and the case for 'allostructions'. In Schönefeld, Doris (ed.), *Constructions all over: Case studies and theoretical implications* [special issue of the e-journal *Constructions*]. Available at: <http://www.constructions-online.de/articles/specvol1/683>
- Dehé, Nicole. 2002. *Particle verbs in English. Syntax, information structure and intonation (Linguistik Aktuell / Linguistics Today 59)*. Amsterdam & Philadelphia: John Benjamins.
- Diessel, Holger & Michael Tomasello. 2005. Particle placement in early child language: A multifactorial analysis. *Corpus Linguistics and Linguistic Theory* 1.1: 89-111.
- Elenbaas, Marion. 2013. Motivations for particle verb word order in Middle English and Early Modern English. *English Language and Linguistics* 17.3: 489-511.
- Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. London and New York: Continuum Press.
- Gries, Stefan Th. 2011. Acquiring particle placement in English: A corpus-based perspective. In Guerrero-Medina, Pilar (ed.), *Morphosyntactic alternations in English: Functional and cognitive perspectives*, 235-263. London & Oakville, CT: Equinox.
- Haddican, Bill & Daniel Ezra Johnson. 2012. Effects of particle verb alternation across English dialects. *University of Pennsylvania Working Papers in Linguistics* 18.2: 31-40.
- Lohse, Barbara, John A. Hawkins & Thomas Wasow. 2004. Domain minimization in English verb-particle constructions. *Language* 80.2: 238-261.

### Densification in Nineteenth-century News Discourse: Noun Phrases with Premodifying Nouns

Erik Smitterberg

Uppsala University, Sweden

Previous research on Late Modern and Present-day news texts has revealed that newspaper English has undergone *densification*: the information density of the linguistic material has increased, as evidenced by rising frequencies of, for instance, noun + noun sequences and s-genitives (Leech et al. 2009: 211–226). Biber (2003: 170) notes that its reliance on “a dense use of nouns and integrated noun phrase constructions” makes present-day newspaper English “very similar to academic prose” in this regard. This similarity is of particular interest because, unlike academic writing, newspaper English exhibits tendencies towards more spoken-like usage in other respects (see e.g. Hundt and Mair 1999).

The aim of this corpus-based study is to answer the following research questions:

1. To what extent do noun phrases that include one or several premodifying nouns contribute to densification in British news texts from the 1800s?
2. What types of premodifying nouns and head nouns occur in such noun phrases, and what are the semantic relationships between them?

Previous research (e.g. Biber and Gray 2012: 322–323) indicates that noun phrases with nouns as premodifiers have become considerably more frequent in newspapers and scientific prose since 1800. Biber and Gray (2012: 326) attribute this change chiefly to the informational purpose of news and science texts, with the specialized readership of academic writing being of secondary importance. However, more research is needed on what types of nouns occurred in such noun phrases. For instance, while Leech et al. (2009: 259–261) find that the use of titular nouns such as *Mr* decreased in late-twentieth-century English, a change which they link to democratization, my results indicate that their frequency rises considerably during the nineteenth century. The present study is based on the Corpus of Nineteenth-century Newspaper English (CNNE), which includes newspapers published in England during the periods 1830–1850 and 1875–1895. Nineteenth-century newspaper English is of considerable interest to historical linguists for several reasons. First, the newspaper was an important medium during the 1800s (see e.g. Lee 1976: 18). Secondly, considerable development took place in this genre; for instance, Brown (1985: 1) claims that “[t]he news’ as we understand it is a nineteenth-century creation”. Thirdly, given that research on newspaper English occupies a prominent place in studies of language change in the twentieth century, it is of obvious interest to examine the extent to which developments that were in progress during the 1900s can be shown to occur in the preceding century as well. The study is based on a tagged version of CNNE, from which noun-phrase heads preceded by nouns could be retrieved automatically. The resulting concordances were then inspected manually in order to include only relevant instances in the counts and in order to classify the relevant instances on parameters such as the type of premodifying noun. I expect to find an increase in the frequency of most types of premodifying nouns.

### References

- Biber, D. 2003. “Compressed Noun-phrase Structures in Newspaper Discourse: The Competing Demands of Popularization vs. Economy”. In: Aitchison, J., and Lewis, D.M. (eds.), *New Media Language*. London: Routledge, 169–181.
- Biber, D., and Gray, B. 2012. “The Competing Demands of Popularization vs. Economy: Written Language in the Age of Mass Literacy”. In: Nevalainen, T., and Traugott, E.C. (eds.), *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, 314–328.
- Brown, L. 1985. *Victorian News and Newspapers*. Oxford: Clarendon Press.
- Hundt, M., and Mair, C. 1999. “‘Agile’ and ‘Uptight’ Genres: The Corpus-based Approach to Language Change in Progress”. *International Journal of Corpus Linguistics* 4 (2), 221–242.
- Lee, A.J. 1976. *The Origins of the Popular Press in England: 1855–1914*. London: Croom Helm.
- Leech, G., Hundt, M., Mair, C., and Smith, N. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.

### Pattern and appraisal: towards a local grammar of JUDGEMENT

Hang Su

University of Birmingham, United Kingdom

It is pointed out that even though the grammar pattern project (Francis et al 1996, 1998; Hunston & Francis 2000) was not envisaged as something related to evaluation, the patterns which in some way convey evaluative meaning inevitably attract our attention (see Hunston 2011: 120). Closely related to pattern grammar is the concept of local grammar which, recently, has also been applied into the study of evaluative language (e.g., Hunston & Sinclair 2000; Hunston 2003, 2011; Bednarek 2008). The results suggest that the development of local grammars of evaluation would be a successful and useful enterprise. The aspect of evaluative language in this paper is restricted to expressions associated with the evaluation of human beings; this restriction facilitates local grammar analysis. In addition, the assessment of human beings has been shown to play a crucial role in social interaction, for example as evidence for the retention and/or promotion in careers (Hyon 2011; also see Martin & White 2005: 53). As for how human beings are evaluated, the JUDGEMENT system identified in the Appraisal Theory systematises the judgemental meanings into five semantic categories, i.e., 'normality' (how unusual someone is), 'capacity' (how capable they are), 'tenacity' (how resolute they are), 'veracity' (how truthful someone is) and 'propriety' (how ethical someone is) (see Martin & White 2005: 52-56). It is arguable that the JUDGEMENT system could be considered as the most fine-grained framework from a linguistic perspective currently available for accounting for the evaluation of human behaviour. This study, bringing together the concept of local grammar and the concept of Judgement, aims to address the question as to what extent and how grammar patterns can be used to parse Judgement expressions, i.e., to build local grammars. The grammar patterns in this study are confined to adjective complementation patterns because adjectives are the word class most frequently associated with evaluation (Martin & White 2005: 58; Hunston 2011: 129). The corpus consulted for the current study is the Corpus of Biography which is compiled drawing texts from the British National Corpus and accessed using the BNCweb interface - the CQP edition (see Hoffman et al 2008). The Judgement expressions are then semi-automatically extracted through performing queries of adjective complementation patterns in the corpus. The local grammar elements (e.g., Evaluator, Target, Scope, etc.) that are involved in judgement are subsequently proposed for a local grammar analysis. It is demonstrated that grammar patterns are helpful in identifying expressions that are associated with one specific meaning area and in specifying the structure of these expressions, which are crucial to local grammar analysis. It is concluded that grammar patterns are conducive to developing local grammars, which is exemplified by the successful development of the local grammar of JUDGEMENT with grammar patterns being taken as the starting point. It is also illustrated that local grammars has the potential to enable the automatic parsing of appraisal expressions.

### References

- Bednarek, M. (2008). *Emotion Talk Across Corpora*. New York: Palgrave Macmillan.
- Francis, G., Hunston, S. and Manning, E. (1998). *Collins Cobuild Grammar Patterns 2: Nouns and Adjectives*. London: Harper Collins.
- Hoffman, S. et al (2008). *Corpus Linguistics with BNCweb - A Practical Guide*. Frankfurt: Peter Lang.
- Hunston, S. (2003). Frame, phrase or function: A comparison of frame semantics and local grammars. In *Corpus Linguistics 2003*, Lancaster University CD-ROM.
- Hunston, S. (2011). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London: Routledge.
- Hunston, S. & Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Hunston, S. & Sinclair, J. (2000). A local grammar of evaluation. In S. Hunston & G. Thompson (eds.) *Evaluation in Texts: Authorial Stance and the Construction of Discourse*, pp.74 – 101. Oxford: Oxford University Press.
- Hyon, S. (2011). Evaluation in tenure and promotion letters: constructing faculty as communicators, stars, and workers. *Applied Linguistics*, 32 (4), 389 – 407.
- Martin, J. R. & White, P. (2005). *The Language of Evaluation: Appraisal in English*. New York: Palgrave Macmillan.



### Investigating interdisciplinary research discourse

*Paul Thompson*

*University of Birmingham, United Kingdom*

Recent corpus research (eg, Hyland 2000, Fløttum, Dahl & Kinn 2006) has revealed much about linguistic variation in and between disciplinary discourses, but little as yet has been written about the features of interdisciplinary research writing. It is probable that researchers working in an interdisciplinary research context frame their messages for a broader audience than they would do when writing for a discipline-specific journal, and consequently we predict that the linguistic features of papers in interdisciplinary research journals are distinctively different from those in monodisciplinary journals. In this talk we report on the first stages of an ESRC funded project carried out at the Centre for Corpus Research, Birmingham, in collaboration with the publisher Elsevier, in which we investigate, using corpus linguistic techniques, the entire holdings (1990 onwards) of a successful journal in an interdisciplinary field: *Global Environmental Change*. The corpus consists of 676 research articles, and approximately 5.5 million tokens. We have clustered the articles variously on different dimensions using Biber's Multidimensional Analysis in order to capture congruences of features between them, and in later stages of the project we will focus on describing certain features through close textual analysis (citation practices, and the representation of writers and readers in the texts). The research questions addressed in the preliminary stage of the project are: - When the articles in the *Global Environmental Change* journal are 'clustered' according to linguistic features, do the clusters that are identified represent communities of practice, disciplinary or otherwise? - Do the clusters remain the same over time? - Do the general linguistic profiles of GEC articles (established through multidimensional [MD] analysis) resemble those of other IDR journals? - Are they distinct from those of specialist disciplinary journals? In order to address the third and fourth of these questions, comparative analyses of corpora of research articles from ten other Elsevier journals are conducted, for the period 2001-2010. Five of the journals are interdisciplinary, and five are specialist disciplinary journals. Through these analyses, we seek to determine the features of writing in the disciplines that come together to work on global environmental change issues, and also to see whether research articles in other interdisciplinary areas (with different collaborations) cluster in the same ways that the GEC articles do, or not. At this point, we cannot present conclusive findings but we expect to find evidence of change over time, as conventions for communication in interdisciplinary settings evolve, and also we expect MD analysis to find variation between the interdisciplinary journals. Our findings are augmented with data from interviews and surveys conducted with editors, reviewers and authors.

### Left-dislocated Noun Phrases in the recent history of English: genre distribution and discourse functions

David Tizón-Couto  
University of Vigo, Spain

This investigation finds a significant decrease of Left Dislocated Noun Phrases (LDNPs such as *After it was dark any Ship that came to us we engaged them*) from the ME period onwards in the examples extracted from the Penn(-Helsinki) parsed corpora of Middle English (PPCME2), Early Modern English (PPCEME), Modern British English (PPCMBE) and the Early English Correspondence Corpus (PCEEC). The decrease (ME: 10.7, EModE: 2.4 and LModE: 1.3 instances per 1,000 words) could be explained as a combination of several factors such as the establishment of the syntactic and orthographical bases of the sentence (Culpeper and Kytö 2010:168) or prescriptivist criticism as regards words “put out of their proper order” (Michael 1970:471). The analysis focuses on the genre distribution and discourse functions of the 989 LDNPs extracted from the Modern English corpora (PPCEME, PPCMBE and PCEEC). As for the distribution across historical genres of a word-order variant such as LD, commonly associated with spoken language (Geluykens 1992, Prince 1997, Gregory and Michaelis 2001), the findings suggest that the frequency of LDNPs in speech-like texts (letters and diaries) has proven lower (n.f. 0.13) than in speech-purposed (drama and sermons; n.f. 0.94) or in mixed (fiction and trial proceedings) and written (biography, educational treatise, handbook, history, law, philosophy, science and travelogue; n.f. 0.64) genres since the ME period. However, concerning their discourse function, those LDNPs that carry out an affective or highlighting function (in the sense of Keenan-Ochs and Schieffelin 1976:245; Geluykens 1992:95; Kim 1995:285), rather than a more neutral discourse-organisational role (Netz et al. 2011), have been found to be more likely in speech-like (58.2%) and speech-purposed (55.8%) genres (only 35.2% in mixed genres and 37.5% in written genres). Additional variables suggest that the form and function of LDNPs reflect differences between speech-related and purely written genres. For instance, a tally of the element which may precede LDNPs (usually a conjunction or a complementiser) shows that 34.2% of all instances of LDNPs preceded by a conversational item such as clause-level *and* (Culpeper and Kytö 2010:166) is attested in letters and diaries (by far the highest percentage for any genre). In addition, bare LDNPs (i.e. with no previous conjunction) are most frequent in speech-purposed (70.2% in sermons and drama) and mixed texts (67.6% in fiction and trial proceedings), while those that have a previous marker of any kind are more likely to convey a highlighting functional shade (44.1% of the total for affective roles) rather than a neutral discourse-organisational function (35.8% of the total for discourse-organisational roles). These two latter findings suggest, respectively, that LDNPs seem to have been particularly useful as deictic rhetorical devices in written-to-be-spoken texts such as sermons and drama, and that other conversational clause-initial markers such as *and* interacted more regularly with LDNPs when the authors/speakers felt freer to innovate (i.e. in genres with less editorial control).

### References

- Anagnostopoulou, Elena, Henk van Riemsdijk and Frans Zwarts, eds., 1997. *Materials on left dislocation*. Amsterdam: John Benjamins.
- Culpeper, Jonathan and Merja Kytö. 2010. *Early Modern English dialogues. Spoken interaction as writing*. Cambridge: Cambridge University Press.
- Geluykens, Ronald. 1992. *From discourse process to grammatical construction: on left dislocation in English*. Amsterdam: John Benjamins.
- Gregory, Michelle and Laura Michaelis. 2001. Topicalization and left-dislocation: a functional opposition revisited. *Journal of Pragmatics* 33: 1665-1706.

### The Empirical Strikes Back: Corpus Linguistics & Adaptation Theory

Joseph Trotta

University of Gothenburg, Sweden

Though corpus methodologies and other empirical approaches have transformed linguistics in recent decades, such methods have not been widely welcomed among researchers in other liberal arts disciplines. In particular, literary scholars have strongly questioned the usefulness of quantitative approaches, with reactions to Franco Moretti's influential *Graphs, Maps, Trees* (2005) clearly demonstrating an ever-growing schism in the field. Some view work like Moretti's as offering confirmable and generalizable 'scientific' results, whereas others, the (apparent) majority, see this hard-science angle on literary analysis as contributing only trivial insights at best, at worst it is viewed as circular since, according to the critics, its results depend on dubious coding and labeling based on a priori understandings of the very concepts under investigation. With this background in mind, the present work-in-progress study interrogates the usefulness of corpus tools in analyzing adaptations (and to some extent remediations), in this case, chiefly of novels to film, but other alternatives are also relevant, such as film adaptations of plays, comic books and video games, to name a few. With the corpus approach advocated here, and with the help of software such as WordSmith and Wmatrix, digitalized versions of the original text and the adapted text are compared in terms central techniques in corpus linguistics, such as key words, semantic fields, lexical density, type-token ratios, and so on, which, in turn, shed light on central issues in adaptation studies such as how elisions, interpolations, 'translations' and interpretations are handled. The study focuses mainly on the 2009 film adaptation of Cormac McCarthy's *The Road*, using examples from this work as a platform for larger discussions on what corpora can offer to the field. Though *The Road* is the primary text examined, cursory references and comparisons are made to other well-known adaptations such as *A Clockwork Orange*, *Fight Club*, *The Heart of Darkness*/*Apocalypse Now*, *V for Vendetta*, among others.

#### References

Moretti, F. (2007). *Graphs, maps, trees: abstract models for literary history*. London; New York: Verso.

### Repairing with hands: comparing gestures of advanced-level Japanese learners of English with basic-I

Keiko Tsuchiya  
Tokai University, Japan

This presentation reports a small scale study comparing gestures of advanced-level Japanese learners of English in a pair conversation with basic-level counterparts in relation to use of *repair* strategies using a time-aligned multimodal corpus approach. Goldin-Meadow (1999) categorises hand gestures into four functions: *iconic gestures* (IG), which describe a picture that the speaker has in mind; *metaphoric gestures* (MG), which are more abstract than iconic gesture and describe speakers' thoughts or idea; *beat gestures* (BG), which can be used to emphasise what the speaker is saying; and *deictic gestures* (DG), in other words, pointing gestures. In more recent studies, Adolphs and Carter (2013), for example, indicate the close relationship between use of particular gestures and contexts. In terms of *repair*, Schegloff (2007) defines *repair* in conversation as 'efforts to deal with trouble-sources or repairables – marked off as distinct within the ongoing talk' (ibid: 101). Schegloff, Jefferson and Sacks (1977) identify four types of repair: self-initiated self-repair (SISR), other-initiated self-repair (OISR), self-initiated other-repair (SIOR), and other-initiated other-repair (OIOR) (also cited in Hutchby and Wooffitt 2008). Referring these previous studies in gestures and repair strategies, three research questions are addressed here: (1) are there any differences in the numbers of word count and the lengths of speaker turns between the two pairs?, (2) how many and what kind of gestures do the learners use in the conversations, and are there any differences between the two levels?, and (3) how do these gestures relate to their repair strategies? However, the main purpose of this preliminary study is to establish and construe the feasibility of the research method.

#### Research method and data:

Two five-minutes-long pair conversations, one of which is a pair of advanced learners (B1 in CEFR) and the other is basic level learners (A2), were recorded at a university in Japan. These two data sets were compared in terms of word count, speaker turns, functions of gestures and repair strategies using a time-aligned multimodal corpus. Each participant's utterances and gestures were transcribed and time-stamped using a software package, Transana (Fassnacht and Woods 2002). Annotations of repair strategies were added manually for qualitative analysis. This methodology was developed based on recent studies adopting multimodality and the integration of corpus linguistics with conversation and discourse analytic approaches (Adolphs 2008; Knight 2011; Tsuchiya 2013).

#### Results:

The results are summarised as follows:

- (1) In terms of turn-taking structure, fewer but longer turns were observed in the advanced pair compared with the basic level pair.
  - (2) More gestures were observed in the basic-level pair than the advanced-level pair. More than half of the gestures in the advanced pair is MG, while the basic learners used IG most in addition to MG and DG.
  - (3) The numbers of instances of repair between these two pairs are almost the same. However, SISRs tend to collocate with MG in the advanced-level pair, but with IG and DG in the basic-level pair.
- In sum, the learners' levels of English seem to affect their use of gestures in repair sequences.

#### References

- Adolphs, S. 2008. *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse*. Amsterdam: John Benjamins Publishing Company.
- Adolphs, S. and R. Carter 2013. *Spoken Corpus Linguistics: From Monomodal to Multimodal*. Abingdon: Routledge.
- Fassnacht, C. and D. Woods 2002. Transana. Version 2.12 - Win. Wisconsin: University of Wisconsin-Medison.
- Goldin-Meadow, S. 1999. "The role of gesture in communication and thinking " *Trends in Cognitive Sciences* 3/11: 419-429.
- Hutchby, I. and R. Wooffitt 2008 [1988]. *Conversation Analysis*. Cambridge: Polity Press.
- Knight, D. 2011. *Multimodality and Active Listenership*. London: Continuum.
- Schegloff, E., A. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis, Volume 1*. Cambridge: Cambridge University Press.
- Schegloff, E., A., G. Jefferson and H. Sacks 1977. "The preference for self-correction in the organization of repair in conversation." *Language* 53: 361-382.
- Tsuchiya, K. 2013. *Listenership Behaviours in Intercultural Encounters: A Time-aligned Multimodal Corpus Analysis*. Amsterdam: John Benjamins Publishing Company.

### What happens in translation?

Åke Viberg

Uppsala University, Sweden

A comparison of original and translated English and Swedish texts containing verbs meaning SIT, STAND and LIE in the English Swedish Parallel Corpus (ESPC)? The present paper combines the approach of contrastive studies (Johansson 2006) with certain aspects of translation studies (Malmkjær & Windle 2011, Munday, J 2008). The postural verbs in English and Swedish are historical cognates which still are very similar in form: *sit/sitta*, *stand/stå* and *lie/ligga*. In spite of that, they are semantically different in many respects (Viberg 2013). The relative frequency of the Swedish verbs is also significantly higher, if the verbs are compared pair by pair. These contrastive relationships form the point of departure for the present study which is concerned with the differences between original and translated texts in the English Swedish Parallel Corpus (ESPC, see Aijmer & Altenberg 2000). By comparing the overall frequencies of the verbs in original and translated texts in the ESPC, it can easily be shown that all three English verbs are overrepresented in translated English texts, whereas the corresponding Swedish verbs are underrepresented. All of these differences are statistically significant. The present study attempts to take the analysis one step further by looking at various uses (or senses) of the verbs. There is a tendency that the differences between originals and translations are more pronounced when the verbs have a human subject than when they have a (concrete) inanimate subject. Examples with abstract subjects and metaphorical uses represent special cases that need to be treated separately in the analysis. It is interesting to do a more fine-grained analysis and look at what happens when the Swedish verbs are used in special constructions such as the presentative construction (*Det ligger en bok på bordet* 'There is a book on the table') and in the so-called pseudo-coordination with a progressive meaning (*Ann sitter och läser* 'Ann is reading'. The progressive meaning is often not as clearly reflected in the translation as in this example.) One problem with the more fine-grained analysis is that the number of examples often is too low to allow statistical testing. The study will also be concerned with restructuring. A relatively frequent type of restructuring is the shift from a spatial PP in the English version to a relative clause containing a postural verb +PP in the Swedish version as in (1).

(1) She gave the children the last of the fishfingers **from the bottom of the freezer**. (FW1)

Hon gav barnen de sista fiskpinnarna **som låg längst ner i frysen**. [that lay furthest down in the freezer]

The ESPC is bi-directional and makes several types of comparison possible: Swedish original –English original, original – translation in the same language. This is important for the methodology of language comparison. Corpus-based typological studies (or multilingual contrastive studies) for practical reasons are often mono-directional and based on source texts in only one language. (Viberg 2013 is based on Swedish originals translated into other languages. Harry Potter is a popular source text in typological studies.) A comparison will be made between the usage pattern of *sit*, *stand* and *lie* in English original and translated texts to see to what extent the methodology based on mono-directional translation corpora gives a correct picture of contrastive-typological relationships.

### References

- Altenberg, B. & Aijmer, K. 2000. The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies, in: C. Mair & M. Hundt (eds.) *Corpus linguistics and linguistic theory*, pp. 15-33, Rodopi, Amsterdam – Atlanta/GA.
- Johansson, S. 2007. *Seeing through multilingual corpora*. Amsterdam: Benjamins.
- Malmkjær, K. & Windle, K. Eds. 2011. *The Oxford handbook of translation studies*. Oxford : Oxford University Press.
- Munday, J. 2008. *Introducing translation studies : theories and applications*. 2nd ed. London : Routledge.
- Viberg, Å. Posture verbs. A multilingual contrastive study. *Languages in Contrast* 13:2, 139- 169

### Rebalancing corpora

Sean Wallis

University College London, United Kingdom

A key distinction between a **natural experiment**, such as a case study or corpus analysis, and a **true experiment** is that in the latter case, the researcher is able to select participants for each experiment, and thereby balance participants demographically. Thus a researcher might ensure that, in an experiment looking at the effect of age on a particular response, each age group was comprised of equal numbers of men and women (or failing this, in a constant ratio). The purpose of this requirement is to ensure that an observed effect attributed to age is not, in fact, potentially gender-related. However, one drawback is that the more variables taken into account, the more difficult it is to balance samples and obtain enough participants. This paper applies the problem of obtaining a balanced sample in corpus linguistics. A corpus sample is not 'representative' *a priori*, but must be representative of something – a **random sample** of the language, a **typological sample** including a wide range of text types, or a **stratified sample** including all permutations of each demographic variable. Each approach has its advantage: a genuine random sample provides the greatest generalisation power to comparable speakers/authors, a typological sample permits investigation of genre-specific phenomena, and a stratified sample allows sociolinguistic phenomena to be readily compared by partitioning data. We suggest ICE-GB and DCPSE are typical of many 'balanced corpora' in that they were collected principally as a typological sample. They were not explicitly stratified, although an attempt was made "to include as full a range as possible of the social variables which define the population" (Nelson, 1996: 28). We begin by examining these corpora empirically. We examine distributions of word count proportions by three key variables: **genre**, **gender** and (for DCPSE) **time**. For example, the proportion of words uttered by women is near-identical for written and spoken ICE-GB subcorpora, but the proportion varies widely *within* both subcategories. As a result, when researchers compare technology writing (0% women) with creative writing (45% women) they may mistakenly attribute gender effects to genre. Although we have examined proportions of **words**, an experimental sample is more precisely defined as consisting of a series of **cases** (Nelson *et al.* 2002, Ch. 9), where each "case" consists of a single instance of the linguistic phenomenon under investigation, e.g. clause patterns, modal verbs, etc. To construct an experimental sample a researcher performs a search procedure to extract only those phenomena under investigation and their permutations. We propose that corpus researchers should be able to identify the extent to which this experimental sample is demographically uneven, by subdividing their sample by multiple demographic variables. Secondly, they need transparent analytical methods to address this unevenness (noting that some data, like ICE-GB's female technology writers, may be unavailable). We consider a number of different approaches to this analysis problem including multivariate analysis and logistic regression, and a step-wise approach working from first principles, which we offer as a more transparent and less 'black box' approach to data analysis than automatic methods.

### References

- Nelson, G. 1996. 'The Design of the Corpus' in Greenbaum, S. (ed) *Comparing English Worldwide*. Oxford: Clarendon Press.
- Nelson, G., Wallis, S.A. and Aarts, B. 2002 (eds). *Exploring Natural Language*. Amsterdam: John Benjamins.

### The semantic patterning of grammatical key words: a cross-institutional, cross-disciplinary study of undergraduate writing

Karin Whiteside

University of Warwick, United Kingdom

This paper is a progress report on PhD research which takes a corpus-driven approach to analysis of lexico-grammatical features of successful undergraduate writing across two degree programmes, History and PIR (Politics and International Relations), at two UK HE institutions. A number of studies have investigated disciplinary distinctions in student writing (e.g. Samraj, 2008; Bruce, 2010; Nesi and Gardner, 2012), but there has been little work to date examining lexico-grammatical features of student writing which may be linked to discipline, nor has there been much research involving detailed contrastive analysis of student writing in subjects which sit quite closely on the disciplinary spectrum. History and PIR are relatively closely-related 'soft' disciplines (Becher and Trowler, 2001) sharing the 'the Essay genre family' (Nesi and Gardner, 2012) as the principle genre in which students write. The cross-institutional dimension of the study is intended to measure the extent to which factors other than discipline may explain differences found in student-level writing. The project follows broadly the approach taken in Groom's (2007) study of disciplinary difference in professional academic writing, that of choosing salient grammatical words as the focus for close analysis. Four discipline-specific specialised corpora were created consisting of recent successful third-year writing from the two target disciplines and institutions (three comprising approximately 90,000 words, one comprising approximately 50,000 words). Using Wmatrix, keyness analyses of the four corpora were undertaken using as a reference corpus the 'BNC Sampler Written'. Items were manually identified and lists were collated of 'grammatical' words (Scott and Tribble, 2006; Groom, 2007) including those that were key:

- i. across all four corpora (most saliently, *of, that, as, and this*)
- ii. in one discipline across both institutions (*is, an, can, its and does* in PIR; *was, were, his and their* in History)
- iii. in only one institution across both disciplines (*by, these and through* in Institution X; *more* in Institution Y)

A qualitative, bottom-up approach is taken to analysis of items on these lists. Groom's (2007, p102) notion of 'semantic motif' is adopted as a starting point for categorisation since the primary goal of analysis of the phraseological patterning of each key word is semantic rather than grammatical. Groupings and resulting categories for each key word are generated from close analysis of randomly selected concordance lines. For each target key word a 300-line random sample is taken from a corpus created by pooling all four sub-corpora. Once firm categories have been established, an analysis of a 300-line random sample from each disciplinary and institutional sub-corpus is undertaken applying pre-generated categories and adding any new categories that emerge. The paper will report results of concordance analyses completed to date.

#### References

- Becher, T. and Trowler, P.R. (2001). *Academic Tribes and Territories: Second Edition*, Buckingham: Open University Press
- Bruce, I. (2010). Textual and discursive resources used in the essay genre in sociology and English. *Journal of English for Academic Purposes*, 9, pp153-166
- Groom, N. (2007) Phraseology and epistemology in humanities writing: a corpus-driven study. Unpublished PhD thesis submitted to the University of Birmingham.
- Groom, N. (2010) 'Closed-class keywords and corpus-driven discourse analysis.' In M. Bondi and M. Scott (eds.) *Keyness in Texts*. Amsterdam and Philadelphia: John-Benjamins.
- Nesi, H., and Gardner, S. (2012) *Genres across the Disciplines: Student writing in higher education*. Cambridge: Cambridge University Press
- Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes*, 7, pp55-67
- Scott, M. and Tribble, C. (2006). *Textual Patterns: Key Words And Corpus Analysis in Language Education*, Amsterdam: John-Benjamins

### Twitter's affordances for animation of reported speech

Peter Wikström  
Karlstad University, Sweden

Reported speech is a common feature of conversational spoken interaction, but has also emerged as an interesting feature of computer-mediated communication. The present study investigates one specific aspect of reported speech, namely animation (Tannen, 2007), as it occurs in a corpus of data from the social networking service Twitter, in relation to a notion of communicative affordances. In line with the theme of ICAME 35, this study locates its interest in an intersection of fields—specifically corpus-based discourse analysis and computer-mediated communication or new media studies. Broadly, *animation* is any strategy used by the speaker or author of a speech report to represent or dramatize non-verbal qualities (e.g. quality of voice or facial expression) on behalf of the purported originator of the reported speech. By means of animation, a quality of voice or the like is not so much described as enacted. Interestingly, animation of reported speech happens not only in face-to-face interaction, but also in typewritten modes of language use online ([Author1], 2013). Communicative *affordances*, in turn, may be broadly defined as structural enablings or constraints of modes and behaviors of communicative interaction (Gibson, 1986; Hutchby, 2001). The aim of the study is to identify and illustrate the range of devices for animation afforded by the medium of Twitter, with a focus on reported speech framed by quotative *like*. As an example, a Twitter user who writes *Ha ha ha when the chinese screamed they were like YAAA OHHHHHHHHHHHHH* is employing two frequently used devices for animation, namely character reduplication and case shifting. In this case, the devices are reasonably understood as animating a vocal quality, dramatically representing the length and intensity of the reported scream. In line with the aim of the study, three research questions are posed. Firstly, what range of devices for animation of reported speech is afforded on Twitter? Secondly, how frequent are the various devices for animation? Thirdly, what are the main communicative functions of the various devices for animation? The material used to answer these questions is a random sample of one million tweets (approximately 13 million words/tokens), collected at regular intervals throughout 2012, and filtered to represent general use of English on Twitter (IllocutionInc.com, ICAME 35 call for papers 2013; C. Darwin, personal communication, May 30, 2013). Preliminary findings indicate that Twitter, while being a formally relatively constrained medium, affords a variety of multimodal, but mostly typographical, devices for animation, such as case variation, emoticons, and embedded images. The devices can serve a wide range of context-sensitive functions, but the general function is often to dramatize some aspect of a bite-sized narrative. The results are likely to have implications for a general linguistic understanding of the forms taken and functions served by reported speech, as well as for ongoing debates about the modal characteristics of online language.

### References

- [Author1]. (2013). & *she was like "O\_O": Animation of reported speech on Twitter*. Manuscript submitted for publication.
- Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hutchby, I. (2001). *Conversation and technology: From the telephone to the internet*. Cambridge, MA: Polity.
- IllocutionInc.com. (2013). Illocution Incorporated Corpora. Retrieved December 1, 2013, from <http://illocutioninc.com/Corpora/>
- Tannen, D. (2007). *Talking voices: Repetition, dialogue, and imagery in conversational discourse* (2 ed.). Cambridge: Cambridge University Press.



### Left dislocation and fronting in spoken varieties of English

*Claudia Winkle*

English word order has often been described as fixed. Yet, there are different syntactic strategies to transform the canonical word order (SVO) for certain pragmatic effects, for example, for organizing the information flow and creating coherence or for putting emphasis on certain elements in a clause. Left dislocations (e.g. *this cat she is fourteen*) and fronting constructions (e.g. *my birthday party you arrange*) belong to this group of pragmatically specialized constructions resulting in an unusual or marked word order. Previous studies have discussed the constructions' discourse functions (e.g. Geluykens 1992, Gregory & Michaelis 2001) and their distribution in different varieties of English. It has been found that they occur particularly frequently in the 'New Englishes' (e.g. Lange 2012) and the 'Celtic Englishes' (e.g. Filppula 2009). It is difficult to directly compare these results, since the studies are based on different datasets and on different definitions of the constructions in question. The present study provides a more comprehensive view by systematically investigating comparable data from a number of first- and second-language varieties of English. In my report I present preliminary results of work in progress, comparing the use of left dislocations and fronting constructions in British English, Irish English, Indian English and Hong Kong English. The data used for the analysis consists of face-to-face interactions and telephone conversations from the private dialogues sections in the relevant components of the International Corpus of English (ICE). Basing my analysis on the ICE family allows me to directly compare the frequencies and usage patterns found in the respective corpora, since they have all been compiled according to the same design. I expect to find quantitative and/or qualitative differences across the varieties analysed and I hypothesize that substrate influence plays a major role in the observed usage patterns. My investigations so far show that speakers of Indian English and Hong Kong English use left dislocations much more frequently than speakers of British English and Irish English. The number of occurrences in the Irish English corpus is very low in comparison to the (South) Asian English corpora. What is also interesting to note is that for fronting constructions high frequencies can only be attested for Indian English but not for Hong Kong English. How can this diverging behaviour of speakers from India and Hong Kong be accounted for? What motivates speakers of Hong Kong English to use left dislocations very frequently but not the - at least syntactically - quite similar fronting constructions? Possible answers to these questions might be found in the influence of the substrate languages and, in addition, in the different discourse functions of the two constructions. The use of comparable data from a number of different varieties of English allows the present study to provide deeper insights not only in the pragmatic functions of left dislocations and fronting constructions but also in the outcomes of language contact in this domain.

### References

- Filppula, Markku. 2009. The rise of *it*-clefting in English: areal-typological and contact-linguistic considerations. *English Language and Linguistics* 13: 267-293.
- Geluykens, Ronald. 1992. From discourse process to grammatical construction of left-dislocation in English. Amsterdam: John Benjamins.
- Gregory, Michelle L. & Laura A. Michaelis. 2001. Topicalization and left dislocation: A functional opposition revisited. *Journal of Pragmatics* 33: 1665-1706.
- Lange, Claudia. 2012. The syntax of spoken Indian English. Amsterdam: John Benjamins.

## Zero Preposition Time Adverbials

Shunji Yamazaki

Daito Bunka University, Japan

As noted by Quirk et al (1985: 692), some time adverbials may be realised by "a noun phrase instead of a prepositional phrase". Such variation can alternatively be described as variable overt vs. zero-marking of the adverbial by a preposition, e.g. *(On) Monday, I attended a long and boring meeting*. Quirk et al further list certain conditions under which the adverbial cannot have an overt preposition, including "before the deictic words last, next, this, that, and before the quantitative words some and every" (*loc. cit.*). Under other conditions, such as when the adverbial denotes delimited periods of time including years, months, weeks, or days of the week, both alternatives are possible. Some researchers have suggested that there may be context- or dialect-sensitive variation in the frequencies of use. For example, Algeo (1988: 14) notes that with such (named) periods of time, "the omitted preposition is Common English", but that there are several areas of difference between British and American English: in some cases, British English "has no preposition, but one would be expected in American" English, and by contrast "British [English] usually requires a preposition (*on*) with days of the week, whereas American [English] can have the preposition or omit it". However, to date, there has been little if any detailed quantitative study of this variable, presumably in large part because the unmarked forms need to be identified by visual inspection and cannot easily be retrieved automatically. The present research compares data from the Brown and LOB corpora to examine the following dimensions of variation in omission of adverbial prepositions:

1. Variation by preposition (some prepositions may be more likely to be omitted than others: for example, as a function of overall preposition frequency, or (conversely) specificity of meaning).
2. Variation with sentence position of the adverbial (preposition omission may be expected with higher frequency in sentence-initial adverbials than in sentence-final adverbials).
3. Variation by genre (in particular, more formal genres should more often favour overt markers).
4. Variation by semantic relationship between the sentence and the adverbial (preposition omission should be more frequent with more general time adverbial meanings, e.g. with expressions of time duration).
5. Variation by lexical item and/or frequency (some frequent expressions may favour preposition omission).
6. Variation by regional standard (British English versus American English), considered in combination with each of the preceding dimensions.

## References

- Algeo, John (1988). British and American grammatical differences. *International Journal of Lexicography* 1: 1-31.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

# Posters

**Compiling The Diachronic Corpus of Hong Kong English (DC-HKE): motivation, progress and challenges**

Carolyn Biewer<sup>i</sup>, Tobias Bernaisch<sup>ii</sup>, Benedikt Heller<sup>ii</sup>, Mike Berge<sup>ii</sup>  
 University of Zurich, Switzerland / University of Giessen, Germany<sup>ii</sup>

Hong Kong English as an emerging postcolonial variety shows remarkable structural differences from Standard British and American English (e.g. Bolton 2002a, Hung 2012), which find reflection on all levels of language organisation, i.e. in pronunciation (e.g. substitution of dental fricatives), in the lexicon (e.g. lexical borrowings from Cantonese such as taipan), as well as on the level of grammar (e.g. zero subjects, the use of will as past habitual marker), lexico-grammar (e.g. omission/insertion of particles) and pragmatics and discourse (e.g. the frequency and functions of actually).<sup>1</sup> These structural intricacies of Hong Kong English are at least partly the result of multiple language contact - with Cantonese (the native language of many Hong Kong Chinese) and Mandarin Chinese (which has gained influence through recent political changes) playing central roles. Other factors that shape the structure of English in Hong Kong are second-language acquisition effects and socio-cultural motivations of language use. Despite the distinctive present-day structural profile of Hong Kong English, its overall sociolinguistic developmental status remains largely unresolved to date. It is currently a hotly debated issue whether Hong Kong English has progressed into the phase of nativization as claimed by Schneider (2007: 137) or whether it has - largely due to the end of British colonial rule in Hong Kong in 1997 and its return to China - taken a step backwards and should rather be considered a foreign as opposed to a second-language variety (for opposite views see for instance Bolton (2002b) versus Görlach (2002)). Given the resulting necessity of an empirically-informed perspective onto the (more recent) structural developments of Hong Kong English, we set out to collect diachronic data for Hong Kong English. On this poster we want to give a work-in-progress report on the compilation of The Diachronic Corpus of Hong Kong English (DC-HKE). The aim is to create a 4-million-word corpus with data representing Hong Kong English in the 1930s, 1960s, 1990s and today which is comparable in its text categories to a) the Brown family and b) other diachronic second-language corpora of Indian and Singapore English currently compiled in Germany. On our poster we will briefly illustrate the colonial history of English in Hong Kong and the changes in the status and function of English since 1997. We will discuss open research questions in the study of Hong Kong English such as the ones mentioned above and we will point out interesting structural characteristics of Hong Kong English today. We will then demonstrate the usefulness of diachronic data to answer these research questions and to understand the origin and development of these features. The focus of the poster will be on methodological issues and challenges in the compilation of DC-HKE. We will also give a state-of-the-art report on the type and amount of data collected so far. Our aim is to introduce DC-HKE as a useful resource for research on Hong Kong English, which will also help us understand more general aspects in the evolution of non-native varieties of English. For more information on these listed features see Setter et al. (2010) and Cheng & Warren (2001).

**References**

- Bolton, Kingsley 2002a (ed.). *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press.
- Bolton, Kingsley 2002b. "The sociolinguistics of Hong Kong and the space for Hong Kong English". In: Kingsley Bolton (ed.). *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press, 29-55.
- Cheng, Winnie and Martin Warren 2001. "The functions of actually in a corpus of intercultural conversations". *International Journal of Corpus Linguistics* 6(2): 257-280.
- Görlach, Manfred 2002. "English in Singapore, Malaysia, Hong Kong, Indonesia, the Philippines ... a second or a foreign language?". In: Manfred Görlach (ed.). *Still More Englishes*. Amsterdam/Philadelphia: John Benjamins, 99-117.
- Hung, Tony T.H. 2012. "Hong Kong English". In: Ee-Ling Low & Azirah Hashim (eds.). *English in Southeast Asia: Features, Policy and Language in Use*. Amsterdam/Philadelphia: John Benjamins, 113-133.
- Schneider, Edgar W. 2007. *Postcolonial English. Varieties Around the World*. Cambridge: Cambridge University Press.
- Setter, Jane, Cathy S. P. Wong & Brian H. S. Chan 2010. *Hong Kong English*. Edinburgh: Edinburgh University Press.

### **'Love be the pilot, it's kamikaze': a comparison of metaphors of love across three music genres**

*Stephanie Furness-Barr*

This research is an investigation into metaphorical expressions of love in the lyrics of Country, R&B, and Rock and Metal music. Given that these genres are musically distinct, and tend to appeal to different audiences, it is reasonable to wonder if their lyrics characterise love, a common theme in popular music, any differently. Thus the aim of this research was to address the following questions:

1. What metaphors of love can be found in the three genres?
2. How are metaphors of love similar and different across the genres?

To find out, I composed three small corpora, consisting of the lyrics from 25 to 27 albums per genre, each of which reached the top, or near the top, of their respective charts in 2011, according to UK based 'The Official Charts Company'. The subsequent exploration focussed on i) discovering the number of instances of LOVE and common expressions associated with it, ii) identifying and categorising metaphorical expressions of love, then iii) comparing and contrasting these expressions across the three genres.

Initial analysis has shown similarity and variation across the three corpora, with the results as follows:

- i) The lyrics of all three genres refer to love, but with a higher number of instances per number of tokens in Country than R&B and a much lower number for Rock and Heavy Metal.
- ii) All three genres make use of LOVE IS A CONTAINER (BE in love) more or less equally. Rock & Metal lyrics include more instances of LOVE IS AN UNCONTROLLED DESCENT INTO A CONTAINER (FALL in love), and possessive LOVE (eg. 'my/your/our' + love) than Country or R&B.
- iii) A total of 229 expressions were identified as metaphorical, labelled and put into nine 'supra-metaphor' categories for comparison. The results show that Country lyrics for this year had the greatest number of metaphorical expressions, and Rock and Metal the least. The supra-metaphor category LOVE IS AN INANIMATE OBJECT had significantly more expressions overall, and was well represented in all three genres and where the majority of the Rock and Metal and R&B examples were found. LOVE IS (A/AN)(UNPREDICTABLE) FORCE was particularly favoured in Country lyrics, but also substantially represented in the other two genres. The third largest category was LOVE IS AN ANIMATE / ORGANIC OBJECT, again predominantly found in Country lyrics, whereas LOVE IS A CONDITION / AFFLICTION was favoured by R&B.

While all three genres have a number of specific metaphors in common, such as LOVE CAN BE GENUINE / FAKE and LOVE IS MENTAL INSTABILITY, Country and R&B lyrics shared significantly more metaphors with each other (e.g. LOVE IS LUCK / A GAMBLE, LOVE CAN BE BROKEN / UNDONE) than either share with Rock and Metal.

Further observations include the fact that:

- Metaphors used in Country music seem to give quite a positive or hopeful view of love.
- R&B lyrics have a fair number of expressions in which LOVE IS AN INTOXICANT.
- Rock & Metal, on the whole, does not have a dominant love metaphor.

## Investigating fluency in post-colonial varieties of English: A corpus-based feasibility study

Sandra Götz, Alicia Eckert, Amelie Waldhauer  
Justus Liebig University Giessen, Germany

Kachru's (1985) distinction between English as a native language (ENL), English as a second language (ESL) and English as a foreign language (EFL) has exerted an enormous influence on the modelling of Englishes worldwide. However, newer research shows that this distinction is far from clear-cut and the recent tendency is to view EFL, ESL and ENL rather as a continuum (e.g. Gilquin & Granger 2011; Housen *et al.* 2011; Mukherjee & Hundt 2011). Corpus-based studies in that vein have mainly focused on written English so far (very few exceptions being Ballier & Martin 2011 or Götz & Schilk 2011). A comparative approach to spoken fluency in EFL vs. ESL vs. ENL has not yet been undertaken; indeed, fluency has not played any role so far in the description of post-colonial varieties of English. Against the background of a 'cline model' of ENL, ESL and EFL forms of English, we would like to present a poster illustrating the findings of a feasibility study of the frequency and functions of selected fluency variables (*viz.* filled and unfilled pauses, mean length of runs, discourse markers, repeats and formulaic sequences) in spoken Sri Lankan and Nigerian English. The aim is (1) to investigate the status of fluency in post-colonial varieties of English for the first time and (2) to compare these findings with fluency variables found in EFL and ENL in order to place ESL on the continuum between EFL and ENL speakers' fluency. More specifically, we will analyze a subset of the spoken parts of the *International Corpus of English* (ICE-SL and ICE-NG) and compare our findings to previous research on fluency in ENL and EFL. To the best of our knowledge, this is the first comparative study on fluency features of Englishes from all three Kachruvian circles. Our feasibility study reveals some interesting observations:

(1) From a quantitative point of view, there is a distinct cline in the frequency of dysfluency features in spoken English. EFL has the highest frequency, followed by ESL, and the lowest frequency is found in ENL.

(2) The qualitative analysis of the functions of fluency enhancement strategies shows that while there is a large area of overlap in the use of some of these strategies in ENL, ESL and EFL forms under scrutiny (e.g. in the function of filled pauses as planning devices at the beginning of utterances or at clause boundaries), some variety/variant-specific functional preferences for how fluency is established also become clearly visible (e.g. a more frequent use of discourse markers in ENL, a preferred use of repeats in ESL, a higher number of filled pauses within constituents or phrases in EFL).

### References

- Ballier, N. & P. Martin (2011): "CV resyllabifications in French learner corpora and syllable-timed varieties of English". Paper presented at the Learner Corpus Research 2011. Université catholique de Louvain. Louvain-la-Neuve. 15-17 September 2011.
- Gilquin, G. & S. Granger (2011) "From EFL to ESL: evidence from the *International Corpus of Learner English*", in: J. Mukherjee & M. Hundt (eds.), *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a paradigm gap*. Amsterdam / Philadelphia: John Benjamins, 55–78.
- Götz, S. & M. Schilk (2011): "Formulaic sequences in spoken ENL, ESL and EFL: Focus on British English, Indian English and Learner English of advanced German learners", in: J. Mukherjee & M. Hundt (eds.), *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a paradigm gap*. Amsterdam: John Benjamins. 79–100.
- Housen, A., E. Schoonjans, S. Janssens, A. Welcomme, E. Schoonheere & M. Pierrard, (2011): "Conceptualizing and measuring the impact of contextual factors in instructed SLA: The role of language prominence", *International Review of Applied Linguistics in Language Teaching* 49(2): 83–112.
- Kachru, B.B. (1985): "Standards, codification and sociolinguistic realism: The English language in the outer circle", in: R. Quirk & H.G. Widdowson (eds.), *English in the World: Teaching and learning the language and literatures*. Cambridge: Cambridge University Press, 11–30.
- Mukherjee, J. & M. Hundt (eds.) (2011): *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a paradigm gap*. Amsterdam / Philadelphia: John Benjamins.

### English-Spanish phraseology in audit reports: the C-AURS Corpus

Marlèn Izquierdo, Isabel Pizarro, Rosa Rabadán

University of the Basque Country, Spain / University of Valladolid, Spain / University of León, Spain

Business Audit Reports (AuRs) are highly conventionalized texts whose communicative purpose is clear within the business speech community. These reports contain very specialized and precise information which has to be expressed unambiguously and concisely. For this, both English and Spanish display a relatively fixed set of syntactic and lexical resources which could be defined as phraseologisms (Gries 2008). This paper seeks to compare English and Spanish Audit Reports on the assumption that differences may appear with regard to i) the nature and number of the elements of phraseological units, ii) their frequency of occurrence and iii) their syntactic and lexical flexibility within, i.e., whether one given phraseologism could be considered as such even if the specified lexical item varies from occurrence to occurrence or the unit displays a differing syntactic arrangement (Gries 2008). The study also aims at iv) examining the distribution of phraseological units in the report, paying attention to their occurrence in the various rhetorical moves of the texts. To this end, a contrastive-rhetoric study has been conducted on an *ad-hoc* comparable corpus, namely, ACTRES C-AURs. The analysis has followed a top-down approach. In a first stage the structure of the text-type has been observed in both subcorpora English and Spanish. Second, a juxtaposition stage has brought to light a macrostructure of AuRs which is functionally common to both English and Spanish. A set of rhetorical labels for 'moves', 'steps' and 'sub-steps' has been defined, drawing on canonical studies (Bhatia, 1993; Connor and Upton, 2007) and previous ACTRES research (Xxx, 2012). ACTRES C-AURS has been tagged at the rhetorical level by means of a semi-automatic tagger developed by the ACTRES group. The rhetorical labelling functions as a *tertium comparationis*, making it possible to proceed from the bigger functional chunks down to lower levels of analysis of greater delicacy in both languages. A third stage is, thus, the actual contrast of phraseological units, at the lexis-syntax interface, as observed in ACTRES C-AURs. At this level, the extraction of units is semi-automatic; the browser retrieves linguistic information *per* rhetorical blocks from which recurrent and co-occurring word patterns will be analysed qualitatively and quantitatively. The study reveals relatively formulaic texts in both English and Spanish, bringing to light clear equivalent phraseological units English-Spanish. These results have been used to feed a computer writing aid addressed to those non-English or non-Spanish native speakers who can communicate fluently in their second language, but lack/need to learn the textual and discourse skills to produce effective Audit Reports.

### References

- Bhatia, V. K. 1993. *Analysing Genre: Language Use in Professional Settings*. Longman.
- Biber, D. U. Connor and T. A. Upton. 2007. (eds). *Discourse on the Move. Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins.
- González del Rey, I. 2008. *A multilingual focus on contrastive phraseology and techniques for translation*. Hamburgo: Kovac.
- Gotti, M. 2003. *Specialized Discourse. Linguistic Features and Changing Conventions*. Bern: Peter Lang.
- Hunston, S. 2011. *Corpus approaches to evaluation: phraseology and evaluative language*. New York: Routledge.
- Xxx. 2012. Designing writing materials for the business English language class. *Intercultural inspirations for language education. Spaces for understanding*, Ilona Semrádová et al. (eds), Univerzita Hradec Kralové (Rep. Checa). (ISBN 978 80 86771 52 6), pp. 46-55.
- Gries. S. Th., 2008. Phraseology and linguistic theory: a brief survey. In S. Granger & F. Meunier (eds.) *Phraseology. An Interdisciplinary perspective*. Amsterdam/Philadelhoa: John Benjamins, pp. 3-26.

### Ongoing grammatical changes and the new Englishes: Towards a set of corpora of English uses in the expanding circle

Mikko Laitinen, Magnus Levin, Alexander Lakaw  
Linnaeus University, Sweden

This poster presents ongoing corpus work that focuses on studying English uses in the expanding circle. The objective is to test the applicability of the methods used in the study of recent and ongoing grammatical changes for the study of new English uses. Much of the research on these changes has been carried out in the core varieties of English, and these investigations have built on intersecting historical linguistics and sociolinguistic variationism and relied on textually-balanced corpora offering equidistant observation points (Mair 2009). According to our current knowledge, the use of English in the expanding circle is often described as norm-dependent, but in today's globalized world, it is equally clear that the farther a language travels from its roots, the more likely it is to be affected by the multilingual settings in which it is used, as suggested by Mair (2013). Our research puts this working hypothesis to the test and aims to contribute to linguistic theories that reflect current knowledge of the nature of language variability and diachronic change. The objective of our corpus work is to create a systematic and replicable sampling frame for a set of corpora of present-day English uses in the expanding circle. We focus on English uses in countries in which the role of English is undergoing a shift from being a foreign language to that of a semi-official second language used as a linguistic resource in various contexts (cf. also the articles in Mukherjee & Hundt, eds. 2011 on seeing learner varieties and legislated second-language English as a continuum). In the current corpus linguistics thinking, English uses in the expanding circle are represented in the well-known learner corpora, such as the ICLE, but increasingly also in English as a lingua franca corpora (Seidlhofer 2011; Mauranen 2012), such as VOICE or ELFA containing spoken texts or in the new WrELFA corpus under compilation at the University of Helsinki. The core notion in our corpus work that the material should enable diachronic and diatopic study of how ongoing grammatical changes are adopted in use in the new Englishes. We feel that our corpus building could lead to a critical rethinking of the material sources in English corpus linguistics. For instance, the sampling frame for text categories used for the ICE family of corpora requires modification for the expanding circle where English tends to be used in various mediated genres, and the frame used for collecting the ICLE family is not suitable for diachronic studies, since the texts only cover limited genre variation. This poster discusses our theoretical assumptions, presents the sampling frame, which we have developed for exploring new English uses in the expanding circle, and explains the parameters for understanding textual variation. It also illustrates the ethnographically-informed present-day philological corpus work that we have engaged in, and also presents our future plans and prospects.

#### References

- Mair, Christian. 2009. Corpora and the study of recent change in language. In Anke Lüdeling & Merja Kytö (eds.) *Corpus Linguistics. An International Handbook*. Vol. 2, 1109-1125. Berlin: Walter de Gruyter.
- Mair, Christian. 2013. The World System of Englishes: Accounting for the Transnational Importance of Mobile and Mediated Vernaculars. *English World-Wide* 34, 253-278.
- Mauranen, Anna. 2012. *Exploring ELF: Academic English Shaped by Non-Native Speakers*. Cambridge: Cambridge University Press.
- Mukherjee, Joybrato & Marianne Hundt (eds.) 2011. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam: John Benjamins.
- Seidlhofer, Barbara. 2011. *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.



### Investigating Phrasal verbs in Late Modern Spoken English

Ljubica Leone

University of Salerno, Italy

The aim of this research is to investigate the syntactic changes that have occurred in Late Modern English (LModE) period (1700-1900) with a focus on Phrasal verbs in a short term span window, specifically in the century that goes from 1750 to 1850. Many studies have been conducted in a diachronic perspective with reference to phrasal verbs (Elenbaas 2007; Thim 2012), multiword verbs in general (Claridge 2000) but less has been done specifically on LModE. This is particularly true when the spoken dimension is involved because of the limitation of data. Few texts are available to analyze English from past times, and, as a consequence, the study of "indirect evidence using speech-based texts" (Biber 1998:252) of written records of real speech events such as trial proceedings and witness depositions, becomes of great interest. From this perspective, the Proceedings of the Old Bailey ([www.oldbaileyonline.org](http://www.oldbaileyonline.org)) are "a valuable textual source" (Huber 2007) and allow us to rebuild language changes providing some insights on how spoken language was like. A corpus, the LModE-OBC, has been specifically compiled with texts all belonging to the Proceedings of the Old Bailey. It consists of a collection of speech based texts, court records and transcriptions of legal proceedings covering the 100 year period from 1750 to 1850. It is divided into two subcorpora and 10 decades and totally brings up 1,008,000 words. This research is configured as a corpus-based investigation which involves both quantitative analysis and qualitative interpretation in the description of language usage. The aim is to explore Phrasal verbs in a historical perspective, and compare our results with the British National Corpus (BNC) for the Present Day English (PDE). Three major research questions will be addressed: (i) Are there any changes in Phrasal verbs along the time line? And what about their frequency? (ii) What are the differences between semantic regular combinations and idiomatic ones? (iii) What are the properties of the particles in phrasal verbs? And what about the particle placement? Are there specific factors that influence this feature?

It is likely that Phrasal verbs will show an increase in the frequency of use, and what we will also consider of great interest is to explain the reasons beyond this trend and analyze, if it will be the case, the stability of particular patterns.

### References

- Biber, D., Conrad, S., Reppen, R. (1998) *Corpus Linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. London: Longman.
- Claridge, C. (2000) *Multi-word Verbs in Early Modern English. A corpus-based study*. Amsterdam: Rodopi.
- Dehé, N. (2002) *Particle Verbs in English. Syntax, information structure and intonation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Elenbaas, M. (2007) *The Synchronic and Diachronic Syntax of the English Verb-Particle Combination*. Utrecht: LOT ed.
- Hickey, R. (2010). *Eighteenth-Century English. Ideology and Change*. Cambridge: Cambridge University Press.
- Hitchcock, T., Shoemaker, R. (2006) Digitising History From Below: The Old Bailey Proceedings Online, 1674–1834. *History Compass*, Vol. 4, Issue 2, pp. 193–202.
- Huber, M. (2007) The Old Bailey Proceedings, 1674-1834. Evaluating and annotating a corpus of 18<sup>th</sup> and 19<sup>th</sup> century spoken English. University of Giessen, *Annotating Variation and Change*.
- Kytö, M., Rydén, M., Smitterberg, E. (2006) *Nineteenth-century English. Stability and Change*. Cambridge: Cambridge University Press.
- McEnery, T., Hardie, A. (2012) *Corpus Linguistics. Method, Theory and Practice*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R., Tonio, Y. (2006) *Corpus-based Language Studies: An Advanced Research Book*. New York: Routledge.
- Nevalainen, T. (2006) *An Introduction to Early Modern English*. Edinburgh: Edinburgh University Press.
- Rydén, M. (1998) *The Study of Eighteenth Century English Syntax*. In: Rydén, M. & Tiekens-Bon Van Ostade, I., Kytö, M. (Eds.) *A Reader in Early Modern English*. Frankfurt: Peter Lang.
- Shoemaker, R. B. (2008) The Old Bailey Proceedings and the Representation of Crime and Criminal Justice in Eighteenth Century London, *Journal of British Studies*, 47 (3) , pp. 559-580.
- Thim, S. (2012) *Phrasal Verbs. The English Verb-Particle Construction and its History*. Berlin/Boston: De Gruyter Mouton.
- Tiekens I., Van Ostade B. (2009) *An Introduction to Late Modern English*. Edinburgh: Edinburgh University Press.
- Wright L. (2000). *The Development of Standard English. 1300–1800. Theories, Descriptions, Conflicts*. Cambridge: Cambridge University Press.

### Approximating the norm: Exploring EFL students' use of formulaic expressions in economics papers

Martti Mäkinen, Turo Hiltunen

Hanken School of Economics, Sweden / University of Helsinki, Finland

Formulaic language has been shown to be an important part of successful academic writing (e.g. Biber and Barbieri 2007), as well a common source of difficulties in L2 writers' texts across the board (e.g. Granger 1998, Coxhead 2008, Ädel and Erman 2012). However, the relationship between formulaic language and the quality of writing is far from simple. For instance, our earlier corpus-based study on formulaic language in L2 student writing (Hiltunen and Mäkinen forthcoming), based on a corpus of essays by students of economics, suggests that EFL writers do not simply underuse formulaic expressions in comparison to published and copy-edited research articles. In fact, formulaic expressions – here understood as items included in the Academic Formulas List (AFL) (Simpson-Vlach and Ellis 2010) – were shown to be almost equally frequent in student papers and published texts, though not necessarily in the same functions. Moreover, our study revealed considerable variation between individual writers, such that some writers use formulaic expressions extremely often, others hardly at all. For these reasons, it is difficult to give pedagogic recommendations about formulaic language that would apply to all situations. In this poster presentation, we complement our previous work, based on instructor-revised essays, in two ways. First, we investigate students' first drafts, produced before the first teacher intervention, in order to access their written production in a form that may give a more accurate view of their language skills. In addition, this approach enables us to assess the teacher-induced production of formulaic expressions in students' writing. Second, we explore instances of formulaic language that are based on our search list in the original study, but were not recoverable by it. For example, a search for the formula *on the other hand* would miss the usage quoted as example (1), which clearly represents an attempt to produce this formula:

- (1) *...in the other hand, not only the confirmatory cases can function as examples of the emergent patterns [...]* (Hanken Corpus: 0045)

Deviations of this kind occur occasionally in our data, as in other types of EFL data, both spoken and written (cf. Mauranen 2009). Our research questions are:

- (1) Are students' first drafts different from their final versions submitted at the end of the course, as far as formulaic expressions are concerned?
- (2) How frequent deviations of the type illustrated in (1) are, and in what kinds of texts do they occur?

In line with our previous work, the present study focuses on formulas included in the AFL, and their approximations. As our data, we use *Hanken Corpus of Academic Written English for Economics*, a POS tagged corpus of c. 280 texts/700,000 words consisting of first and final versions of end-of-term papers. Our main hypothesis, based on a preliminary analysis, is that students' first drafts are different from the finalised versions, especially in the frequency of discourse marker and topic introduction formulas. The aim of this poster is to investigate this hypothesis both in quantitative and qualitative terms.

#### References

- Ädel, A. and B. Erman, (2012). Recurrent Word Combinations in Academic Writing by Native and Non-native Speakers of English: A Lexical Bundles Approach. *English for Specific Purposes* 31 (2), 81–9.
- Biber, D., and F. Barbieri, (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26 (3), 263-286.
- Coxhead, A., (2008). Phraseology and English for Academic Purposes: Challenges and Opportunities. In Meunier, F. and S. Granger, (eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, 149-161.
- Granger, S., (1998). Prefabricated patterns in advanced EFL writing: collocations and lexical phrases. In Cowie, A.P. (ed.) *Phraseology: Theory, Analysis and Applications*, Oxford: Oxford University Press, 145-160.
- Hanken Corpus = *Hanken Corpus of Academic Written English for Economics*, (in preparation). Mäkinen, M. and T. Hiltunen (compilers). Hanken School of Economics and University of Helsinki.
- Hiltunen, T. and M. Mäkinen, (forthcoming). Formulaic language in L2 academic writing for business studies and economics. In Gotti, M. and D. Giannoni, *Corpus Analysis for Descriptive and Pedagogic Purposes: English Specialised Discourse*. Berlin: Peter Lang.
- Mauranen, A. (2009) Chunking in ELF: Expressions for managing interaction. *Intercultural Pragmatics* 6 (2), 217–233.
- Simpson-Vlach, R. and N. Ellis. (2010). An Academic Formulas List: New methods in phraseology research. *Applied Linguistics* 31 (4), 487-512.

## **Wow!-A Corpus-Based Study on Differences in the Use of Emotive Interjections in Child-and Adult-Directed Speech**

Ulrike Stange

Johannes Gutenberg Universität Mainz, Germany

The present paper, which subscribes to a register-based analysis of corpus data, addresses the question to what extent adults differ in their use of the interjections *Ow!*, *Ouch!*, *Ugh!*, *Yuck!*, *Whoops!*, *Whoopsadaisy!* And *Wow!*, depending on whether they are interacting with young children (1;8-3;0) or with other adults (14+). Corpus-based analyses of the spoken section of the BNCweb (subset conversations, adult-directed speech, 3.3 million words) and the Manchester corpus (CHILDES, child-directed speech, 1.3millionwords) have revealed a number of interesting findings. First, adults were much more likely to use interjections when engaging with children. In child-directed speech the frequencies ranged between 19.5 interjectional occurrences (IO) pmw (*Ouch!*) and 796 IO pmw (*Whoops!*), while the most frequent interjection in adult-directed speech had 53 IO pmw (*Ugh!*). Second, *Whoopsadaisy!* occurred only once in adult-directed speech (0.3 IO pmw), which provides empirical evidence for the claim that its use is inappropriate among adults, and that they virtually only use it in child-directed speech. Third, concerning the contexts in which the adults used emotive interjections, the abstract use of interjections of pain and disgust was restricted to adult-directed speech (this includes references to unpleasant jobs -*Yuck!*, running into ex-lovers -*Ouch!*, etc.), as this particular context of usage was entirely absent from the *CHILDES* data. Evidently, care givers refrain from using *Ow!* Or *Yuck!* in abstract contexts when engaging with young children based on the intuition that they cannot understand that yet. Furthermore, empathetic uses of *Ow!* and *Ouch!* were considerably more frequent in child-directed speech (for experienced, imaginary and anticipated pain). As regards the use of spill cries *Whoops!* and *Whoopsadaisy!*, they referred to the child most of the time in that the spill cries were mainly found in the contexts "failure of other's intention" (i.e. the child's) and "situation change noticed"(i.e. it had been caused by the child). The caregivers commonly used *Wow!* to compliment their interlocutor (i.e. the child) on doing something, a context of usage which was rare in adult-directed speech. To sum up, in adults' child-directed speech the use of emotive interjections was not only considerably more frequent, it also turned out to be predominantly other-oriented in that the use of emotive interjections referred to the child's pain, mishap, or actions, where as in adult-directed speech it was significantly more egocentric (that is, referring to speaker's own pain, mishap, or actions). What is more, my paper has shown that the use of interjections cannot adequately be described without accounting for situational parameters such as age of participants (adults, children), self-or other orientation, and audience design (child-directed or adult-directed).

### **References**

BNCweb<http://corpora.lancs.ac.uk/BNCweb/>  
Manchester<http://childes.psy.cmu.edu/data/Eng-UK-MOR/>

## Academic argument and other beasts: a CL and SFL investigation of disciplinary seminar discussion.

Elizabeth Tanguay

Aston University, United Kingdom

'The teacher sits at the head of the classroom, feeling pleased with herself and her class. The students are engaged in a heated debate. The very noise level reassures the teacher that the students are participating, taking responsibility for their own learning. Education is going on. The class is a success.' (Tannen, 1998:263).

Research into academic seminar discussions has investigated the function of lexical or grammatical items through the use of corpus techniques in order to draw comparisons with other academic registers (Biber, 2006). Alternatively, approaches to Discourse Analysis have been used which allow the investigation of sequential patterns of interaction over a number of turns (looking at for example initiating, responding and feedback moves in exchanges (Basturkmen, 2002), sometimes combining Conversation Analysis with CL in order to look at sequences in interaction (O'Keeffe, & Walsh, 2010). This poster considers how corpus techniques can be fruitfully combined with a different approach to DA – a genre framework from Systemic Functional Linguistics (following Eggins & Slade, 2004), in order to characterise the functional nature of longer stretches of discussion across the disciplines in UK Higher Education. The part of the study presented here aimed to investigate whether academic seminar discussion in UK HE is usefully characterised as argument, or whether there are other ways of conceptualising academic seminar discussion. Data used in the study is drawn from the seminar component of the British Academic Spoken English corpus (BASE). Six types of discussion, termed discussion macro genres, identified across three disciplinary areas in the corpus are outlined. Preliminary results from investigations of these discussion macro genres call into question notions of academic discussion as argument. There are indications that while a part of the seminar discussion in the corpus is indeed oppositional in nature (with participants arguing about whether x is y or z), a much greater proportion is dialogic, with seminar participants working together to reason towards a common goal. It is hoped that findings will be useful for those researching and teaching spoken academic discourse.

### References

- Basturkmen, H. (2002). Negotiating meaning in seminar-type discussion and EAP. *English for Specific Purposes*, 21, 233-242.
- Biber, D. (2006). *University language : a corpus-based study of spoken and written registers*. Amsterdam; Philadelphia: J. Benjamins.
- Eggins, S., & Slade, D. (2004). *Analysing casual conversation*. London: Equinox Publishing Ltd.
- O'Keeffe, A. a. W., S. (2010). Investigating higher education seminar talk. *Novitas-ROYAL(Research on Youth and Language)*. 4(2), 141-158.
- Tannen, D. (1998). *The argument culture : moving from debate to dialogue*. New York: Random House.

### Foreign Words and Phrases in English Historical Texts: Retrieval methods and quantitative findings

Jukka Tyrkkö, Arja Nurmi, Päivi Pahta Anna Petäjaniemi  
University of Tampere, Finland

Recent research has shown that multilingual practices can be found in a wide variety of written registers of English throughout the centuries (see, e.g., Kopaczyk 2013, Nurmi & Pahta 2013, Pahta 2012, and studies in Schendl & Wright 2011). However, at present most of what we know in the emerging field of historical multilingualism is largely based on the qualitative analysis of original texts or relatively small corpora, such as the *Helsinki Corpus*, *Corpus of Early English Correspondence* or *Anglo-Norman Correspondence Corpus*. To form a more coherent view of how frequent multilingual practices have been in English and how they vary between genres and registers across centuries, we need a large-scale survey using historical mega-corpora and appropriate retrieval methods. This paper discusses two distinct but related topics. First, we will survey the precision and recall of four different methods of identifying foreign words in a corpus otherwise written in standard English: function words (see Johnson 1993), foreign tags in a POS-annotated corpus (see Grefenstette 1995), dictionary-based retrieval (see Alex 2005) and non-standard letter clusters (see Andersen 2011). However, unlike with most previous work in this area, our objective is not to identify the language of individual texts or to identify individual words to improve natural language processing (see Alex et al. 2007), but rather to identify such instances for the purpose of further analysis as evidence of multilingual practices (see Andersen 2012). The fact that we focus on Late Modern and Early Modern English introduces several additional challenges such as non-standard spelling. We will also discuss the pros and cons of quantifying language alternation on the level of words vs. chunks and demonstrate the latest version of our *Multilingualiser* corpus tool. Second, using the methods outlined above, we present findings from the 34-million-word *Corpus of Late Modern English 3*, enhanced with new descriptive metadata. We will finish by discussing the significance of these early results to the more detailed corpus study of multilingual practices and outline the next steps in our long-term project, *Multilingual Practices in the History of Written English*, funded by the Academy of Finland.

#### References

- Alex, Beatrice (2005). An unsupervised system for identifying English inclusions in German text. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, Student Research Workshop. 133–138.
- Alex, Beatrice, Amit Dubey and Frank Keller (2007). Using foreign inclusion detection to improve parsing performance. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning 2007 (EMNLP-CoNLL 2008)*, 151–160. Available online at <<http://aclweb.org/anthology/D/D07/D07-1016.pdf>>
- Andersen, Gisle (2011). Corpora as lexicographical basis – The case of anglicisms in Norwegian. In Rayson Paul, Sebastian Hoffmann and Geoffrey Leech (eds.), *Methodological and Historical Dimensions of Corpus Linguistics. Studies in Variation, Contacts and Change in English* 6. Available online at <http://www.helsinki.fi/varieng/series/volumes/06/andersen/>.
- Andersen, Gisle (2012). Semi-automatic approaches to Anglicism detection in Norwegian corpus data. In Furiassi, Cristiano, Virginia Pulcini and Félix Rodríguez González (eds.), *The Anglicization of European Lexis*. Amsterdam: John Benjamins, 111–130.
- Corpus of Late Modern English 3 (CLMET3)*. Compiled by Hendrik De Smet, Hans-Jürgen Diller and Jukka Tyrkkö. See <<https://perswww.kuleuven.be/~u0044428/>>.
- Grefenstette Gregory (1995). Comparing two language identification schemes. In *Proceedings of Analisi Statistica dei Dati Testuali (JADT)*, 263–268.
- Kopaczyk, Joanna (2013). Code-switching in the records of a Scottish brotherhood in early modern Poland-Lithuania. In *Poznań Studies in Contemporary Linguistics* 49 (3), 281-319.
- Nurmi Arja & Päivi Pahta (2011). Multilingual practices in women's English correspondence 1400–1800. In Mark Sebba, Shahrzad Mahootian & Carla Jonsson (eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse*. New York / London: Routledge, 44-67.
- Pahta Päivi (2012). Code-switching in English of the Middle Ages. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*. New York: Oxford University Press, 528-537.
- Schendl, Herbert & Laura Wright (eds.) (2011). *Code-switching in Early English*. Berlin/Boston: De Gruyter Mouton.

### A diachronic perspective on the English modal system: *had to* and *must* in the 18<sup>th</sup> and 19<sup>th</sup> centuries

Bianca Widlitzki

Justus Liebig University Giessen

This paper investigates variation between *must* and *had to* as verbal expressions of obligation and logical inference in the past tense. In contrast to contemporary English, where *must* can only serve as a present-tense verb, earlier stages of English allowed *must* with past reference. The following examples from 1783 and 1880 illustrate this:

- (1) [M]y landlady came in and told me I **must** move [...]. (OBC, t17830430-30)
- (2) I could see that the packing-case **must** be resting on something in the trap [...]. (OBC, t18800524-529)

Most studies of variation between *must* and *have to* concentrate on ongoing changes in the modal system and on the present tense (cf. e.g. Smith 2003, Close and Aarts 2010). This investigation, in contrast, focuses on the 18<sup>th</sup> and 19<sup>th</sup> centuries, when *had to* gradually replaced *must* in the past tense, leading to the current division of labour between *must* in present and alternative forms in past contexts. Using a quantitative and qualitative corpus-linguistic approach, the study explores the internal and external factors that shaped this development: in particular, the type of modality and sociobiographical factors are discussed. As both are highly relevant for current developments concerning present-tense *have to* and *must* (cf. Close and Aarts 2010: 176f. on type of modality, Tagliamonte and Smith 2006: 371 on speakers' gender and age), they might also provide explanations for past developments. As a data source, the Old Bailey Corpus (OBC; <http://www.uni-giessen.de/oldbaileycorpus>), containing data from 1720 to 1913, is well-suited: Its sheer size (14 million words) permits fine-grained analyses and its extensive mark-up (e.g. on age, gender, class) facilitates sociolinguistic investigations. As the corpus is based on trial proceedings (*The Proceedings of the Old Bailey*; <http://www.oldbaileyonline.org>), it closely reflects the spoken language of the period. All relevant tokens of *must* (in the past tense) and *had to* from the corpus are coded for type of modality, distinguishing root meaning (=obligation, see (1)) and epistemic meaning (=logical inference, see (2)). A first analysis confirms that the 18<sup>th</sup> and 19<sup>th</sup> centuries do indeed constitute a decisive phase in the replacement of *must* by *had to*. In the first third of the period spanned by the corpus (1720-1789), *must* is still the preferred option and used in more than 97% of the contexts where *must* or *had to* are possible. In the last third (1850-1913), the rate of *must* falls to 23%. Preliminary results indicate that this change is led by the lower social classes. This study will also investigate the role of other social variables and type of modality in the replacement process, thus presenting a comprehensive picture of the change. It is hoped that this will also add to our understanding of contemporary variation between present-tense *must* and *have to*.

#### References

- Close, Joanne and Bas Aarts (2010). "Current change in the modal system of English: A case study of *must*, *have to*, and *have got to*." Eds. Ursula Lenker, Judith Huber and Robert Mailhammer, *English Historical Linguistics 2008.: Volume 1: The history of English verbal and nominal constructions*. Amsterdam: Benjamins. 165–181.
- Smith, Nicholas (2003). "Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English." Eds. Roberta Facchinetti, Manfred G. Krug and Frank Palmer, *Modality in contemporary English*. Berlin: Mouton de Gruyter. 241–267.
- Tagliamonte, Sali and Jennifer Smith (2006). "Layering, competition and a twist of fate: Deontic modality in dialects of English." *Diachronica* 23 (2): 341–380.

# Software demonstrations

## FRED online

*Katharina Ehret, Bernd Kortmann*  
*University of Freiburg, Germany*

This paper introduces and describes software which will allow researchers to access and query the Freiburg English Dialect Corpus (FRED) online. Hitherto, FRED is—due to copyright issues—only available for on-site usage at the University of Freiburg, and of-site research is limited to a small subset of the corpus. To remedy this situation, a web application for FRED was developed, which provides tools for linguistic corpus analyses, while protecting copyrighted material. FRED is a monolingual corpus of regional English dialects and samples 372 interviews with native speakers, amounting to approximately 2.5 million words of transcribed text and 300 hours of recorded speech. The corpus is organised into nine major dialect areas covering 43 counties and 163 locations in England, Wales, Scotland, the Hebrides and the Isle of Man. Thus, FRED provides extensive geographical coverage of traditional varieties in the British Isles spoken during the second half of the twentieth century (Hernandez 2006). In short, the Freiburg English Dialect Corpus is a unique and invaluable resource for typological and dialectological research. Based on the emulation software of the project "Baden-Württemberg Functional Long-Term Archiving" (<http://bw-fla.uni-freiburg.de/>), an online application was designed for searching the corpus while, at the same time, preventing editing or copying of the original data. On a more technical note, the corpus text files and an open source concordance programme will be installed on an emulated operating system which researchers can start in their web browser (<http://bw-fla.uni-freiburg.de/>). In other words, FRED can be locally accessed through any web browser and researchers can, for example, conduct simple concordance searches, create sorted frequency lists or establish the type-token ratio of selected corpus texts with the concordance programme provided in the emulation. Hence, the corpus can be edited and analysed without damage to the original data. In summary, we demonstrate how the Freiburg English Dialect Corpus can be readily accessed and analysed from the web, thus enabling researchers to conduct of-site research. In more general terms, we present suitable software to make corpus data whose distribution is limited by copyright publicly available for research.

## References

Nuria Hernandez. User's guide to FRED : Freiburg corpus of english dialects, 2006. URL  
<http://www.freidok.uni-freiburg.de/volltexte/2489/>.



## CAT: an advanced environment for the manual annotation of text and corpora

Giovanni Moretti, Matteo Fuoli, Rachele Sprugnoli

Fondazione Bruno Kessler, Italy / Lund University, Sweden / University of Trento, Italy

This software demonstration will provide an overview of the functionality of the Content Annotation Tool – CAT, a general-purpose web-based tool for manual text annotation that can be successfully employed in corpus-based analyses of semantic/pragmatic and discourse phenomena. CAT can also aid the creation of databases of annotated corpus examples for multivariate corpus-based analyses (e.g. Geeraerts et al 1994, Gries 1999, Divjak 2006, Glynn 2009). CAT provides a user-friendly interface for annotating text spans of variable length on the basis of an annotation scheme fully defined by the user. Text annotation is performed by highlighting a text span and manually assigning the desired category labels to it. Annotated data can be exported in stand-off XML format or, alternatively, in tabular 'case-by-variable' format, which can be used with spreadsheet and statistical software (e.g. *R*) for further processing and analysis. Finally, the program features a statistics module that calculates the frequency of annotated types and chance-corrected agreement between independent annotators (Dice coefficient). Among the major strengths of CAT are its ease of use and flexibility. CAT does not require any programming skills or prior knowledge of XML for its installation and use and allows users to freely define and dynamically change the annotation scheme as the project progresses. Compared to similar software, e.g. the UAM Corpus Tool (O'Donnell, 2008), CAT offers several advantages. Most notably, it is web-based, so different people in different locations can work on the same annotation project simultaneously. Further, CAT allows to annotate discontinuous text spans and to export the annotation results in a case-by-variable format, facilitating sophisticated statistical analyses. CAT has already been used in various Natural Language Processing projects. It has been successfully tested on TimeML annotation for the creation of part of the Ita-TimeBank, the largest Italian corpus annotated with information for temporal processing (Caselli et al., 2011). CAT has also been used to perform a semantic annotation of children's stories within the TERENCE European project and to manually annotate customer interactions within the EXCITEMENT European project. Recently, CAT has been chosen as the tool for the annotation of temporal information, semantic roles and intra-document co-reference within the NewsReader European project (Fokkens et al., 2013). While so far it has been mainly used to develop resources for training and evaluation of automatic NLP systems, CAT finds application in the field of corpus linguistics as well. As part of the software demonstration, we will show a concrete example of the use of CAT in a corpus-based multifactorial analysis of *evaluation* (Bednarek, 2006; Hunston, 2011; Martin and White, 2005) in a small-sized specialized corpus of business reports. The case study will be used to demonstrate the advantages of using manual annotation and CAT for the quantitative analysis of evaluation and to show that insightful multivariate analyses can be performed on the basis of richly annotated corpora.

## References

- Bartalesi Lenzi, V., Moretti, G., Sprugnoli, R., 2012. 'CAT: the CELCT Annotation Tool'. In *Proceedings of LREC 2012*, Istanbul.
- Bednarek, M. (2006). *Evaluation in media discourse: analysis of a newspaper corpus*. London & New York: Continuum International Publishing Group.
- Caselli, T., Bartalesi Lenzi, V., Sprugnoli, R., Pianta, E., & Prodanof, I. 2011. 'Annotating Events, Temporal Expressions and Relations in Italian: the It-Timeml Experience for the Ita-TimeBank'. In *Proceedings of LAW 5*, 143-151.
- Divjak, D. 2006. 'Ways of intending: A corpus-based cognitive linguistic approach to near-synonyms in Russian'. St. Th. & A. Stefanowitsch (eds). *Corpora in Cognitive Linguistics*. Berlin: Mouton de Gruyter, 19–56.
- Fokkens A., van Erp, M., Vossen, P., Tonelli, S., van Hage, W., Serafini, L., Sprugnoli, R., Hoeksema, J. 2013. 'GAF: A Grounded Annotation Framework for Events'. In *Proceedings of Workshop on Events: Definition, Detection, Coreference, and Representation*, NAACL HLT 2013, 11–20.
- Geeraerts D., Grondelaers, S. & Bakema, P. 1994. *The Structure of Lexical Variation. Meaning, naming, and context*. Berlin: Mouton de Gruyter.
- Glynn, D. 2009. 'Polysemy, syntax, and variation. A usage-based method for Cognitive Semantics'. V. Evans & S. Pourcel (eds). *New Directions in Cognitive Linguistics*. Amsterdam: John Benjamins, 77–104.
- Gries, Stefan Th. 1999. 'Particle movement: A cognitive and functional approach'. *Cognitive Linguistics* 10: 105-145.
- Hunston, S. 2011. *Corpus approaches to evaluation: phraseology and evaluative language*. New York & London: Routledge.
- Martin, J. & White, P. 2005. *The language of evaluation: Appraisal in English*. London & New York: Palgrave Macmillan.
- O'Donnell, M. 2008. 'Demonstration of the UAM corpus tool for text and image annotation'. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Demo Session*, pages 13–16. Association for Computational Linguistics.

## The Multidimensional Analysis Tagger - A tool for genre and text type analysis

Andrea Nini  
Aston University, United Kingdom

The present paper will introduce the Multidimensional Analysis Tagger (MAT), a program for the analysis of corpora or single texts following the model proposed by Biber (1988). The program uses the Stanford Tagger to generate an initial tagged version of the input which is then used to find and count Biber's (1988) linguistic features. The program then plots the text or corpus on to Biber's (1988) Dimensions and assigns it a text type as proposed by Biber (1989). Finally, MAT offers a tool to visualise the features of each Dimension in the text. The software was tested for reliability by comparing the Dimensions scores produced by MAT against the ones obtained by Biber (1988) in the analysis of the LOB corpus. This analysis show that MAT can largely replicate the results found in Biber's original study. The software was also tested on the Brown corpus. The analysis of the Brown corpus not only confirms the reliability of MAT in calculating the Dimension scores but also suggest that Biber's (1988) Dimensions and text types can be generalised and applied to the Brown corpus as well. Finally, as an example of a MAT analysis, the study of a corpus of threatening and abusive letters will be reported. This corpus did not contain enough texts and the balanced genre categories required to perform a factor analysis for the purpose of a multidimensional analysis of genre variation. However, by plotting the texts in pre-defined dimensions of variations such as the ones proposed by Biber (1988), MAT allowed a genre and text-type analysis of the corpus. In addition, the genre analysis performed by MAT allowed a comparison between threatening or abusive letters and other genres of the English language. The findings of this study indicate that threatening or abusive letters are similar to personal and professional letters, both in terms of Dimension scores and text types. Moreover, this study suggests that the generalizability of Biber's (1988) Dimensions can go as far as helping in the analysis of other genres. It is possible to conclude from the analyses explored in the present paper that MAT is a useful tool for performing a genre or text type analysis when it is impossible to carry out a factor analysis or cluster analysis, such as in cases of insufficient data or imbalanced categories. MAT is also useful when the aim is to compare a particular text or corpus against Biber's (1988; 1989) reference genres and text types.

## References

- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.  
Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1), 3–43.

# List of participants

## List of participants

Aarts, Bas	b.aarts@ucl.ac.uk	UCL
Ådel, Annelie	annelie.adel@du.se	Dalarna University
Aijmer, Karin	karin.aijmer@eng.gu.se	University of Gothenburg
Alsop, Sian	alsops@uni.coventry.ac.uk	Coventry University
Andersen, Gisle	gisle.andersen@nhh.no	NHH
Anglemark, Linnéa	Linnea.Anglemark@nordiska.uu.se	Uppsala University
Archer, Dawn	dearcher@uclan.ac.uk	University of Central Lancashire
Baron, Alistair	a.baron@lancaster.ac.uk	Lancaster University
Bech, Kristin	kristin.bech@ilos.uio.no	University of Oslo
Bernaisch, Tobias	Tobias.J.Bernaisch@anglistik.uni-giessen.de	University of Giessen
Berzlanovich, Ildiko	berzildi@yahoo.com	Eötvös Loránd University
Biewer, Carolin	carolin.biewer@es.uzh.ch	University of Zurich
Blass, Anne-Katrin	bllass@uni-trier.de	University of Trier
Bredenbröcker, Martina	martina.bredenbroecker@upb.de	University of Paderborn
Brems, Lot	lbrems@ulg.ac.be	University of Liège
Brezina, Vaclav	v.brezina@lancaster.ac.uk	Lancaster University
Brookes, Gavin	Gavin.Brookes@nottingham.ac.uk	University of Nottingham
Brunner, Marie-Louise	ml.brunner@t-online.de	Saarland University
Busse, Beatrix	beatrix.busse@as.uni-heidelberg.de	Heidelberg University
Carey, Ray	ray.carey@helsinki.fi	University of Helsinki
Carter, Ronald	ronald.carter@nottingham.ac.uk	University of Nottingham
Cermakova, Anna	anna.cermakova@ff.cuni.cz	Charles University, Prague
Close, Joanne	j.close@chester.ac.uk	University of Chester
Conklin, Kathy	kathy.conklin@nottingham.ac.uk	University of Nottingham
Davydova, Julia	jdavydov@mail.uni-mannheim.de	University of Mannheim
De Cock, Sylvie	sylvie.decock@uclouvain.be	Université catholique de Louvain
De Dios, Tanya	-	University of Santiago de Compostela
De Felice, Rachele	r.defelice@ucl.ac.uk	University College London
De Haan, Pieter	p.dehaan@let.ru.nl	Radboud University Nijmegen
Defour, Tine	Tine.Defour@UGent.be	Ghent University
Degaetano-Ortlieb, Stefania	s.degaetano@mx.uni-saarland.de	Saarland University
Demjén, Zsófia	zsofia.demjen@open.ac.uk	Open University
Demmen, Jane	j.demmen@lancaster.ac.uk	Lancaster University
Deuber, Dagmar	deuber@uni-muenster.de	University of Muenster
Diemer, Stefan	s.diemer@mx.uni-saarland.de	Saarland University
Dirdal, Hildegunn	hildegunn.dirdal@ilos.uio.no	University of Oslo
Donaldson, Christopher	c.donaldson@lancaster.ac.uk	Lancaster University
Dupont, Maïté	maite.dupont@uclouvain.be	Université catholique de Louvain
Eckert, Alicia	alicia.eckert@lehramt.uni-giessen.de	University of Giessen
Edler, Sabrina	sabrina.edler@uni-muenster.de	University of Muenster
Egan, Thomas	Thomas.Egan@hihm.no	Hedmark University College
Ehret, Katharina	katharina.ehret@hpsl.uni-freiburg.de	University of Freiburg
Elsness, Johan	johan.elsness@ilos.uio.no	University of Oslo
Evans, Dave	david.evans@nottingham.ac.uk	University of Nottingham
Fang, Alex Chengyu	alex.fang@cityu.edu.hk	City University Hong Kong
Filppula, Markku	markku.filppula@uef.fi	University of Eastern Finland

## List of participants

Fisher, Tony	anthony.fisher@nottingham.ac.uk	University of Nottingham
Flowerdew, Lynne	flowerdewlynn@gmail.com	City University of Hong Kong,
Flowerdew, John	lclynne@ust.hk	City University of Hong Kong,
Freddi, Maria	maria.freddi@unipv.it	University of Pavia
Frisk, Irina	irina.frisk@miun.se	Mid-Sweden University
Fruttaldo, Antonio	antonio.fruttaldo@unina.it	University of Naples 'Federico II'
Frye, Jessica	j.frye@mmu.ac.uk; ljfrye@hotmail.co.uk	Manchester Metropolitan University
Fuchs, Robert	robert.fuchs@uni-muenster.de	University of Muenster
Fuoli, Matteo	Matteo.Fuoli@englund.lu.se	Lund University
Garretson, Gregory	gregory.garretson@engelska.uu.se	Uppsala University
Gee, Matt	matt.gee@bcu.ac.uk	Birmingham City University
Geng, Yifan	yifan.geng@warwick.ac.uk	University of Warwick
Gerckens, Caroline	caroline.gerckens@engsem.uni-hannover.de	University of Hannover
Gilquin, Gaëtanelle	gaetanelle.gilquin@uclouvain.be	University of Louvain
Götz, Sandra	Sandra.Goetz@anglistik.uni-giessen.de	Justus Liebig University Giessen
Grafmiller, Jason	Jason.Grafmiller@kuleuven.be	KU Leuven
Gregory, Ian	i.gregory@lancaster.ac.uk	Lancaster University
Gries, Stefan Th.	stgries@gmail.com	University of California, Santa Barbara
Grieve, Jack	j.grieve1@aston.ac.uk	Aston University
Gupta, Kat	kat.gupta@nottingham.ac.uk	University of Nottingham
Gut, Ulrike	gut@wwu.de	University of Münster
Hackert, Stephanie	Stephanie.Hackert@anglistik.uni-muenchen.de	LMU Munich
Hang, Su	hxs201@bham.ac.uk	University of Birmingham
Hardie, Andrew	a.hardie@lancaster.ac.uk	Lancaster University
Harvey, Kevin	kevin.harvey@nottingham.ac.uk	University of Nottingham
Hasselgård, Hilde	hilde.hasselgard@ilos.uio.no	University of Oslo
Heller, Benedikt	mail@benedikt-heller.de	KU Leuven
Heyvaert, Liesbet	liesbet.heyvaert@arts.kuleuven.be	University of Leuven
Hiltunen, Turo	turo.hiltunen@helsinki.fi	University of Helsinki
Hoffmann, Sebastian	hoffmann@uni-trier.de	University of Trier
Hofland, Knut	knut.hofland@uni.no	Uni Research Computing
Hu, Xianyao	x.hu4@lancaster.ac.uk	Lancaster University
Huber, Magnus	magnus.huber@anglistik.uni-giessen.de	University of Giessen
Hunston, Susan	s.e.hunston@bham.ac.uk	University Of Birmingham
Hunt, Sally	s.hunt@ru.ac.za	Rhodes University
Ifukor, Presley	pifuk_01@uni-muenster.de	University of Muenster
Izquierdo, Marlén	marlen.izquierdo@ehu.es	University of the Basque Country
Jaworska, Sylvia	s.jaworska@reading.ac.uk	University of Reading
Johannsen, Berit	berit.johannsen@fu-berlin.de	FU Berlin
Kaunisto, Mark	mark.kaunisto@uta.fi	University of Tampere
Kehoe, Andrew	andrew.kehoe@bcu.ac.uk	Birmingham City University
Kimps, Ditte	ditte.kimps@arts.kuleuven.be	University of Leuven
Kirk, John	jk@etinu.com	Independent Researcher, Belfast
Koller, Veronika	v.koller@lancaster.ac.uk	Lancaster University
Kortmann, Bernd	bernd.kortmann@anglistik.uni-freiburg.de	University of Freiburg

## List of participants

Kranich, Svenja	kranich@uni-mainz.de	University of Mainz
Kreyer, Rolf	kreyer@uni-marburg.de	University of Marburg
Kunz, Kerstin	kerstin.kunz@iued.uni-heidelberg.de	University of Heidelberg
Kytö, Merja	merja.kyto@engelska.uu.se	Uppsala University
Lahey, Ernestine	e.lahey@ucr.nl	University College Roosevelt
Lakaw, Alexander	alexander.lakaw@lnu.se	Linnaeus University
Landert, Daniela	daniela.landert@es.uzh.ch	University of Zurich
Lapshinova-Koltunski, Ekaterina	e.lapshinova@mx.uni-saarland.de	Saarland University
Larsson, Tove	tove.larsson@engelska.uu.se	Uppsala University
Lavid, Julia	lavid@ilol.ucm.es	Universidad Complutense de Madrid
Leech, Geoffrey	g.leech@lancaster.ac.uk	Lancaster University
Lehto, Anu	anu.lehto@helsinki.fi	University of Helsinki
Leone, Ljubica	lleone@unisa.it	University of Salerno
Levin, Magnus	magnus.levin@lnu.se	Linnaeus University
Littre, Damien	damien.littre@uclouvain.be	Université Catholique de Louvain
Liu, Ying	liuying1213@gmail.com	City University Hong Kong
López-Couso, María José	mjlopez.couso@usc.es	University of Santiago de Compostela
Louw, William (Bill)	godelkreiss@gmail.com	University of Zimbabwe
Maekelberghe, Charlotte	charlotte.maekelberghe@kuleuven-kulak.be	University of Leuven
Mahlberg, Michaela	michaela.mahlberg@nottingham.ac.uk	University of Nottingham
Mair, Christian	christian.mair@anglistik.uni-freiburg.de	University of Freiburg
Mäkinen, Martti	martti.makinen@hanken.fi	Hanken School of Economics
Malá, Markéta	marketa.mala@ff.cuni.cz	Charles University in Prague
Mastropiero, Lorenzo	lorenzo.mastropiero@nottingham.ac.uk	University of Nottingham
McEnery, Tony	a.mcenery@lancaster.ac.uk	University of Lancaster
Meierkord, Christiane	christiane.meierkord@rub.de	Ruhr-University Bochum
Méndez-Naya, Belén	belen.mendez@usc.es	University of Santiago de Compostela
Meyer, Thomas	thomas.meyer@epfl.ch	École Polytechnique Fédérale de Lausanne
Milojkovic, Marija	marija.okon@gmail.com	University of Belgrade
Mindt, Ilka	mindt@mail.uni-paderborn.de	University of Paderborn
Molino, Alessandra	alessandra.molino@unito.it	University of Turin
Montoro, Rocio	rmontoro@ugr.es	University of Granada
Moratón, Lara	-	Universidad Complutense de Madrid
Murrieta-flores, Patricia	p.murrietaflores@chester.ac.uk	University of Chester
Nacey, Susan	susan.nacey@hihm.no	Hedmark University College
Nedoluzhko, Anna	nedoluzko@ufal.mff.cuni.cz	Charles University, Prague
Nesi, Hilary	h.nesi@coventry.ac.uk	Coventry University
Nilsson-Schwarz, Sarah	sarah.nilsson@engelska.uu.se	Uppsala University
Nini, Andrea	ninia@aston.ac.uk	Aston University
Nisco, Maria Cristina	mariacristina.nisco@unina.it	Università di Napoli Federico II
Nissel, Magnus	magnus.nissel@anglistik.uni-giessen.de	University of Giessen
Nordrum, Lene	lene.nordrum@englund.lu.se	Lund University
Novák, Michal	mnovak@ufal.mff.cuni.cz	Charles University, Prague
Nurmi, Arja	arja.nurmi@uta.fi	University of Tampere
Otchere, Gloria	gloria.otchere@ilos.uio.no	University of Oslo

## List of participants

Pace-Sigge, Michael	michael.pace-sigge@uef.fi	University of Eastern Finland
Pahta, Päivi	paivi.pahta@uta.fi	University of Tampere
Palacios Martínez, Ignacio	ignacio.palacios@usc.es	University of Santiago de Compostela
Peck, Rebecca	rebecca.peck@nottingham.ac.uk	University of Nottingham
Pellicer-Sánchez, Ana	Ana.Pellicer-Sanchez@nottingham.ac.uk	University of Nottingham
Petäjaniemi, Anna	anna.petajaniemi@uta.fi	University of Tampere
Philip, Gill	g.philip.polidoro@gmail.com	University of Macerata
Pizarro, Isabel	pizarro@lia.uva.es	University of Valladolid
Plevoets, Koen	koen.plevoets@ugent.be	University of Ghent
Priestley, Joel	joel.priestley@iln.uio.no	University of Oslo
Qiu, Christina Yinqing	christinasisu@hotmail.com	Chinese University of Hong Kong
Rabadán, Rosa	rosa.rabadan@unileon.es	University of León
Rayson, Paul	p.rayson@lancaster.ac.uk	Lancaster University
Reichardt, Renate	renate_reichardt@hotmail.com	University of Birmingham
Renouf, Antoinette	antoinette.renouf@btinternet.com	Birmingham City University
Richard, Sophie	sophie.richard.linguist@gmail.com	University of Western Australia
Römer, Ute	uromer@gsu.edu	Georgia State University
Rodríguez-Puente, Paula	paula.rodriguez@unican.es	University of Cantabria
Ronan, Patricia	marionpatricia.ronan@unil.ch	Université de Lausanne
Rørvik, Sylvi	sylvi.rorvik@hihm.no	Hedmark University College
Röthlisberger, Melanie	melanie.roethlisberger@access.uzh.ch	KU Leuven
Rothstein, Björn	bjoern.rothstein@rub.de	Ruhr Universität Bochum
Rudanko, Juhani	juhani.rudanko@uta.fi	University of Tampere
Rühlemann, Christoph	chrisruehlemann@googlemail.com	Philps University Marburg
Ruskan, Anna	ania_r_lt@yahoo.co.uk	Vilnius University
Rusten, Kristian	kristian.rusten@if.uib.no	University of Bergen
Rysová, Magdaléna	magdalena.rysova@ufal.mff.cuni.cz	Charles University, Prague
Salway, Andrew	andrew.salway@uni.no	Uni Research
Sand, Andrea	sand@uni-trier.de	Universität Trier
Schaub, Steffen	schaubst@staff.uni-marburg.de	University of Marburg
Schilk, Marco	schilk@uni-hildesheim.de	Stiftung Universität Hildesheim
Schmidt, Selina	hslinguistics@gmail.com	Saarland University
Schneider, Gerold	gschneid@es.uzh.ch	University of Zurich
Schreier, Daniel	schreier@es.uzh.ch	University of Zurich
Schützler, Ole	ole.schuetzler@uni-bamberg.de	University of Bamberg
Semino, Elena	e.semino@lancaster.ac.uk	Lancaster University
Seoane, Elena	elena.seoane@uvigo.es	University of Vigo
Shank, Christopher	c.shank@bangor.ac.uk	Bangor University
Sikveland, Rein	Rein.Sikveland@nottingham.ac.uk	University of Nottingham
Smith, Nicholas	ns359@le.ac.uk	Leicester University
Smitterberg, Erik	erik.smitterberg@engelska.uu.se	Uppsala University
Song, Myoung Hyoun	y720four@yahoo.com	Seoul National University
Stange, Ulrike	stangeu@uni-mainz.de	University of Mainz
Stenbrenden, Gjertrud Flermoen	g.f.stenbrenden@ilos.uio.no	University of Oslo
Suárez-Gómez, Cristina	cristina.suarez@uib.es	University of the Balearic Islands
Szmrecsanyi, Benedikt	benzsm@kuleuven.be	KU Leuven

## List of participants

Taavitsainen, Irma	irma.taavitsainen@helsinki.fi	University of Helsinki
Tagg, Caroline	c.tagg@bham.ac.uk	University of Birmingham
Tanguay, Elizabeth	tanguaye@aston.ac.uk	Aston University
Teubert, Wolfgang	teubertw@bham.ac.uk	University Of Birmingham
Thompson, Paul	p.thompson@bham.ac.uk	University of Birmingham
Tizón-Couto, David	davidtizon@uvigo.es	University of Vigo
Tranchese, Alessia	alessiatranchese@gmail.com	-
Trotta, Joseph	joe.trotta@sprak.gu.se	University of Gothenburg
Tsuchiya, Keiko	tsuchiya.keiko@gmail.com	Tokai University, Japan
Tyrkkö, Jukka	Jukka.Tyrkko@uta.fi	University of Tampere
Van Hattum, Marije	vanhatm@hope.ac.uk	Liverpool Hope University
Van Rooy, Bertus	bertus.vanrooy@nwu.ac.za	North-West University
Venuti, Marco	venuti@unina.it	Università di Napoli Federico II
Viberg, Åke	Ake.Viberg@lingfil.uu.se	Uppsala University
W. Schneider, Edgar	edgar.schneider@ur.de	University of Regensburg
Waldhauer, Amelie	Amelie.waldhauer@lehramt.uni-giessen.de	University of Giessen
Walker, Brian	b.d.walker@hud.ac.uk	University of Huddersfield
Wallis, Sean	s.wallis@ucl.ac.uk	University College London
Wang, Amy	y.wang@mmu.ac.uk	Manchester Metropolitan University
Wei, Naixing	nxwei@buaa.edu.cn	Beihang University
Werner, Valentin	valentin.werner@uni-bamberg.de	University of Bamberg
Whiteside, Karin	karin.whiteside@rhul.ac.uk	University of Warwick
Widlitzki, Bianca	bianca.widlitzki@anglistik.uni-giessen.de	University of Giessen
Wiegand, Viola	aexvw1@nottingham.ac.uk	University of Nottingham
Wiemeyer, Leonie	leonie_w@gmx.net	University of Bremen
Wikstrom, Peter	peter.wikstrom@gmx.com	Karlstad University
Winkle, Claudia	claudia.winkle@anglistik.uni-freiburg.de	University of Freiburg
Xiao, Richard	r.xiao@lancaster.ac.uk	Lancaster University
Yamazaki, Shunji	yamazaki@ic.daito.ac.jp	Daito Bunka University
Zehentner, Eva	eva.zehentner@univie.ac.at	University of Vienna
Zhang, Cathy Guiping	zgp3021@126.com	Chinese University of Hong Kong
Zinsmeister, Heike	heike.zinsmeister@uni-hamburg.de	University of Hamburg





The University of  
**Nottingham**

UNITED KINGDOM • CHINA • MALAYSIA