Adami, Elisabetta: To each reader his, their or her own pronoun: Prescribed, Proscribed and Disregarded Uses of Generic Pronouns in English

For a number of years grammarians, linguists and teachers have debated on which English pronoun should be used to refer individually to gender-indefinite or sex-mixed human categories and roles. When sexism in language was a major issue, both the long-lasting prescription of "generic *he*" and the proscription of the so-called "singular *they*" were questioned and various gender-fair alternatives, such as "*he/she*", were suggested.

Nowadays "The Great *He/She* Battle" seems to have exhausted its ink-munitions and, in absence of an agreed solution, "recast the sentence into the plural" remains the handbooks most suggested strategy in order to avoid the problem.

Yet so far no corpus-based study has apparently ascertained the current use of generic pronouns. In order to fill this gap, after a brief overview of the existing literature, this paper will investigate the use of generic pronouns in the academic and journalistic sections of the *BNC* and the *ANC*.

The analysis aims to evaluate both the current status of the proscription of "singular *they*" and the extent of the introduction of gender-fair neologisms such as "*he or she*". The two phenomena will be investigated in relation to possible variations in use between British and American English (the latter generally considered more conservative with generic pronouns).

Furthermore the paper will advance hypotheses on the rhetorical and pragmatic effects deriving from the rare, yet attested, "generic *she*", the use of which seems to give a peculiar perspective to the utterance with respect to logical categories such as general vs. particular, abstract vs. concrete, generalization vs. exemplification.

Finally, the results of the analysis will, we hope, lead to a fruitful comparison between language in use and grammatical descriptions and prescriptions of English generic pronouns.

References:
Baron, Dennis E., 1986, *Grammar and Gender*. New Haven: Yale University Press.
Bodine, Anne, 1975, "Androcentrism in Prescriptive Grammar: Singular 'they', Sex-indefinite 'he', and 'he or she'", *Language in Society*, 4, 129-146.
Cole, Peter, 1974, "Indefiniteness and Anaphoricity", *Language*, 50/4, 665-674.
Cooper, Robert L., 1984, "The avoidance of androcentric generics." *International Journal of the Sociology of Language,* 50, 5-20.
Corbett, Greville G., 1991, *Gender*. Cambridge: Cambridge University Press.
Jochnowitz, George, 1982, "Everybody Likes Pizza, Doesn't He or She?", *American Speech*, 57/3, 198-203.
Lakoff, Robin, 1975, *Language and woman's place*. New York: Harper & Row.
Martyna, Wendy, 1989, "The psychology of the generic masculine", in S. McConnell-Ginet, R. Borker, & N. Furman (eds.), *Women and language in literature and society*. New York: Praeger Publishers. 69-78.
Meyers, Miriam Watkins, 1993, "Forms of *they* with singular noun phrase antecedents: Evidence from current educated English usage", *Word*, 44/2, 181-192.
Mühlhäusler, Peter & Rom Harré, 1991, *Pronouns and People: The Linguistic Construction of Social and Personal Identity*. Oxford: Basil Blackwell.
Newman, Michael, 1996, *Epicene Pronouns: The Linguistics of a Prescriptive Problem*. London: Garland Publishing.
Nilsen, Alleen Pace, 1984, "Winning the Great He/She Battle", *College English*, 46/2, 151-157.
Sklar, Elizabeth S., 1988, "The Tribunal of Use: Agreement in Indefinite Constructions", *College Composition and Communication*, 39/4, 410-422.
Zuber, Sharon, & Ann M. Reed, 1993, "The Politics of Grammar Handbooks: Generic *He* and Singular *They*", *College English*, 55/5, 515-530.

Aijmer, Karin: Are there modal particles in English?

Modal particles are said to be a feature of some, mainly Germanic, languages (e.g. German, Swedish, Dutch, Danish, Norwegian).The definition of the term (modal) particle relies on a number of criteria such as position in the clause and the lack of stress. The modal particles do for instance not occupy initial position in the clause but 'particle' in the relevant languages has a fixed position in the verbal complex (the middle field), a topological notion referring to the position after the initial element of a complex verbal element. In addition modal particles are characterised as frequent and polyfunctional. The formal criteria are fairly rigid and are influenced by the German tradition of Partikelforschung (Diewald 2006).

A problem which has long tortured linguists has to do with whether modal particles can be defined functionally. A related question is whether modal particles have correspondences in all languages. My ideas in this paper are inspired by Hoye (1997) and his suggestion that modal adverbs in English should be treated as modal particles, 'primarily those adverbs with only faint shades of meaning'.

Following Hoye "it would not be implausible to redefine subjuncts expressing modality as 'modal particles', subdivided into the following categories: evidential particles (*clearly*, *obviously*); hearsay particles (*apparently*); reinforcement or emphasising particles (*certainly, surely, well*); and focus particles (*only, simply*)".

The aim of my presentation is to consider whether there is room for a category of modal particles in English and what their properties are. Features which will be considered are position and prosody (loss of accent). Special focus will be on function and on explaining the functions of modal particles in the perspective of grammaticalization theory (Waltereit 2001, 2006). A comparison will be made with equivalents in Swedish on the basis of the English-Swedish Parallel Corpus (Altenberg & Aijmer 2001).

References

Altenberg, Bengt and Karin Aijmer (2001). The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies. In: Mair, Christian and Marianne Hundt (eds) *Corpus linguistics and linguistic theory. Papers from the 20th International Conference on English Language Research on Computerized Corpora (ICAME 20) Freiburg im Breisgau 1999*. 15-33. Amsterdam & Philadelphia: Rodopi.

Diewald, Gabriele. 2006. Discourse particles and modal particles as grammatical elements. In: Fischer, Kerstin (ed.) *Approaches to discourse particles*. 403- 425. Amsterdam, etc: Elsevier.

Hoye, Leo. 1997. *Adverbs and modality in English.* London & New York: Longman.

Waltereit, Richard. 2001. Modal particles and their functional equivalents: A speech-act-theoretical approach. *Journal of Pragmatics* 33: 1391-1417.

Waltereit, Richard. 2006. *Abtönung. Zur Pragmatik und historischen Semantik von Modalpartikeln und ihren funktionalen Äquivalenten in romanischen Sprachen.* Tübingen: Niemeyer, 2006

Breban, Tine: Subjectification and leftward movement in the English noun phrase

Adamson (2000) proposed that in the English noun phrase (NP) processes of grammaticalization and subjectification are structurally accompanied by leftward movement. That is to say, when an element develops a new, more subjective meaning, it relocates to a more leftward position. This proposal fits in with the more generally claimed idea that subjectivity and objectivity are associated with the left and the right pole of the NP respectively (amongst others Quirk et al. 1972; Hetzron 1978; Dixon 1982).

In my paper, I will present the findings of an in-depth corpus study of the adjectives *very* and *true*, investigating whether subjectification as a diachronic hypothesis goes together with relocation to the left in the English NP and whether more advanced subjectification also entails further leftward movement. From a semantic point of view *very* and *true*, which share the same objective source semantics, have developed a range of grammaticalized/subjective uses in the NP (cf. Adamson and Diaz 2005 on *very*), including amongst others the degree modifier use, e.g. *a very long train*, an emphasizing postdeterminer use, as in *the very person*, and an evaluating modifier use, e.g. *a true gentleman*. I will start my paper with a comparative analysis of the developments of *very* and *true* as they can be traced in diachronic and synchronic corpus data. Then I will discuss how they relate to the leftward movement hypothesis on the basis of their occurrences as part of a group of premodifying adjectives in a wide range of historical (and synchronic) corpora.

References

Adamson, S. 2000. "A lovely little example: word order options and category shift in the premodifying string". In O. Fischer, A. Rosenbach and D. Stein, eds. *Pathways of Change: Grammaticalization in English*. Amsterdam/Philadelphia: John Benjamins. 39-66.

Adamson, S. and V. González-Diáz. 2005. "'And I were your very very Rosalind': the role of hyperbole in the grammaticalisation of intensifiers". Paper presented at Fitigra (From Ideational to Interpersonal: Perspectives from Grammaticalization), Leuven, 10-12 February 2005.

Dixon, R.M.W. 1982. *Where Have all the Adjectives Gone? And Other Essays in Semantics and Syntax*. Berlin: Mouton.

Hetzron, R. 1978. "On the relative order of adjectives". In H. Seiler, ed. *Language Universals: Papers from the Conference held at Gummersbach/Cologne, Germany, October 3-8, 1976*. Tübingen: Gunter Narr. 165-184.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1972. *A Grammar of Contemporary English*. London: Longman.

Calle-Martín, Javier & Antonio Miranda-García: On the use of split infinitives in English

A cleft or a split infinitive construction denotes a particular type of syntactic tmesis in which a word or phrase, especially and adverb or an adverbial phrase, occurs between *to* and the infinitival form of a verb. Although occasional from a statistical viewpoint, the earliest instances of the split infinitive date back to the 13th century, in which a personal pronoun, an adverb or two or more words could appear in such environments (Visser 1984, II: 1038-1045). Notwithstanding that the usage of the split infinitive is rare between the 16th to the 18th centuries, it begins to gain ground again throughout the 18th and 19th centuries, hence resisting the severe criticisms of grammarians. Nowadays, however, a search for these types of constructions in a present-day English corpus reveals that the prejudice against splitting infinitives is receding in view of the number of occurrences found. In light of this, in this paper these constructions will be investigated in order to a) provide their statistics and b) examine the types of adverbs which usually appear in these syntactic contexts. The corpus used as source of evidence has been extracted from the *British National Corpus* (2000).

References
Baker, Sheridan. 1980. "Comments on Comments on Splitting Infinitives". *College English* 42.1: 90-91.
Close, R.A. 1987. "Note on the Split Infinitive". *Journal of English Linguistics* 20.2: 217-229.
Crystal, David. 1985. "A Case of the Split Infinitives". *English Today* 3: 16-17.
Curme, George O. 1927. "The Split Infinitive". *American Speech* 2.8: 341-342.
Foster, David William. 1978. "A Transformational Justification for English Split Infinitives". In Dietrich Nehls (ed.). *Studies in Descriptive English Grammar*. Heidelberg: Groos, pp. 85-93.
Nagle, Stephen. 1994. "Infl in Early Modern English and the status of *to*". In Dieter Kastovsky (ed.). *Studies in Early Modern English*. Berlin and New York: Mouton de Gruyter, pp. 233–242.
Neely, Peter M. 1978. "To Split or Not to Split". *College English* 40.4: 402-406.
Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
Visser, Fredericus Theodorus. 1984. *An Historical Syntax of the English Language*. 4 vols. Leiden, Brill.

Claridge, Claudia: Variation in the use of pragmatic markers

The paper will investigate a group of pragmatic markers which partly overlap in their meanings and functions, namely *as it were*, *so to speak*/*say* and *if you like*. All of them can be used as hedges, providing metalinguistic comments on the applicability of the concepts or the type of expression used by the speaker. In addition, they fulfil to varying extent also interpersonal functions, for example in the context of politeness strategies, and textual functions with regard to, for instance, indicating information structure and focus. None of these forms is very common (in BrE), but with roughly forty instances per one million words *as it were*/*if you like* are much more frequent than *so to speak/say* with only about four instances per million. Furthermore, the distribution of these forms shows some interesting variation. While *as it were* and *if you like* are clearly spoken forms, *so to speak*/*say* is found to an almost equal extent in spoken and written contexts. Within the spoken domain, all these forms show a clear preference for content-oriented, informational and more formal contexts (as found in the context-governed part of the *BNC*/spoken part), in contrast to some other markers' preference for colloquial contexts. The users of these forms are overwhelmingly male speakers, which raises questions as to the role of gender for the use of hedges. Finally, there are also noticeable inter-varietal differences: while *if you like* seems to be an exclusively British variant and *as it were* is used more often in British English than in American English, *so to speak*/*say* is clearly preferred by American speakers. The paper will discuss some potential explanations for the above-mentioned distributions.

References
Brinton, Laurel. 1996. *Pragmatic markers in English: Grammaticalization and discourse functions.* Berlin/New York: Mouton de Gruyter.
Fraser, Bruce. 1990. "An approach to discourse markers." *Journal of Pragmatics* 14, 383-395.
Holmes, Janet. 1990. "Hedges and boosters in women's and men's speech." *Language and Communication* 10, 185-205.
James, Allan R. 1983. "Compromisers in English: A cross-disciplinary approach to their interpersonal significance." *Journal of Pragmatics* 7, 191-206.
Lakoff, George. 1973. "Hedges: A study in meaning criteria and the logic of fuzzy concepts." *Journal of Philosophical Logic* 2, 458-508.
Markkanen, Raija & Hartmut Schröder (eds.). 1997. *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts.* Berlin/New York: Walter de Gruyter.

Davies, Mark: A balanced 100+ million word corpus of American magazines, 1900-1999

One of the serious gaps in terms of corpora of English is the early and mid-1900s. Starting in the late 1950s and into the early 1960s, of course, we find small corpora like London-Lund, Brown, and the LOB. But for the sixty or so years before that, we have almost nothing. Work is progressing on "backward" extensions of the LOB for 1931 and 1901, but each of these is only one million words in size, and these are solely for British English.

In this presentation, we will discuss a 100+ million word corpus of American magazines from the 1900s, which we have recently created. The corpus includes 10 million words from each decade in the 1900s, and includes more than 300,000 articles in total. In addition, each article is annotated for subject matter, and this can be used as part of the query.

The corpus uses essentially the same architecture and interface as the VIEW interface to the BNC (http://view.byu.edu). This architecture allows researchers to quickly and easily find the following:

The frequency of words and phrases from 1900-2007, in either table or chart format. Users can search for specific words and phrases, alternates, or use wildcards and part of speech: e.g. *dis*tion, completely/totally* [vvd], *put * the* [nn*], etc.

Words and phrases that have a given frequency in one period, as opposed to another. For example, all words that appear for the first time in 1975, or all words that are used in the late 1940s but then disappear after 1950.

The collocates of a given word over time, and limited (if desired) by time period. For example, z-score ranked collocates of *woman* by decade, all collocates near *engine* that occur for the first time in the 1990s, or all words in the slot *I'm – that* that disappear after the 1930s.

Finally, we will briefly consider the role this corpus plays in a larger 200+ million word corpus of English (1500s-1900s), which we are currently constructing.

De Cock, Sylvie: Contracted forms in NS and NNS speech and writing

The use of contracted forms of auxiliary verbs is widely presented as typical of informal speech and writing. Learners of English, on the other hand, appear to be somewhat reluctant to use contracted forms in informal speech while they do not hesitate to use them in formal writing (De Cock 2003). This paper sets out to investigate the use of contracted forms in native and non-native informal spoken English and in native and non-native formal argumentative essay writing. The corpora used include comparable native and non-native corpora of informal spoken English, namely LOCNEC (the Louvain Corpus of Native English Conversation) and LINDSEI (Louvain International Database of Spoken English Interlanguage), and comparable native and non-native corpora of formal argumentative essay writing, namely ICLE (International Corpus of Learner English) and LOCNESS (Louvain Corpus of Native English Essays) (Granger 1998, De Cock 2004).

After giving an overall picture of the use of contracted forms in NS and NNS speech and writing, the paper focuses more specifically on the contracted forms of *be* in negative constructions and compares the use of '*not*-contractions' with 'operator contractions' (i.e. it **isn't** vs. it**'s not**) (Biber et al. 1999) in informal NS and NNS speech. Biber et al. (1999) and Carter and McCarthy (2005) have shown that 'operator contraction' is the preferred form when the subject is a pronoun and when the contracted form does not occur is clause-initial position. This study seeks to find out whether the use of '*not*-contractions' and 'operator contractions' is determined by the same factors in LOCNEC and in the spoken productions of advanced EFL learners of a variety of mother tongue backgrounds (e.g. French, German or Chinese).

References
De Cock, S. (2003) *Recurrent sequences of words in native speaker and advanced learner spoken and written English.* Unpublished PhD dissertation. Louvain-la-Neuve: Centre for English Corpus Linguistics: Université catholique de Louvain.
De Cock, S. (2004) 'Preferred sequences of words in NS and NNS speech'. *Belgian Journal of English Language and Literatures*. Academia Press. 225-246.
Granger, S. (1998) 'The computerized learner corpus: A versatile new source of data for SLA research'. In S. Granger (ed.) *Learner English on Computer*. London: Longman. 3-18.
Biber, D., Johansson, S., Leech, G., Conrad, S. and E. Finegan (1998) *Longman grammar of spoken and written English.*
Carter, R. and McCarthy, M. (2005) *Cambridge Grammar of English*. Cambridge: Cambridge University Press.

de Klerk, Vivian: The features of 'teacher talk' in a corpus-based study of Xhosa English

It has long been an accepted tenet of language teaching and language learning that teachers are key linguistic 'gatekeepers' (alongside dictionaries, textbooks and the media etc.) in entrenching the approved or 'standard' model of language that is passed on from one generation to the next. In South Africa, though, there is more than one model of English which is taught in our schools, given the complex socio-political history of the country: traditional Standard English, and various Black South African Englishes (BSAE) - generally regarded today as the varieties of English commonly used by mother-tongue speakers of South Africa's indigenous African languages in areas where English is not the language of the majority. This paper concerns itself with a particular sub-type of BSAE, Xhosa English, and with the nature of patterns of interaction in English classrooms where teachers are Xhosa speakers. The paper is based on analyses of two separate corpora: one is a corpus of the spontaneous spoken English of mother-tongue Xhosa speakers (de Klerk 2006), and the second is a (smaller) corpus of the English of mother-tongue Xhosa-speaking teachers in the classrooms in which they teach (in schools in Grahamstown).

In the first part of this paper, I report on usage of selected linguistic characteristics in the corpus of teacher talk, many of which have been identified as prototypical features of BSAE to a greater or lesser degree (e.g., Gough 1996, Mesthrie 1997, van der Walt and van Rooy 2002; de Klerk 2003). The second section reports on usage of selected discourse characteristics typically associated with classroom talk (as described, for example, by Crawford 2004) and the specific focus is again on the fact that the teachers in this corpus are L2 teachers. The question underlying the analysis presented here is whether their discourse is the same as that of L1 teachers

References:

Crawford Camiciottoli, B. 2004. Interactive discourse structuring in L2 guest lectures: some insights from a comparative corpus based study. *Journal of English for Academic Purposes* 3: 39-54.

De Klerk, V. 2003. Towards a norm in South African Englishes: the case for Xhosa English. *World Englishes* 22 (4): 463-481.

De Klerk, V. 2006. *Corpus Linguistics and World Englishes: An Analysis of Xhosa English*. London: Continuum.

Gough, D. 1996. Black English in South Africa. In V. de Klerk, ed. *Focus on South Africa*. Amsterdam: John Benjamins, 53-78.

Mesthrie, R. 1997. A sociolinguistic study of topicalisation phenomena in South African Black English. In E. W. Schneider, ed. *Englishes Around the World: Studies in Honour of Manfred Görlach*. Amsterdam: John Benjamins, 119-140.

van der Walt, J. and B. van Rooy. 2002. Towards a norm in South African Englishes. *World Englishes,* 21 (1): 113-128.

Egan, Thomas: Complement constructions with let and allow: similarities and differences

The first definitions of the word *let* in four corpus-based English dictionaries, *The New Oxford Dictionary of English*, *Collins Cobuild English Dictionary*, the *Cambridge International Dictionary of English* and the *Longman Dictionary of Contemporary English* all contain the word *allow*. One might therefore be forgiven for thinking that *let* and *allow* are near synonyms. This, indeed, is precisely the contention of Hudson *et al.* (1996), who argue that the meanings of the two are so closely related as to put "the onus […] on those who think otherwise not only to demonstrate the differences in meaning but also to show why the syntactic differences follow from them" (p. 440).

Other scholars, including Cotte (1982), Mittwoch (1990) and Duffley (1992), maintain that there is a basic difference in meaning between the two forms. In this paper I present the results of a study of 1,000 tokens of *let* and *allow* downloaded at random from the BNC. I show that one of the distinctions posited by Cotte, Duffley and Mittwoch, related to the eventual realisation or non-realisation of the situation in the complement clause, is not reflected in the corpus data. On the other hand, the *let* and *allow* constructions clearly differ with respect to the distinction drawn by Talmy (1986, 2000) between what he calls *onset* and *extended letting*.

References
Cotte, P. (1982). Autour de 'to'. *CIEREC Travaux, xxxv*, 57-80.

Duffley, P. J. (1992). *The English infinitive*. London: Longman.

Hudson, R., Rosta, A., Holmes, J., & Gisborne, N. (1996). Synonyms and syntax. *Journal of Linguistics, 32*(2), 439-446.

Mittwoch, A. (1990). On the distribution of bare infinitive complements in English. *Journal of Linguistics, 26*(7), 103-131.

Talmy, L. (1986). Force dynamics as a generalization over 'causative'. In D. Tannen & J. E. Alatis (Eds.), *Languages and linguistics : the interdependence of theory, data, and application* (67-85). Washington, D.C.: Georgetown University Press.

Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, Mass.: MIT Press.

Evert, Stefan: Context graphs

Many, if not most corpus linguists still rely on sorted concordances as their main tool of data analysis, especially when they study the usage and meaning of a word or a phrase. Reading such concordances properly is a difficult art, though, and considerable experience is required to pick out the major usage patterns quickly without losing sight of less frequent constructions. With rapidly increasing corpus sizes (news collections and Web corpora may comprise billions of words of text nowadays), it becomes more and more important to abstract away from individual concordance lines.

The most widely accepted method of summarising concordance data is collocation analysis. However, this approach puts a disproportionate emphasis on isolated collocates, while their syntagmatic and paradigmatic interconnections are lost. For instance, the noun "bucket" collocates with "brigade" and "algorithm" in the BNC, but the collocation analysis gives no indication that both collocates belong to the same compound term "bucket brigade algorithm". A further drawback of collocation analysis is that only the most common usage patterns are reflected in collocation lists, while infrequent examples that fall outside the norm are hidden from view (for linguistic purposes, these examples are often the most rewarding ones).

In my talk, I will suggest an alternative solution that "folds" large concordances into trees and more general graph structures. Such "context graphs" provide a compact representation and visualisation of concordance data, and have the potential to highlight frequent constructions as well as infrequent "outliers" at the same time. A simple implementation of the context graph approach collapses the left or right context of the node into a tree that is similar to, but more concise than a sorted concordance. More sophisticated implementations building on statistically salient collocates lead to rich and complex graph representations.

Facchinetti, Roberta: The pragmatic force of interrogatives in different varieties of English

It is widely acknowledged that a large number of questions are not to be interpreted as mere interrogations, since they convey a set of illocutionary values which range from polite request to order and even rebuke. Indeed, such illocutionary values tend to be intrinsically face-threatening and thus require strategic redress leading to conventionalised indirectness, so as to reduce the impact of the infringement.

Bearing this in mind, the aim of the present paper is to highlight (a) what types of illocutionary acts are preferably carried out in questions in English and (b) if and to what extent such illocutionary acts can be generalized among different varieties of English or rather partly differ from one variety to the other.

To carry out my analysis, I have focussed on the four mental verbs *know*, *think*, *see*, and *want* – which are among the most frequent lexical verbs in English – and have studied their occurrences in collocation with a central modal verb (e.g.: *can you think...?, wouldn't one want...?*"). Indeed, the four elements (a) subject, (b) modal verb, (c) mental verb, and (d) interrogative-negative form are bound in a force-dynamic relationship which conveys a higher level of subjectivity than the potency profiled by each single component of the cluster considered separately.

The patterns will be studied in the seven components of the *International Corpus of English* currently available, namely the two native varieties British English and New Zealand English and the five non-native varieties East African English, Hong Kong English, Indian English, Philippines English, and Singapore English. Each component covers a total of one million words distributed in a number of samples from a great variety of textual types.

First, the data will be analysed considering all the seven million words together, so as to highlight the most typically recurring patterns of modalized mental state verbs in interrogative contexts. Then, I will focus on every single variety in order to identify possible differences among the varieties themselves.

It will be seen that in interrogative contexts the modalized mental verbs under scrutiny are relatively limited in number in all the seven corpora, the Asian varieties being the ones which exhibit the highest frequencies; moreover, collocations with *can* and *would* are the most common; finally by far the most frequent cluster is '*can + see*' with subject *you* both in positive and in negatively polarised contexts. The semantic-pragmatic analysis of the data will highlight the following recurrent illocutionary values: (a) invitation to consider / to act, (b) suggestion, (c) rebuke / accusation. Such uses will be discussed in detail with reference both to their distribution within the different varieties and to their syntactic realizations.

References
Benjamin, Jessica (1998) *Shadow of the Other: Intersubjectivity and Gender in Psychoanalysis*, New York/London: Routledge.
Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan (1999) *Longman Grammar of Spoken and Written English*, London: Longman.
Clayman, S. and J. Heritage (2002) "Questioning presidents: Journalistic deference and adversarialness in the press conferences of U.S. presidents Eisenhower and Reagan". *Journal of Communication* 52 (4), pp. 749-775.
Emmertsen, Sofie (2007) "Interviewers' challenging questions in British debate interviews". *Journal of Pragmatics* 39: 570-591.
Heritage, John (2002) "The limits of questioning: negative interrogatives and hostile question content". *Journal of Pragmatics* 34: 1427-1446.

Fodor, Alexandra: Reduced adverbial clauses introduced by a subordinator: a diachronic study from Middle English to Modern English

My paper reports on a corpus-based study of reduced adverbial clauses introduced by a subordinator. My analysis of the corpus material so far points to the likelihood of Middle English being the period in which adverbial clauses of this type first arose in English. My purpose is to trace the development of these clauses from the earliest beginnings to Modern English.

The clauses I discuss represent two main types: they are either verbless clauses (e.g. *when young*) or non-finite clauses (e.g. *if not checked*). Both types are introduced by a variety of subordinators, such as *although, as if, as though, if, once, though, when, whenever, where, wherever, while, whilst.* In the verbless clauses the subordinator is mostly followed by an adjective.

In my analysis I pay attention to the proliferation of the different types and to developments in their frequencies. My hypothesis is that the types of adverbial clauses I discuss become increasingly frequent from Late Middle English onwards. My analysis will also address syntactic matters, such as the position of the reduced clause in a sentence. Possible variation according to text type will be studied as well as the role of linguistic contacts.

The corpora used in the present study are The Helsinki Corpus of English Texts, the Century of Prose Corpus, and the ARCHER corpus.

References
The Cambridge History of the English Language, Volume II-IV. Cambridge: CUP.
Görlach, Manfred (1997). *The Linguistic History of English*. London: Macmillan.
Huddleston, R. and G.K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge, CUP.
Kytö, Merja (1996). *Manual to the Diachronic Part of The Helsinki Corpus of English Texts*. Helsinki: University of Helsinki.
McIntosh, A., M.L. Samuels, and M. Benskin (1986). *A Linguistic Atlas of Late Mediaeval English*. Aberdeen: Aberdeen University Press.
Mossé, Ferdinand (1952). *A handbook of Middle English*. Baltimore: The Johns Hopkins Press.
OED= The Oxford English Dictionary (1961). Ed. by James A. H. Murray – H. Bradley – W. A. Craigie – C. T. Onions. Oxford: Clarendon Press.
Quirk, R. et al. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Fries, Udo: The Shipping News

ZEN, The Zurich English Newspaper Corpus, is a corpus of about 1.5 million running words of texts from English newspapers from the late 17th to the end of the 18th century. With a few exceptions, e.g. poetry, lists of figures (the lottery, lists of stocks and other financial information) and lists of names (lists of assizes, sheriffs, preachers) every text was included irrespective of whether it happened to be a news report, a proclamation, an advertisement, or any other report. In order to make comparisons between different newspapers, different years, and perhaps texts from other corpora (e.g. the *Rostock Newspaper Corpus*) possible, each individual text was marked as belonging to a specific text class, among them home news, foreign news, births, deaths, lost and found, reviews, and also ship news.

News about movements of ships were sent to London from all the major ports of England. The importance of these news items can be gathered from the fact that they were often separated from other news by a special headline 'ship news'. They show a very restricted vocabulary and syntax.

> *Portsmouth,* Jan. 7.
> This Morning sail'd for Lisbon, Sir John Jennings, with 7 English and 3 Dutch Men of War, the Transports, and a Fleet of Merchant Ships. The New-England Fleet are now coming in under Convoy of the Norwich, last from Plymouth. (1711 Post Boy 2443)

In this paper a full text-linguistic description of the text class 'ship news' will be given, but more general problems, such as the delimitation of text classes in corpora and their relation to texts outside a corpus will be addressed.

References
Fries, Udo. 2001. "Text Classes in Early English Newspapers". European Journal of English Studies 5: 167-180.
Schneider, Kristina. 2002. *The Development of Popular Journalism in England from 1700 to the Present: Corpus Compilation and Selective Stylistic Analysis*. Ph.D. Dissertation, University of Rostock.
ZEN, *Zurich English Newspaper Corpus*. 2004. (Online edition). http://es.unizh.ch.

Gehweiler, Elke: The parallel grammaticalization of adjectives in English and German: the case of *mere* and *lauter*

Both the English adjectival intensifier *mere* 'nothing more than' (cf. (1)) and its (now obsolete) Early New High German counterpart *lauter* 'nothing more than' (cf. (2)), have developed from adjectives with the meaning 'pure'. Unlike *mere*, *lauter* has further grammaticalized into a determiner and is now used to mean 'many, only' (cf. (3); see also Eckardt 2006).

(1)   The point has a particular relevance to the work of those who believe, like Foucault, following Nietzsche, that truth is a *mere* rhetorical device, employed in the interests of oppression, and say so at length.
[*BNC*: A1A]

(2)   Die grossen Herrn aber ... das all jhr thun / nichts denn *lauter* heucheley sey / aussen gleissen sie / aber inwendig sind sie aller su^enden vnd vnreinigkeit vol.
'But the great men ... that all their deeds / are nothing more than hypocrisy / they shine from the outside / but from within they are full of sins and impurity.'
[*BK*: 1578, DietrichSummaria]

(3)   Gerade … zum Jahresende kann man damit beginnen, mit Reizen wohldosiert umzugehen. Die Kinder haben so viele Weihnachtsfeiern … besucht. Und draußen sieht man vor *lauter* Lichterketten bald kein Licht mehr
'Especially ... towards the end of the year one should start to reduce the stimuli. The children have gone to so many Christmas parties. And outside one cannot see clearly because of the many fairy lights'
[*DWDS*: Frankfurter Rundschau 28.12.1998]

The present paper will first present a short corpus-based synchronic analysis of PDE *mere* and PDG *lauter*. It will then trace the grammaticalization paths of *mere* and *lauter* and attempt to explain why both expressions have come to be grammaticalized in a similar way. It will be argued that both expressions have grammaticalized from attributive uses of the respective adjective and that here the ambiguity of adjectives in attributive position, described by Bolinger (1967) as *reference* vs. *referent modification*, has played a crucial role (cf. also Huddleston/Pullum 2005: 553ff., Raffelsiefen 2004). The fact that other "privative" adjectives have grammaticalized with similar negative meanings (cf. English *sheer*, German *schier*, English *bare(ly)*, German *bloß*) shows the influence of the semantics of the source adjectives on the outcome of grammaticalization. It also indicates that a certain implicature, i.e. that a lack of diversity and too much uniformity are bad, has played a role in the semantic development of these forms.

Corpora used
ARCHER
British National Corpus
Bonner Frühneuhochdeutschkorpus (*BK*). Online: www.ikp.uni-bonn.de/dt/forsch/fnhd/.
DWDS-Corpus. Online: www.dwds.de/.
Oxford English Dictionary
References
Bolinger, Dwight. (1967). 'Adjectives in English: Attribution and Predication'. In: Lingua 18, 1-34.
Eckardt, Regine. (2006). *Meaning Change in Grammaticalization. An inquiry into semantic reanalysis*. Oxford: Oxford University Press.
Huddleston, Rodney/Pullum, Geoffrey. (2005). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Raffelsiefen, Renate. (2004). *Determinsm in Grammaticalization*. Manuscript. FU Berlin

Geisler, Christer & Christine Johansson: Relativizers in British and American speech and writing

When differences between British and American English are described, it is mainly vocabulary, pronunciation and spelling that are discussed. However, there are also important differences in syntax between the two national varieties. One such syntactic difference, rarely mentioned in grammars (but see Biber et al. 1999), is the frequent use of the relativizer *that* in certain written American English genres, in restrictive relative clauses referring to a non-personal noun (see for example Jacobson 1989). British English has the choice of *that* and *which*, the latter being preferred in formal contexts.

Geisler and Johansson (2002) have shown that the relativizer *that* was used frequently in *spoken* formal and informal American English also: it occurred in 80%–96% of the restrictive relative clauses. The data was drawn from CSPA (Corpus of Spoken Professional English), the ANC (American National Corpus) and the SBC (The University of Santa Barbara Corpus).

The present study aims at giving a comprehensive description of the use of *that* versus the use of the *wh*-forms and, perhaps most importantly, the use of the overt markers versus the zero-construction. The study will include formal and informal spoken and written genres of American English and compare these with the corresponding genres in British English. The data will be drawn from the CSPA, the SBC, ANC and the BNC (British National Corpus). Our hypothesis is that the relativizer *that* will be frequent in both American speech and writing and that it is becoming more frequent in different genres of British English too, possibly as a result of the influence of American English.

The purpose of the study is to show important differences in the use of relativizers in British and American English, which should be highlighted when new English grammars are written and in EFL teaching in general.

References
Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
Geisler, Christer and Christine Johansson. 2002. "Relativization in formal spoken American English". In: Marko Modiano (ed.), *Studies in Mid-Atlantic English* (HS-institutionens skriftserie 7). Gävle: Gävle University Press, 87–109.
Jacobson, Sven. 1989. "Some grammatical trends in American newspaper language". In: Leiv Egil Breivik, Arnoldus Hille and Stig Johansson (eds.), *Essays on English language in honour of Bertil Sundby* (Studia Anglistica Norvegica 4). Oslo: Novus forlag, 145–154.

Gilquin, Gaëtanelle & Magali Paquot: Making the most of Contrastive Interlanguage Analysis

Since the advent of learner corpus research in the late 1980s, the learner corpus has established itself as major resource in interlanguage analysis, gradually coming to maturity and earning respectability among SLA practitioners. One method of analysis that has been very popular among researchers is Contrastive Interlanguage Analysis (CIA), which compares native and learner varieties of language (see Granger 1996). This method has highlighted an unprecedented number of features that characterize learner interlanguages. To date, however, most studies have failed to exploit the full potential of the CIA model, focusing on the comparison between a learner corpus and a native reference corpus, but neglecting the comparison of different learner corpora of the same target language. In our presentation, we will report on a large scale study of the lexical means used in the sixteen L1 sub-corpora available in the second edition of the *International Corpus of Learner English* (Granger et al. forthcoming) to fulfil organizational or rhetorical functions typically prominent in academic writing, e.g. contrasting (*on the other hand, in contrast to*), exemplifying (*for example*, X *is an example of* Y) or concluding (*in conclusion, it can be concluded that*). We will show that comparisons of several learner corpora are indispensable if we want to identify the distinguishing features of learner language at a given stage of development (cf. Bartning 1997) and that learner language "looks rather different when we look at a lot of it at once" to use Sinclair's (1991:100) phrasing. By way of illustration, we will compare data from one learner corpus with results from the ICLE as a whole. We will also show that the type of reference corpus chosen (e.g. a corpus of native student vs. professional writing) may have an influence on the results and that this choice should be made in accordance with one's research question.

References

Bartning I. (1997) L'apprenant dit avancé et son acquisition d'une langue étrangère: tour d'horizon et esquisse d'une caractérisation de la variété avancée. *AILE* 9 : 9-50.

Granger S. (1996) From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In Aijmer K., Altenberg B. and Johansson M. (eds) *Languages in Contrast. Text-based cross-linguistic studies. Lund Studies in English 88.* Lund: Lund University Press, pp. 37-51.

Granger, S., E. Dagneaux, F. Meunier & M. Paquot (forthcoming) *International Corpus of Learner English. Handbook & CD-ROM*. Version 1.2. Presses universitaires de Louvain: Louvain-la-Neuve.

Sinclair J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Gries, Stefan Th.: Finding groups in chronologically-ordered corpus data: variance-based neighbor clustering

Much corpus-linguistic research is concerned with the development of particular parameters over time. For example, in L1/L2 acquisition/learning, the syntactic development of a child/learner is approximated on the basis of how mean lengths of utterance or *t*-unit-based measures as derived from corpora change over time (cf. Shirai and Andersen 1995, Ortega 2003). Similarly, in historical linguistics, the degree of grammaticalization of an expression is approximated on the basis of how the percentages of the use of an expression as a lexical or grammatical element change over time; cf. (Svensson 2000).

It is probably fair to say that most such studies do not just want to represent this variation as continuous, but also in terms of stages. The probably best-known example is that of Brown's (1973) MLU stages, which underlie most work on L1 acquisition, but the situation is similar in most areas (cf., say, Hilpert 2006 for a diachronic example). Thus, all of these approaches must somehow identify the different stages, but so far no principled method for this has been developed. In work on language acquisition, much work uses Brown's essentially arbitrary cut-off points; elsewhere, the data are just split up into *n* equally-sized parts, where *equally-sized* variously refers to amounts of time or numbers of items included in parts.

In this paper, I introduce a completely data-driven clustering method, variance-based neighbor clustering (VNC) that solves this problem. VNC takes as input chronologically-ordered corpus data and identifies the most useful chronological stages in the data in a bottom-up fashion and on the basis of the phenomenon in question. VNC is similar to standard clustering approaches because grouping is performed objectively using quantitative information and represented graphically in the form of dendrograms. VNC differs from standard approaches because it only clusters neighboring data points, thus preserving the data points' temporal sequence. I will discuss results from two submitted case studies, one based on the tense-aspect acquisition from the Stoll corpus of Russian L1 acquisition, the other based on data regarding the prepared infinitives following *shall* from the Penn Parsed Corpora of Historical English.

References

Brown, Roger. 1973. *A first language: the early stages*. Cambridge, MA: Harvard University Press.

Hilpert, Martin. 2006. Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory* 2.2.

Ortega, Lourdes. 2003. Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. *Applied Linguistics* 24.4:492-518.

Shirai, Yasuhiro and Roger W. Andersen. 1995. The acquisition of tense-aspect morphology: a prototype account. *Language* 71.4:743-62.

Svensson, Patrik. 2000. Grammaticalization of partitive constructions. Seminar given at the Department of English, Uppsala University.

Hardie, Andrew & Shelley Bowdler: Quantifying syntactic structures for keyness analysis

Key items are those that are relatively more frequent in a text than in some reference corpus, to a level that can be shown to be significant by means of some statistical test. Keyness analyses based on wordforms, part-of-speech tags and semantic tags have been applied in numerous studies (e.g. Rayson et al. 1997, Sawyer et al. 2002, Baker 2004). However, to date keyness analysis has not been extended to the level of syntactic structure – that is, parsing.

This may be due to difficulties in identifying the elements to be counted for a syntactic keyness analysis. In the case of wordforms or tags, it is very clear what should be counted – the number of instances of each type/tag. However, parsing structures do not constitute a closed set of discrete units: instances of a structure may overlap, may contain one another, and may differ greatly in their internal composition. Thus it is not straightforward to determine exactly what should be counted.

In this paper we discuss three separate methods for delineating, and thus quantifying, instances of syntactic structures, based on skeleton parsing annotation (Leech and Garside 1991) and using a manually-parsed subsection of the BNC. This dataset is 50% spoken and 50% written, and so provides a good test bed for identifying syntactic features which, when significantly frequent, may serve to differentiate text types – in this case, looking at parse structures which are key when speech and writing are compared.

We argue that a combination of two of these methods allows the calculation of frequencies of parse structures which can meaningfully be used in a quantitative characterisation of the grammar of a text or text type. This methodology of quantifying parse structures points the way to the possibility of automatically characterising significant grammatical patterns in a text.

References

Baker, P. (2004) Querying keywords: questions of difference, frequency and sense in keywords analysis. Journal of English Linguistics, 32(4): 346-359.

Leech, G. and Garside, R. (1991). Running a Grammar Factory: on the compilation of parsed corpora, or 'treebanks'. In: S. Johansson and A.-B. Stenström (eds), *English Computer Corpora: Selected Papers and Research Guide*. Berlin: Mouton de Gruyter. pp15-32.

Rayson, P., Leech, G., and Hodges, M. (1997). Social differentiation in the use of English vocabulary: some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics* 2(1): 133-152.

Sawyer, P., Rayson, P., and Garside, R. (2002) REVERE: support for requirements synthesis from documents. *Information Systems Frontiers Journal* 4(3): 343-353.

Hasselgård, Hilde: Adverbial usage across text types

Adverbials are known to vary in realisation, placement, frequency and the meanings expressed. This study looks at adjunct adverbials in the light of all these parameters. Different semantic types of adjunct have different positional preferences and realisations, and furthermore, that realisation and placement are closely connected. Furthermore, text types differ as regards their frequency of adjuncts and their selection of semantic types. This is connected not only to differences in subject-matter across the text types but also to different conventions as to how things are expressed. One example is the relatively high frequency in academic prose of respect adjuncts elaborating the matrix predicate.

(1)   A survey of consultants' views on systems implementation (Campbell, 1989) revealed this lack of awareness to be a recurrent problem *in terms of effective choice and implementation of systems.* <ICE-GB:W2A-011>

There is also variation across text types as regards the use of the various adverbial positions. This is connected to information structure as well as discourse characteristics of the different positions. For example, sports commentaries have a characteristic information structure because they are largely directly broadcast play-by-play commentaries, thus favouring a word order reflecting *experiential iconicity* (Enkvist 1981), as in (2), where the direction adjuncts follow the trajectory of the football.

(2)   Plays it *square || infield || to Parker who's pushed forward* <ICE-GB:S2A-001>

Other text types utilise adverbial positions to navigate e.g. contingency relations in the text or cohesive relations across sentences and utterances.
        The investigation is based on selections of the ICE-GB corpus including the following text types: conversation, sports commentary, social letters, press reportage, fiction and academic writing. The interpretative framework is functional, taking as its starting point Quirk et al (1985) and Biber et al (1999), but also drawing on insights from systemic-functional grammar (Halliday 2004).

References
Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan. 1999.
        Longman Grammar of Spoken and Written English. London: Longman.
Enkvist, Nils Erik. 1981. Experiential iconicism in text strategy. *Text*, 1 (1), 97-111.
Halliday, M.A.K. 2004. *An Introduction to Functional Grammar.* 3rd edition, revised by
        Christian M.I.M. Matthiessen. London: Arnold.
Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik. 1985. *A Comprehensive
        Grammar of the English Language*. London: Longman.

Heuberger, Reinhard; Manfred Markus & Alexander Onysko: Joseph Wright's English Dialect Dictionary in Electronic Form: a critical discussion of its parameters and query routines

While the entries in Wright's substantial dictionary (1898-1905) are highly variable and confusing at first sight, they allow systematic interpretation on second or third thought. Our paper will present the main problems caused by the structure of the entries, with their eight slots, and offer tentative solutions in view of computerisation of the dictionary. The 'head' of the entries ideally comprises the lemma followed by information on part of speech, usage labels, dialect markers and by the phonetic transcription. In the first part of the presentation, we will particularly focus on the markers of dialect, giving a survey of Wright's understanding of English dialect and of the possibility to map this information electronically. Then we will be concerned with the 'body' of the entries, which contains the meaning of a lemma, citations exemplifying its use, information on Wright's sources, and 'comments'. The paper will here again be selective and mainly deal with what Wright means by 'meaning'. Finally, we. will convey an idea of the Oracle and Java-based search mask and provide a glimpse of its layout and functions. The paper is also a state-of-the-art report on our project, which pursues the primary aim to allow convenient computer access to Wright's large storehouse of English dialect data.

References

Francis, W. Nelson 1983. *Dialectology. An Introduction*. New York: Longmans.

Holder, R.W. 2004. *The dictionary men. Their lives and times.* Bath: Bath University Press.

Lancashire, Ian, and Katharine Patterson 1997. "Learning from the Early Modern English Dictionaries Corpus," in *Tracing the Trail of Time. Proceedings from the Second Diachronic Corpora Workshop*, eds. Hickey, Raymond, Merja Kytö, Ian Lancashire, and Matti Rissanen. Amsterdam, etc.: Rodopi. pp. 47-61.

Romaine, Suzanne ed. 1998. *The Cambridge History of the English Language, IV: 1776-1997.* Cambridge: UP.

Viereck, Wolfgang, and Heinrich Ramisch 1991/1997. *The Computer Developed Atlas of England*, vol. 1 (1991) and vol. II (1997). Tübingen: Niemeyer.

Wright, Joseph 1898-1905. *The English Dialect Dictionary*. 6 vols. Oxford: Henry Frowde.

Wright, Joseph 1905. *The English Dialect Grammar*. Oxford, 1905, repr. 1968.

McArthur, Tom ed. 1992. *The Oxford Companion to the English Language*. Oxford, New York: UP.

Orton, Harold, Sanderson, Stewart and John Widdowson. 1978. *The linguistic atlas of England*. London: Croom Helm.

Heyvaert, Liesbet; Kristin Blanpain & An Laffut: A corpus-based 'lexical syllabus' for Business English. Practical issues and challenges

In this paper, we report on the development of Collex-biz, a corpus-driven learning platform of general economic English for first year university students in Economics. Business English course materials tend to pay considerable attention to methodological issues (e.g. exercise typology, inclusion of authentic texts), but the selection of items still often seems (partly) random, while low-frequency, 'difficult' words remain prominently present and the 'semantic prosody' (Sinclair 1991) of lexical items—often problematic for non-native speakers of English—is insufficiently dealt with. In Collex-biz, the ideas behind the 'lexical syllabus' set out in Sinclair & Renouf (1988:148), with its emphasis on "the commonest word forms in the language", "their central patterns of usage" and "the combinations which they typically form" were taken as starting points. Our paper starts by discussing the procedures followed in processing the data of the specialized, six-million word corpus which we compiled. It will then illustrate how the principles of frequency, collocation and colligation were effectively implemented, with special attention to methodological problems and challenges and a number of factors which had a fundamental impact on our final selection. Issues that will be touched on include: the role of register and genre in the compilation of our corpus; the impact of didactic concerns on the selection and grouping of items, the role of colligational patterns or 'textual priming' (Hoey 2004) in the design of the exercises (e.g. *a rise of a quarter point* vs. *a quarter point rise*); the relativity of numerical cut-off points when dealing with frequency lists (see also Baker 1995); the need for continuous reassessment of statistical information; and the role of the WorldWideWeb as a back-up.

References

Baker, Paul (1995). *Using Corpora in discourse analysis.* London: Continuum.

Hoey, M. (2004) The textual priming of lexis. In Aston, G., Bernardini, S. & D.Stewart (eds) *Corpora and Language Learners*. Amsterdam: Benjamins, 21-41.

Sinclair, J. & A.Renouf (1988) A lexical syllabus for language learning. In Carter, R. & M.McCarthy (eds) *Vocabulary and Language Teaching*. London: Longman, 140-160.

Sinclair, J. (1991) *Corpus Concordance Collocation*. Oxford: OUP.

Hiltunen, Turo & Jukka Tyrkkö: *Tis commonly known to Barbers and Laundresses*:
Knowledge and knowing in English medical texts from the Middle Ages to the Present

The discursive representation of knowledge, the fundamental objective of scientific inquiry, reflects underlying epistemic conditions of scientific thought (Bates 1995). Knowledge is communicated in scientific writing by means of lexical choice, discourse conventions and the organization of information. Over the long history of vernacular medicine, the writers of each era – from scholasticism and empiricism to evidence based medicine – have had their own perspectives to knowledge, revealed by the discursive practices they employed.

In a study presented at the 14–ICEHL conference, we examined explicit references to the concept of knowledge synchronically in medical texts written during the High Middle Ages. The present study takes a diachronic perspective on the same discursive field. We investigate the use of lexical items referring to the concept of knowledge (*e.g.* knowledge, information, doctrine) from the late Middle English period to Present-day English. We analyse variation and change in the lexicon of knowledge and analyse the discursive contexts in which the terms appear, showing how these have changed over time in different subgenres within learned medicine.

The study makes use of several medical corpora compiled by members of the Scientific Thought-styles project at the VARIENG research unit. The MEMT is used for the Middle English period, and a selection of texts from the EMEMT corpus (articles from the *Philosophical Transactions* and other contemporary medical texts) represent the Early Modern English period. For the PDE period, we use a selection of research articles from academic journals and texts from the MEDICOR (Vihla 1998). The total word count of these corpora is roughly one million words.

References

Bates, D. 1995. Scholarly Ways of Knowing: An Introduction. In Bates, D. (ed) *Knowledge and the Scholarly Medical Traditions*. Cambridge: CUP.

EMEMT= Early Modern English Medical Texts. In preparation.

MEMT= Middle English Medical Texts. 2005. Compiled by Taavitsainen I., P. Pahta and M. Mäkinen. CD-ROM. Amsterdam: Benjamins.

Vihla, M. 1998. "Medicor: A corpus of contemporary American medical texts". ICAME Journal 22: 73-80.

Hoffmann, Sebastian & Gunnel Tottie: The meaning and functions of tag questions in the history of English

> (1) It's raining, *isn't it*?
> (2) He doesn't know, *does he*?
> (3) He smokes, *does he*?

Sentences (1) – (3) are variable tag questions of the type characteristic of spoken Present-Day English. With a few exceptions (especially Ukaji 1998 and Salmon 1987a, 1987b) little research has been done concerning their history, and until recently, nothing on their diachrony. Based on the Chadwyck-Healey drama corpus, Hoffmann (2006) was able to show that there has been a tremendous increase in the use of tag questions from their first appearance around 1500 to c. 1900: after having an overall frequency of c. 50 per million words (pmw) around 1600, they show a sharp rise from 1750 on to 400 pmw around 1900 in drama overall, and 1,300 pmw in comedy. Although not genuine spoken language, drama is by far the genre that best represents speech, and we can be confident that this rise mirrors the rise in spoken language: the frequency of tag questions in the demographic sample of the British National Corpus is 4,383 pmw, as shown by Tottie and Hoffmann (2006).

How can we explain the rise in the use of variable tag questions in English? In the present paper we seek to answer this question by focusing on the pragmatic functions of tag questions between 1600 and 1900, using the framework of Tottie and Hoffmann (2006). We show that although tag questions were at first primarily used to seek information and confirmation, other uses, such as expressing speaker attitude, and softening or emphasizing requests were current from the very beginning. We follow the development of attitudinal tag questions into stance markers and also examine the "facilitating" use of tag questions to involve interlocutors in conversation. Our findings are then correlated with earlier research on the increase of stance meanings as documented most recently by Biber (2004).

References

Biber, Douglas. 2004. Historical patterns for the marking of stance. A cross-register comparison. *Journal of Historical Pragmatics* 5:107–136.

Hoffmann, Sebastian. 2006. Tag questions in Early and Late Modern English: Historical description and theoretical implications. *Anglistik* 17:35–55.

Salmon, Vivian. 1987a. Elizabethan colloquial English in the Falstaff plays. In *A Reader in the Language of Shakesperean Drama*, eds. Vivian Salmon and Edwina Burness, 37–60. Amsterdam: Benjamins.

Salmon, Vivian. 1987b. Sentence structures in colloquial Shakesperian English. In *A reader in the Language of Shakesperean Drama*, eds. Vivian Salmon and Edwina Burness, 265 - 317. Amsterdam: Benjamins.

Tottie, Gunnel, and Hoffmann, Sebastian. 2006. Tag questions in British and American English. *The Journal of English Linguistics* 34:283-311.

Ukaji, Masatomo. 1998. Tag questions in Early Modern English. Paper presented at *10th International Conference on English Historical Linguistics*, Manchester.

Huber, Magnus: News from Old Bailey: The development of negative interrogatives in a corpus of spoken English, 1674-1834

The Proceedings of the Old Bailey, London's central criminal court, were published from 1674 to 1834 and contain over 100,000 trials. The corpus includes about 41 million words of direct speech and thus lends itself to the analysis of everyday language in a period that has been neglected with regard to the description of the structure, variability, and change of English.

This paper will investigate the development of negative interrogative structures in the 18th and early 19th centuries. So far, these structures have not attracted much attention in the literature, although classic studies like Jespersen (1917) and Ellegård (1953) as well as some recent studies (e.g. Hoffmann 2006) have addressed aspects of the topic.

Three competing negative interrogative structures are found in the Old Bailey Corpus, "AUX *not* NP?", "AUX NP *not*?", and the contracted form "AUX*n't* NP?" The paper will analyse the distribution and development of these variants as well as the linguistic (e.g. past vs. present), pragmatic (e.g. scope of negation, implicature) and possibly social independent variables (e.g. gender and social class) determining their distribution. Particularly striking is the virtual absence from the corpus of the contracted past forms *wasn't*, *weren't*, *hadn't*, *didn't*, *couldn't*, *mightn't*, *wouldn't*, and *shouldn't*, while contracted present forms are well attested. I will consider possible reasons for this and will also have a look at the differential development of negative interrogatives in tag questions. Another point that will be addressed is the puzzling fact that "AUX*n't* NP", the preferred form in spoken English today, gets more and more marginalized during the period under investigation, largely at the expense of "AUX *not* NP", which is marginal to unacceptable today.

References

Ellegård, Alvar. 1953. *The auxiliary* do. *The establishment and regulation of its use in English*. Stockholm: Almqvist & Wiksell.
Hoffmann, Sebastian. 2006. "Tag questions in Early and Late Modern English: Historical description and theoretical implications". *Anglistik. Mitteilungen des Deutschen Anglistenverbandes* 17: 35-55.
Jespersen, Otto. (1917) 1963/9. "Negation in English and other languages", *Selected writings by Otto Jespersen*. London: Allen & Unwin, 3-151.

Iyeiri, Yoko; Michiko Yaguchi & Yasumasa Baba: Turn-initial words in the Corpus of Spoken Professional American English

The present paper discusses how speakers initiate their turns in English discourse by investigating twelve frequently occurring turn initiators in The Corpus of Spoken Professional American English (CSPAE): (1) *I, my*; (2) *we, our*; (3) *it, its*; (4) *yes, yeah*; (5) *well*; (6) *okay*; (7) *no*; (8) *the*; (9) *and*; (10) *this, these, that, those*; (11) personal names (e.g. *John*, *Gary*, and *David*); and (12) *so*. The CSPAE includes transcripts of conversations held in four separate settings (i.e. White House press conferences, faculty meetings at the University of North Carolina, national committee meetings on mathematics tests, and national committee meetings on reading tests), allowing us to investigate differences due to different conversational types. Our multivariate analysis reveals that turn initiators like *I, my, we, our, well, the,* and *no* are typically associated with the public-speaking style of White House press conferences. The other initiators are, by contrast, more characteristically related to more spontaneous interactions as found in the committee meetings on mathematics and reading tests. Furthermore, in relation to these committee meetings, there is an additional, though slight, tendency for female speakers to use interpersonal linking devices, beginning their turns with such elements as *okay* (addressed to the previous speaker) and vocative personal names like *John*, *Gary*, and *David*. This inclination related to the gender of the speakers is seldom observed in formal settings like White House press conferences and faculty meetings. All in all, the difference in terms of settings has a stronger effect upon the way of initiating turns than the difference of the gender of the speaker.

References

Anping, He and Graeme Kennedy. 1999. "Successful Turn-bidding in English Conversation". *International Journal of Corpus Linguistics* 4: 1-27.

Tao, Hongyin. 2003. "Turn Initiators in Spoken English: A Corpus-based Approach to Interaction and Grammar", in *Corpus Analysis: Language Structure and Language Use*, ed. Pepi Leistyna & Charles F. Meyer, pp. 187-207. Amsterdam: Rodopi.

Kaltenböck, Gunther: Comment clauses in spoken English

This paper investigates the occurrence and communicative properties of spoken comment clauses in a corpus of English. Comment clauses are defined narrowly as reduced parenthetical clauses (Schneider 2006), i.e. asyndedic structures, and delimited from related categories such as discourse markers, matrix clauses and reporting clauses (cf. Kaltenböck 2005). Typical examples are given in (1).

(1) a.   You've got to *I suppose* have something very special to offer (s1a-033-154)
    b.   She's the first English girl I've spoken to for about three or four years *I think* (s1a-020 -28)
    c.   His problem *it seems* is insoluble (s2b-039-31)

The database for the investigation is provided by the British component of the *International Corpus of English* (ICE-GB), which yields a total of 830 instances of spoken comment clauses.

The analysis of the communicative function of comment clauses draws on several parameters: their occurrence in different text types, the semantic nature of their predicate and, most crucially, the scope of the comment clause and its position in the host construction. It will be shown, for instance, that there are essentially two types of scope, phrasal and clausal, determined largely by the position of the comment clause in the host construction as well as by prosodic boundedness to the left or the right. The two types of scope can be shown to differ with regard to communicative function: while comment clauses with phrasal scope can be classified as 'approximators' (in Prince et al.'s 1982 terms) and as such operate proposition-internally, mitigating the 'phrastic' (cf. Hare 1970, Schneider 2006), comment clauses with clausal scope represent 'shields' (Prince et al. 1982) and as such express degree of speaker commitment with regard to the proposition expressed and mitigate the 'neustic' (cf. Hare 1970, Schneider 2006)

References
Hare, R. M. (1970) "Meaning and speech acts". *Philosophical Review* 79: 3-24.
Kaltenböck, G. (2005) "Charting the boundaries of syntax: a taxonomy of spoken parenthetical clauses". *Vienna English Working Papers* (VIEWS) 14 (1): 21-53 http://www.univie.ac.at/Anglistik/Views0501ALL.pdf
Prince, E. et al. (1982) "On hedging in physician-physician discourse", in: Di Pietro, R. J. (ed.) *Linguistics and the professions. Proceedings of the second annual Delaware symposium on language studies*. Norwood, N.J.: Ablex, 83-97.
Schneider, S. (2006) *Reduced parenthetical clauses as mitigators. A corpus study of spoken French, Italian and Spanish*. Amsterdam: Benjamins.

Kennedy, Graeme: Semantic preference and high frequency lexical verbs in the British National Corpus

Following earlier work on the semantics of collocation by Palmer (1933), Firth (1957), Sinclair (1991), and Louw (1993), among others, Stubbs (2001:65) has described the phenomenon of 'semantic preference' as "the relation, not between individual words, but between a lemma or word-form and a set of semantically related words". The relationship in multi-word units between the different concepts of 'semantic prosody' and 'semantic preference' for example, has been described by Partington (2004) and others. Description of different kinds of semantic relationships in multi-word units has been facilitated in recent years by corpus-based research using large structured corpora and more powerful analytic tools.

The present paper builds on research reported in Kennedy (2003) by describing the sets of semantically-related words associated with a number of high frequency lexical verbs in the British National Corpus. These verbs (including *give, receive, enjoy, send, start, begin, stop, end, finish, like, love, want, prefer, hate, understand, follow, remember*) are part of a study of the linguistic ecology of high frequency English lexical verbs. The study uses frequency of occurrence and strength of collocational bonding, as revealed by the Mutual Information measure, to identify semantic relationships between the verb lemmas and the company they keep.

References

Firth, J. (1957). *Papers in linguistics, 1934-1951.* London: Oxford University Press.

Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English language teaching. *TESOL Quarterly, 37,* 467-487.

Louw, B. (1993). Irony in the text or insincerity in the writer?- the diagnostic potential of semantic prosodies. In Baker, M., Francis, G., & Tognini-Bonelli, E. (eds.). *Text and Technology*: in honour of John Sinclair. Amsterdam: Benjamins.

Palmer, H.E. (1933). *Second interim report on English collocations.* Tokyo: Kaitakusha.

Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, *9:1*. 131-156

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Stubbs, M. (2001). *Words and Phrases.* Oxford: Blackwell.

Kjellmer, Göran: Awful adjectives: A type of semantic change in present-day corpora

Semantic change is taking place in the present-day language, in the language spoken all around us and manifested in our modern corpora. The Saussurean division of language study into synchronic and diachronic is therefore not relevant to the investigation of present-day sense developments.

Semantic change observable in isolated linguistic items is both frequent and interesting in itself. More interesting, perhaps, are cases of structural change, i.e. cases where one and the same tendency can be discerned in a related group of words.
This paper will use modern corpus material in order to sketch the development of one such group, words meaning 'frightening', and will suggest that they all follow the same trend in the direction of 'impressive, overwhelming' although they differ with respect to how far they have advanced along that route. The semantic changes of some 25 words in the chosen area have been studied and their development has been illustrated with corpus material. It is concluded that their rate of semantic progress is partly dependent on the time when they entered the semantic field. The paper deals with the adjectives in the group and leaves the adverbs, although equally interesting, out of account for a later study.

Kranich, Svenja: The grammaticalization of the progressive function in Modern English: evidence from ARCHER-2

The progressive in English today mainly serves as a grammaticalized expression of progressive aspect. As such, it is clearly a part of English grammar and its use is obligatory in certain contexts e.g. in expressions referring to a dynamic event ongoing at the time of utterance as in *It is raining right now* (*It rains right now*). The impossibility of using the simple form here is, however, of quite recent date: we can see that in the nineteenth century the simple form was still acceptable in such contexts: *now I will return to Fanny – it rains* (Keats, in a letter written in July 1818, cited in Arnaud 1998:123). In order to study the process by which the progressive has acquired its more fixed status in the English tense-aspect system, I have used the *ARCHER-2* (*A Representative Corpus of Historical English Registers 2*), which covers British and American English data from the time-span from 1600-1999. The corpus includes almost 2,5 million words, distributed in a balanced way across eight different text-types. The corpus thus starts with a time span in which the progressive occurs still rather infrequently, and ends, more or less, with the present state of affairs. The development which the progressive undergoes in this period can be viewed as a secondary grammaticalization process, i.e. the development by which an already grammaticalized construction undergoes further grammaticalization, e.g., as in this case, acquiring more clearly grammatical functions (cf. Traugott forthcoming).

In order to gain a better understanding of this particular secondary grammaticalization process, all the progressives found in the British part of the ARCHER-2 have been studied one by one in their context so that their meaning could be determined. The changing distribution of the meanings of the progressive in these four centuries shall not only enable us to get a better idea of how the grammaticalization of the progressive function exactly went about, but will also allow some general comments on the use of corpora for the study of grammaticalization.

References

Arnaud, René. 1998. "The development of the progressive in 19th century English: A quantitative survey". *Language Variation and Change* 10: 123-152.
Traugott, Elizabeth Closs (forthcoming). "From ideational to interpersonal. A reassessment". In: Hubert Cuyckens, Kristin Davidse and Lieven Vandelanotte (eds.), *Subjectification, intersubjectification and grammaticalization.*(Topics in English Linguistics.) Berlin/New York: Mouton de Gruyter

Lefer, Marie-Aude & Jennifer Thewissen: Orthographic and morphological errors in learner writing. Automatic and manual annotation methods: a match made in heaven?

In this paper, we analyse the orthographic and morphological errors in the 16 subcorpora of the *International Corpus of Learner English* (ICLE) (Granger et al. forthcoming). We then zoom in on specific mother tongue backgrounds for a more fine-grained analysis of error profiles. The aim of our study is twofold. First, we undertake a comparison of two approaches to the extraction of orthographic and morphological errors: we compare a fully **corpus-driven automatic** approach and a fully **manual** approach to the annotation of errors also known as error tagging. The corpus-driven approach is inspired by Granger & Wynne (1999) in which the authors investigated the lexical richness displayed in advanced EFL writing. This approach consists in the automatic extraction of all the words that are not recognised by the USAS semantic tagger (Archer et al. 2002). Many of these words contain orthographic and/or morphological errors but the extracted items may also be instances of lexical innovation. The error tagging approach is based on the error tagging guidelines developed in Louvain (Dagneaux et al. 1998). A major difference between our two extraction methods, for instance, is that while the automatic extraction process only detects errors pertaining to non-existent words, e.g. **ininterested* instead of *uninterested*, the manual error tagging enables us to find words which exist but which are used incorrectly in a particular context, e.g. *lawless* where *unlawful* should have been used.

The second aim of our study is to distinguish between errors that are cross-linguistic invariants and which therefore occur irrespective of the learners' mother tongue (e.g. derivatives such as **happyness*, **powerfull*) and errors that are the result of transfer, a phenomenon which has remained largely unexploited in the word-formation literature (Štekauer et al. 2005:17). As noticed by Granger & Wynne, EFL learners with different mother tongue backgrounds have their own specific problems (e.g. **adecuate* for Spanish learners, compounds such as **bankaccount* for Dutch learners). For our classification purpose, we have developed a categorisation of errors which is partly based on that used by Laufer (1990). Our results have a series of pedagogical implications which range from the elaboration of L1-specific remedial CALL programmes to the improvement of the emerging field of automatic rating procedures.

References
Archer D., A. Wilson and P. Rayson (2002) Introduction to the USAS category system. Benedict project report, October 2002. Available from http://www.comp.lancs.ac.uk/ucrel/usas/
Dagneaux E., S. Denness and S. Granger (1998) Computer-aided Error Analysis. *SYSTEM: An International Journal of Educational Technology and Applied Linguistics* 26 (2), 163-174.
Granger S. and M. Wynne (1999) Optimising measures of lexical variation in EFL learner corpora. In Kirk J. (ed.) *Corpora Galore*. Amsterdam and Atlanta: Rodopi, 249-257.
Granger S., E. Dagneaux, F. Meunier and M. Paquot (forthcoming) *International Corpus of Learner English. Handbook & CD-ROM. Version 1.2*. Presses universitaires de Louvain: Louvain-la-Neuve.
Laufer, B. (1990) Sequence and order in the development of L2 lexis: some evidence from lexical confusions. *Applied Linguistics* 11(3), 281-296.
Štekauer P., D. Chapman, S. Tomascikova and Š. Franko (2005) Word-formation As Creativity within Productivity Constraints: Sociolinguistic Evidence. *Onomasiology Online* 6: 1-55.

Levin, Magnus: *Pled guilty* and *hoved into view*: Using the Web to explore morphological variation

Since the advent of modern standard corpora, one major area of study has been the differences between various varieties and genres of English. Increased corpus size and corpora from more varieties have enabled new comparisons to be made. However, there are still many very rare phenomena which require enormous amounts of data to produce statistically significant results. Corpus linguists have therefore increasingly turned to the use of the Internet as a corpus (cf. the special issue on the Web as a corpus of *Computational Linguistics* 29(3) in 2003). The present study deals with two low-frequency instances of morphological variation, *pleaded/pled guilty* and *hove/hoved into view* in British, American and Australian English.

Pled is mentioned in Quirk et al (1985: 113) as an alternative AmE form to the normally regular *pleaded*. In contrast, the preterite form *hoved* instead of *hove* is not recorded in modern standard grammars. *Heave* is described by Quirk et al (1985: 113) as being usually regular, while Huddleston & Pullum (2002: 1604) suggest that this verb is irregular (*hove*) only in its nautical sense. Nevertheless, *hoved* occurs even in edited writing (Kington 2000).

Preliminary results suggest that *pled* is rarely used in formal written AmE, and that it is virtually non-existent in BrE and AusE, *pleaded* being the preferred choice. The shorter, irregular form *pled* appears to be more common in the past participle than in the preterite, which is also the case with other variable verbs in English (e.g. *burnt* vs. *burned* (Levin forthcoming)). *Hove (into view)* is the preferred preterite form in AmE and AusE, while, interestingly, the unexpected form *hoved* seems to be the most common in BrE.

The results from the Internet searches using WebCorp (http://www.webcorp.org.uk) will be compared with the data found in newspaper CD-ROMs. These comparisons will indicate to what extent the findings from newspaper CD-ROMs can be supplemented by Web searches, and this paper will also discuss the (dis-)advantages of using the Web in corpus studies.

References

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: CUP.

Kington, Miles. 2000. The dangers of digging up the past participle. *The Independent* on CD-ROM, August 22.

Levin, Magnus. Forthcoming. The formation of the preterite and the past participle. To appear in Rohdenburg, Günter and Julia Schlüter (eds.) *One Language, Two Grammars? Differences between British and American English*. Cambridge: CUP.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Lindquist, Hans: *As far as* **revisited**

English has a number of items which restrict the topic or domain of a statement, e.g. *morally*, *from a moral point of view*, *jobwise* (Huddleston and Pullum 2002: 765–766). Rickford et al (1995) used a variety of methods to study the deletion of the verbal element in the topic restricting expression *as far as X is concerned* (e.g. *as far as money, he is OK*), finding among other things that the use of the verb-less variant accelerated after 1960. Although their material includes some British instances it is predominantly American. In a recent study of the rise of another type of topic-restrictor, viewpoint-adverbs in *-wise*, Lindquist (forthcoming) demonstrated that the development of that form started in American English and spread to British English where it eventually became more frequent than in the original regional variety.

The present paper investigates to what extent verbless *as far as*-constructions are used in present-day British English, and whether the variation between the forms in British English is parallel to that in American English or not. Buchstaller (2006:375) has shown that "travelling features are actively re-evaluated and manipulated" in the receiving language context. The aim is to determine if the development in British English displays independent traits, perhaps being part of the same general drift but at a different speed and with its own special features (cf. Mair 1998:155).

Data from the British National Corpus will be compared with the age-graded observational data of Rickford et al. Furthermore, a short-range comparative diachronic study of the construction will be made based on *The New York Times* 1990–2000 and *The Independent* 1990–2000, and finally WebCorp will be used to retrieve even more recent data.

The contextual constraints favouring loss of the verbal coda discussed by Rickford et al (syntactic complexity of the NP, mode (speech or writing), age and sex of speaker/writer, and position of *as far as* in the sentence) will be taken into account.

References

Buchstaller, Isabelle. 2006. Social stereotypes, personality and regional perception displaced: attitudes towards the "new" quotatives in the U.K. *Journal of Sociolinguistics* 10 (3) 362–381.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Lindquist, Hans. Forthcoming. Viewpoint *-wise*: The Spread and Development of a New Type of Adverb in American and British English. To appear in *Journal of English Linguistics*.

Mair, Christian. 1998. Corpora and the major varieties of English: Issues and results. In Lindquist, Hans, Staffan Klintborg, Magnus Levin, and Maria Estling (eds.) *The Major Varieties of English*. Växjö: Acta Wexionensia

Rickford, John, Norma Mendoza-Denton, Thomas Wasow, and Julie Espinoza. 1995. Syntactic Variation and Change in progress: Loss of the Verbal Coda in Topic-restricting as far as Constructions. *Language* 71:102–131.

López Couso, María José & Belén Méndez Naya: Tracking down minor functions: A close look at the declarative complementizer *how* in the history of English

The extensive literature on finite complementation has paid little attention to minor complementizers, that is, those connectives usually associated with other types of subordinate clauses, but which are occasionally used to introduce finite declarative complements. In our earlier research we dealt with the declarative use of a number of originally adverbial links which, over the course of time, developed a secondary use as declarative complementizers: *but* (*that*) (López-Couso & Méndez-Naya 1998), *if* and *though* (López-Couso & Méndez-Naya 2001), *lest* (López-Couso in press) and *as if* and *as though* (López-Couso & Méndez-Naya in preparation).

The focus of the present paper will be on another of these neglected minor links, namely *how (that)*, as in *He told me how he had read about it in the newspapers*. In contrast to the aforementioned minor connectives, recruited from adverbial subordination, the origin of declarative *how* is found within the domain of complementation itself, as its basic function is that of introducing dependent questions and exclamations, as in *I asked him how the papers could be sent to him*.

Making use of different electronic corpora covering the history of English, we shall trace the development of declarative *how* from Old English times to the present day, and explore the factors which may account for the selection of this minor link over the major complementizers *that* and zero. The use of multi-genre corpora such as the Helsinki Corpus and ARCHER will allow us to examine the potential influence not only of structural factors, such as the types of predicates and the syntactic contexts favouring its selection, but also of textual determinants of variation, such as text-type or degree of formality.

References:
Huddleston, Rodney. 1971. *The sentence in written English*. Cambridge: CUP
Lakoff, Robin. 1968. *Abstract syntax and Latin complementation*. Cambridge, Mass.: The MIT Press.
López-Couso, María José & Belén Méndez-Naya. 1998. "On minor declarative complementisers in the history of English: The case of *but*." In Jacek Fisiak and Marcin Krygier (eds.). *Advances in English Historical Linguistics*. Berlin & New York: Mouton de Gruyter: 161-171.
López-Couso, María José & Belén Méndez-Naya. 2001. "On the history of *if-* and *though-*links with declarative complement clauses." *English Language and Linguistics* 5/1: 93-107.
López-Couso, María José (in press). "Adverbial connectives within and beyond adverbial subordination: The history of *lest*." In Ursula Lenker & Anneli Meurman-Solin (eds.). *Clausal Connection in the History of English*. Current Issues in Linguistic Theory. Amsterdam & Philadelphia: John Benjamins.
López-Couso, María José & Belén Méndez-Naya (in preparation). "*It looks as if it's a complementizer*: On the use of (*as*) *if* and (*as*) *though* in declarative complement clauses."
Warner, Anthony. 1982. *Complementation in Middle English and the methodology of historical syntax*. London & Canberra: Croom Helm.

Lozano, Cristóbal & Amaya Mendikoetxea: Learner corpora and second language acquisition: a study of the production of Verb-Subject structures in L2 English

Though there is a long tradition of research into word order phenomena in Second Language (L2) acquisition, this area of enquiry has recently been given a new impetus both from theoretical developments on the form-function interplay and, crucially, from the emergence of learner corpora. We seek to identify the conditions for the production of postverbal subjects in L2 English by speakers of languages allowing 'free inversion' of the subject in V(erb) S(ubject) structures, such as Spanish and Italian (see Rutherford 1989, and, more recently, Oshita 2004).

We analyse VS vs. SV structures in the Italian and Spanish subcorpora of ICLE (Granger *et al.* 2002) and we compare our results with those obtained from a similar native English corpus (LOCNESS), thus incorporating some of the fundamental tenets of what is known as **Contrastive Interlanguage Approach** (see Granger 1996 and Gilquin 2001). Based on previous findings and our analysis of VS order in L1 English, Spanish and Italian, we hypothesize that our learners will produce VS structures only with unaccusative verbs denoting existence or appearance, (Hypothesis 1). Additionally, we predict a tendency in VS structures for S to be long or 'heavy' (Hypothesis 2) and to express new (or relatively new) information (Hypothesis 3). That is, unaccusativity is a necessary, but not a sufficient condition for learners to produce postverbal subjects.

Our three hypotheses are confirmed by the analysis of the data extracted from the corpus, as it will be illustrated with plenty of examples. This means that VS structures in our learners' interlanguage have exactly the same properties as VS structures in L1 English, as shown in the example given in (1) (Birner 1994: 254), though a large proportion of the learners' VS structures were found to be ungrammatical for other reasons. Comparison with the results obtained from LOCNESS further indicate that our learners 'overuse' the construction, VS order being relatively infrequent in L1 English.

(1)   Michael puts loose papers like class outlines in the large file-size pocket. He keeps his checkbook handy in one of the three compact pockets. The six pen and pencil pockets are always full and <u>in the outside pocket</u> go <u>his schedule book, chap stick, gum, contact lens solution and hair brush</u>. [Land's End March 1989 catalog. p. 95]

References

Birner, B. (1994), Information status and English inversion, *Language* 70: 233-259.

Gilquin, G. (2001), **'**The Integrated Contrastive Model. Spicing up your data', *Languages in Contrast* 3(1): 95-123.

Granger, S. (1996), 'From CA to CIA and back: An integrated approach o computerized bilingual and learner', in: K. Aijmer, B. Altenberg B. and M.Johansson M. (eds), *Languages in Contrast. Text-based cross-linguistic studies. Lund Studies in English 88.* Lund: Lund University Press, 37-51.

Granger S., E. Dagneaux and F. Meunier (2002), *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Oshita, H. (2004), 'Is there anything there when *there* is not there? Null expletives in second language data', *Second Language Research* 20: 95-130.

Rutherford, W. (1989). 'Interlanguage and pragmatic word order', in: S. Gass and J.

Schachter (eds.), *Linguistic Perspectives on Second Language Acquisition*, Cambridge:

Cambridge University Press, 163-182.

Lyne, Susanna: Subjects of verbal gerunds in Late Modern English: Variation between genitive / possessive and common-case / objective forms

The paper focuses on the variation between genitive and common-case subjects of verbal gerunds in Late Modern British English, as in *having insisted on her brother's/her brother sending in his poem*. The corresponding variation with personal pronouns involves possessive and objective forms: *...his/him sending in his poem*.

The overarching aim is to give a Late Modern English diachronic overview of this aspect of variation. The study fills a gap in previous research on the topic, which has investigated, on the one hand, the development of the verbal features of the English gerund (e.g. Fanego 2004), and, on the other hand, present-day usage from a genre perspective (Heyvaert et al 2005, Lyne 2006).

Based on a large-scale study of British English fiction published in the latter halves of the 18th, 19th and 20th centuries, the paper draws attention to the overwhelming majority of genitive forms in the earliest period and the steep decline seen in the 20th century. This development is discussed in the light of the prescriptive rule put forth by several 18th-century grammarians, stating that the genitive form was the only correct and acceptable variant.

In the study, the usage of possessive personal pronouns is compared with the usage of genitive NPs. Further, internal features of the NP (plurality, animacy, etc.) are seen to have had played a role in the decline of the genitive.

As regards methodology, the paper discusses the style-shift the fictional genre has undergone during the period of interest (cf. Biber and Finegan 1989), and how this development might have been related to the frequency of the verbal gerund clause in general, and specifically the use of the genitive NP or possessive pronoun as subject.

References

Biber, Douglas, and Edward Finegan. 1989. "Drift and the evolution of English style: A history of three genres". *Language* 65: 487–517.

Fanego, Teresa. 2004. "On reanalysis and actualization in syntactic change. The rise and development of English verbal gerunds". *Diachronica* 21 (1): 5-55.

Heyvaert, Liesbet, Hella Rogiers and Nadine Vermeylen. 2005. "Pronominal determiners in gerundive nominalization: A 'case' study". *English Studies* 86 (1): 71-88.

Lyne, Susanna. 2006. "The form of the pronoun preceding the verbal gerund. Possessive or objective?" *ICAME Journal* 30: 37-53.

Mahlberg, Michaela & Dan McIntyre: Casino Royale: a Case for Corpus Stylistics

Sinclair (2004: 51) points out: "[l]iterature is a prime example of language in use; no systematic apparatus can claim to describe language if it does not embrace the literature also". Corpus linguistics stresses the importance of data as evidence of language use. The study of literary texts, however, has not received much attention yet within corpus linguistics, not least because it seems that for the corpus methodology, literature presents a difficult task. Although creativity and literariness can be described against the background of norms of language, the features that are crucial for one specific text are not always straightforward to identify and describe. The present paper looks at the example of Ian Fleming's novel *Casino Royale* and explores corpus stylistic approaches to the study of this text. We will look at issues that are raised by frequency information for a single novel, and we will discuss routes for analysis that are suggested by computer key words and the comparison of one novel with a general purpose corpus such as the BNC and more specific collections of prose fiction. The paper is not only interested in the description of a particular novel, but also aims to discuss general theoretical questions. It will be argued that local categories of description are crucial to capture textual features in literary texts, and – at the same time – have to play a part in the systematic apparatus that corpus linguistics suggests to describe language.

References

Carter, R. 2004. *Language and creativity. The art of common talk.* London: Routledge.

Hardy, D. 2004. 'Collocational Analysis as a Stylistic Discovery Procedure: The Case of Flannery O'Connor's Eyes', *Style* 38(4): 410-27.

Korte, B. 1997. *Body Language in Literature*. Toronto: University of Toronto Press.

Mahlberg, M. 'Corpus stylistics: bridging the gap between linguistic and literary studies' in Hoey, M., Mahlberg, M., Stubbs, M., Teubert, W. *Text, Discourse and Corpora.* London: Continuum. Scott & C. Tribble. 2006. *Textual Patterns. Key Words and Corpus Analysis in Language Education*. Amsterdam: Benjamins

Sinclair, J. 2004. *Trust the Text. Language, corpus and discourse.* London: Routledge.

Stubbs, M. 2005. 'Conrad in the computer: examples of quantitative stylistics Methods', *Language and Literature*, 14 (1): 5-24.

Marqués Aguado, Teresa: A Corpus-based analysis of modality in specialised discourse

In the last decades, there has been a growing interest in how specialised discourse is constructed (e.g. Gotti and Dossena 2001). In spite of the objectivity which is usually attributed to specialised discourse, subjectivity eventually surfaces when using modality, whereby the author takes a stance on what is being argued, thus revealing his/her commitment to the factuality of the statements exposed.

Accordingly, the aim of the present paper is to discuss the use of modality in specialised discourse by resorting to a corpus of research articles (abstracts included) compiled from several journals devoted to the field of Engineering. This corpus of written texts, which is a sizeable sample of more than 3 million words, will allow us to analyse how several linguistic strategies (that is, nominal, adjectival, adverbial and verbal) other than modal verbs are used to express modality. Similarly, we will also examine the particular frequency and distribution of such meanings along the various sections of the article, highlighting any possible difference observed between them.

References
Facchinetti, Roberta, Manfred Krug & Frank Palmer. 2003. *Modality in Contemporary English*. Berlin – New York: Mouton de Gruyter.
Gotti, Maurizio & Marina Dossena. 2001. *Modality in Specialised Discourse*. Bern: Peter Lang.
Palmer, Francis R. 2001. *Mood and Modality*. Cambridge: Cambridge University Press.

Meierkord, Christiane: Interactions across second language Englishes in South Africa - Issues of corpus compilation and annotation

Corpora of international varieties of English have increasingly been collected over the last few decades, both in their written (e.g. the Kolhapur Corpus and the subcorpora of ICE) as well as in their spoken form (the subcorpora of ICE, and e.g. most recently De Klerk 2006). Interactions between speakers of different international Englishes, however, have only seldom been attended to by corpus linguists.

This presentation introduces a corpus which has been collected, transcribed an analysed over the last three years. It consists of 25 conversations between speakers of Black South African English, Cape Flats English, and Afrikaans English. The data was compiled to investigate whether the individual varieties of English that exist in South Africa cross-influence each other after interethnic contact has increased in the post-apartheid era.

In the first part of the presentation, the focus will be on the sociolinguistic and corpus linguistic considerations which were taken into account to approach the above research question and which underlie corpus design and compilation as well as data collection and transcription.

I shall then move on to describe how the corpus is currently annotated to allow for more detailed and less time-consuming analyses of processes which resemble those that have been found to characterise situations of dialect contact in the UK (Kerswill 2003). Following Dagneaux et al. (1996) and van Rooy and Schäfer (2002), a tag set for handling variety-specific characteristics is being developed.

References

Dagneaux E., Denness S., Granger S. and Meunier F. (1996). *Error Tagging Manual Version 1.1*. Centre for English Corpus Linguistics. Université Catholique de Louvain, Louvain-la-Neuve.

De Klerk, V. (2006). *Corpus Linguistics and World Englishes. An Analysis of Xhosa English.* London: Continuum Press.

Kerswill, P. (2003). Dialect levelling and geographical diffusion in British English. In D. Britain and J. Cheshire (eds.) *Social Dialectology. In Honour of Peter Trudgill.* Amsterdam: Benjamins. 223–243.

Meierkord, C.; Minow, V.; Morreira, K. (2006&2007). *The South African Corpus of Second Language Englishes.* University of Münster.

Van Rooy, Bertus & Schäfer, Lande (2002). The effect of learner errors on POS tag errors during automatic POS tagging. *Southern African Linguistics and Applied Language Studies*, 20(4). 325-335.

Meyer, Charles: Reading between the lines: how to use corpora to study pragmatics

For the past 20+ years, I have been teaching introductory courses in linguistics, with a focus either exclusively on English or on English and other languages. During this period, I have never used an introductory text that I was particularly satisfied with: (1) many texts take a bottom-up approach to language, starting with the smallest unit of language (the phoneme) and progressively introducing larger units; and (2) while all texts provide examples of the linguistic phenomena that they describe, the examples are usually invented by the author and do not accurately reflect patterns of usage. Because of my dissatisfaction, I have begun writing an introduction to English linguistics that rectifies these problems: it takes a top-down approach to describing English, and wherever possible includes examples (e.g. text excerpts, sentences) taken from a variety of different corpora. However, basing the book on corpus data has proven to be a much more challenging undertaking than I had envisioned. In particular, it has been difficult using corpora to illustrate various phenomena falling under the rubric of pragmatics: speech acts, the cooperative principle, politeness, and conversation analysis.

While it is possible to select examples from Vanderveken's (1990) taxonomy of performative verbs in English and search for them in corpora, it turns out that with the exception of the expressives *apologize* and *promise*, performative verbs are quite rare in corpora. Thus, in finding examples of speech acts, one is left with the task of guessing speaker intentions in transcriptions of conversations with no accompanying audio. Likewise, there are no algorithms available to search for violations of the various maxims of Grice's (1989) Cooperative Principle. Since most corpora consist of texts excerpts, it is difficult to use them to do the kind of conversation analysis described in Sacks et al. (1974). But while these areas posed challenges, other areas, such as politeness, did not. Politeness is such a pervasive feature of English that corpora provide excellent resources for studying such phenomena as Leech's (1986) notion of tact. In addition to discussing my use of corpora to study pragmatics, I want to consider two questions: (1) if such phenomena as performative verbs are lacking in corpora, what does this say about the viability of speech act theory in general? and (2) is it possible to design corpora in a manner that makes them better suited for pragmatic analysis?

References

Grice, H. Paul. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Leech, Geoffrey N. 1983. *Principles of Pragmatics*. London: Longman.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A Simplest Systematics for the Organization of Turn-Taking for Conversation" *Language* 50. 696-735.

Vanderveken, Daniel. 1990. *Meaning and Speech Acts: Principles of Language Use*, Vol. 1. CUP.

Mindt, Ilka: Assessing the status of *make certain*, *make clear*, and *make sure*: are they multi-word verbs?

This paper deals with the question whether *make certain*, *make clear*, and *make sure* are multi-word verbs and as such constitute one verb that consists of two elements: the verb *make* (and any of its morphological forms) and an adjective as in (1) to (3). The examples for this study are taken from the British National Corpus (BNC).

(1) *We'll **make certain** that Lewis knows all the facts and can't twist them.* (JXV 1010)

(2) *We have already **made clear** that we are determined to encourage management employee buy-outs.* (G3H 1388)

(3) ***Make sure** that the mug sides are vertical and the top bows out.* (A70 1876)

There is no corpus-based study up to now that has considered the status of *make certain*, *make clear*, and *make sure* in greater detail. They are described in two different ways so far:

1. *Make clear* (see (2)) is considered as a variant of *make it clear* ((4) below), where the pronoun *it* need not be present. The *that*-clause (analysed as the direct object) is postponed by extraposition (e.g. Huddleston, 1984: 453).

(4) *The spokesman **made it clear** that Diana was welcome at Sandringham.*(CBF 1644)

2. *Make certain* and *make sure* have been referred to as collocations (Quirk et. al, 1985: 1198), complex verbs (Biber et al., 1999: 932), or verbal idioms (Huddleston & Pullum, 2002: 289f; 978).

In my paper I will consider both descriptions and relate them to corpus data. An analysis of the six possible combinations (1. *make it certain*, 2. *make certain*, 3. *make it clear*, 4. *make clear*, 5. *make it sure*, 6. *make sure*) accounts for their frequencies as well as semantics. The verbal status of *make certain*, *make clear*, and *make sure* will be assessed in terms of syntactic as well as semantic characteristics.

References
Biber, Douglas/Johansson, Stig/Leech, Geoffrey/Conrad, Susan/Finegan, Edward (1999) *Longman Grammar of Spoken and Written English*. London: Longman.
Huddleston, Rodney (1984) *Introduction to the Grammar of English*. Cambridge: CUP.
Huddleston, Rodney/Pullum, Geoffrey K. (2002) *The Cambridge Grammar of the English Language*. Cambridge: CUP.
Quirk, Randolph/Greenbaum, Sidney/Leech, Geoffrey/Svartvik, Jan (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

Moessner, Lilo: How representative are the Philosophical Transactions of the Royal Society of 17th-century scientific writing?

Representativeness, which was stressed as a necessary requirement of corpora by the pioneers of sociolinguistics (Labov 1966, Trudgill 1974), was the explicit aim of the compilers of ARCHER (Biber/Finegan/Atkinson 1994: 4f.). Diller (2001:18), who raises some doubt about the representativeness of Biber's PDE corpus, is much more confident about the representativeness of ARCHER because of its "carefully controlled sampling procedure". But his arguments against Bazerman's analysis of articles from the *Physical Review* can also be transferred to the register 'science' in ARCHER, which is composed of articles of the *Philosophical Transactions of the Royal Society* (PTRS) only.

In my paper I will test this hypothesis. The subcorpus of ARCHER which contains the texts of the register 'science' for the period 1650-1699 will be complemented by a parallel corpus of scientific texts which matches the ARCHER subcorpus in size and domain. The control corpus contains 10 scientific texts produced by members of the Royal Society in the second half of the 17th century, but not published in the Society's journal. Both corpora will be analysed with the method of multidimensional analysis. The resulting dimension scores of individual texts and the mean register scores of the two corpora will be compared.

Results of a pilot study (Moessner forthc.) allow the conclusion that the register scores will differ especially on the dimensions 'situation-dependent vs elaborate reference' and 'non-impersonal vs impersonal style'. These differences will be explained as a consequence of the purpose of the PTRS. They were intended as brief newsletters for the members of the Royal Society and comparable learned circles on the continent (Gotti 2006: 45). They are not more representative of the register 'science' than the longer and carefully planned publications of the parallel corpus.

References

Bazerman, Charles. 1988. *Shaping Written Knowledge. The Genre and Activity of the Experimental Article in Science*. Madison: The University of Wisconsin Press.

Biber, Douglas. 1993. Representativeness in Corpus Design". *Literary and Linguistic Computing* 8: 241-57.

Biber, Douglas, Edward Finegan and Dwight Atkinson. 1994. "ARCHER and its challenges: Compiling and exploring a representative corpus of historical English register". In *Creating and Using English Language Corpora. Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora, Zürich 1993*, Udo Fries, Gunnel Tottie and Peter Schneider (eds.), 1-13. Amsterdam/Atlanta, GA: Rodopi.

Gotti, Maurizio. 2006. "Disseminating Early Modern Science: Specialized News Discourse in the *Philosophical Transactions*". In *News Discourse in Early Modern Britain. Selected Papers from CHINED 2004*, Brownlees, Nicholas (ed.), 41-70. Bern: Peter Lang.

Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Moessner, Lilo. forthcoming. "The Influence of the Royal Society on 17th-Century Scientific Writing".

Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge: University Press.

Mollin, Sandra: The Hansard parliamentary transcripts: Basis for a new mega-corpus of spoken British English?

It is a corpus linguist's dream to be able to download large amounts of already transcribed spoken text from the internet in order to compile a large corpus of speech, even if it represents only one single genre (e.g. Hoffmann 2007). One such opportunity could be the Hansard transcripts of parliamentary proceedings in the British House of Commons. Based on the number of sitting hours (House of Commons 2005: 10) and an assumed average of 10,000 words produced per hour, the Hansard website provides an estimated 200 million words from the past 18 years for download.

This paper will firstly introduce technical issues and their solutions concerning the downloading of the transcripts and their conversion into a searchable text corpus, focusing mainly on issues of speaker identification.

Secondly, the paper will concentrate on the accuracy of the transcripts made by the Hansard staff. Slembrouck (1992) has already indicated that the transcripts may differ from the original proceedings in several respects. While Slembrouck has analysed a small number of examples from a discourse analyst point of view only, I will give a detailed quantitative analysis of the changes introduced by the transcribers based on an accurate transcription made from a recording of four hours of proceedings in the House of Commons. There are three categories of 'normalisations': (1) the 'correction' of performance characteristics of spoken language such as repetitions, reformulations or incomplete sentences; (2) the omission of extra-factual material like hedges or metacomments on turn-taking; (3) the alteration of grammatical choices by speakers towards more conservative preferences, for instance regarding concord with collective nouns, the distribution of modals of necessity (*must* vs. *have to*), or choosing a bare infinitive versus a *to*-infinitive after the verb *help*.

The paper will, finally, discuss ways in which Hansard may be used as a corpus despite transcription inaccuracies, including the automatic annotation of normalisations.

References
Hoffmann, Sebastian. 2007. "From web page to mega-corpus: the CNN transcripts." In: *Corpus Linguistics and the Web*. Eds Marianne Hundt, Nadja Nesselhauf and Carolin Biewer. Amsterdam: Rodopi, 69-85.
House of Commons. 2005. "Factsheet P4. House of Commons Information Office. Sittings of the House." Available from <http://www.parliament.uk/factsheets>.
Slembrouck, Stef. 1992. "The Parliamentary Hansard 'Verbatim' Report: The Written Construction of Spoken Discourse." *Language and Literature* 1(2): 101-119.

Mukherjee, Joybrato; Sebastian Hoffmann & Marianne Hundt: Indian English – an emerging epicentre? Insights from web-derived corpora of South Asian Englishes

In research into New Englishes, it has been suggested that English has turned into a genuinely pluricentric language in the late 20th century and that various regionally relevant norm-developing centres have emerged that exert an influence on the formation and development of the English language in neighbouring areas, e.g. Australian English as an 'epicentre' for the South Pacific (cf. Leitner 1992). In the present paper, we focus on Indian English, the largest institutionalised second-language variety of English, and its potential role as an emerging epicentre in South Asia. Specifically, we are interested in determining to what extent Indian English as the dominant variety in the region shapes the norms in Standard(ising) Englishes in the neighbouring countries. To this end, we will report on findings obtained from large web-derived corpora with texts from national English-medium newspapers in India, Pakistan, Bangladesh and Sri Lanka – countries that all once formed part of British India and that have retained the English language as a communicative vehicle, albeit to different extents. Of particular interest for our comparative study are linguistic features that cluster around the lexis-grammar interface, e.g. verb complementation. Our insights from web-derived corpora will help to describe the closeness and distance between Indian English on the one hand and English in Pakistan, Bangladesh and Sri Lanka on the other. In a wider setting, we expect our findings to have implications for the future description of the interrelation between Indian English and Englishes in the Indian diaspora, e.g. South Africa, Trinidad and Fiji.

Reference
Leitner, G. (1992): "English as a pluricentric language", *Pluricentric Languages: Differing Norms in Different Nations*, ed. M. Clyne. Berlin: Mouton de Gruyter. 179-237.

Nakamura, Junsaku & Michiyo Kasahara: What's good and what's bad in the BNC World Edition? Cross examination of multivariate analyses revisited

The aim of the present study is to find out which multivariate analysis, among Principal Component Analysis (PCA) used in the works by Burrows and by Tabata (See References), Principal Factor Analysis (PFA) demonstrated to be effective in many works by Biber and Quantification of Contingency Table (QCT) employed in the works by Nakamura, is the best to describe the relationships among word samples, the relationships among text categories and the relationships between word samples and text categories when a frequency table of samples across categories is given. One and the same data is fed into these three statistical techniques to compare their results.

Collocates of *good* and *bad* are extracted from the BNC World Edition. These two words are chosen to see what is considered to be good or bad in the British society in general and specifically to see their usage in different textual domains. Extraction of collocates is conducted in terms of z-scores with the span of two words before and after the node word. All the collocates with their z-scores higher than 1.96 and their frequencies equal to or more than 2 are collected, resulting in the total number of collocates of *good* to be 1,252 and the total number of collocates of *bad* 517. The cross tables showing the distributions of these collocates across 14 textual categories (9 domains in the written part and 5 domains in the spoken part) are then created and fed into statistical procedures.

According to the results obtained from the three techniques, QCT seems to be proven the best technique to reveal the relationships among textual categories, among collocates and between textual categories and collocates.

References

Biber, D. (1985) 'Investigating macroscopic textual variation through multi-feature/multi-dimensional analysis,' *Linguistics*, 23: 337-360.

———(1986a) 'Spoken and written textual dimensions in English: resolving the contradictory findings,' *Language*, 62: 384-414.

———(1986b) 'On the investigation of spoken/written differences,' *Studia Linguistica*, 40: 1-21.

———(1987) 'A textual comparison of British and American writing,' *American Speech*, 62: 99-119.

———(1988) *Variation across speech and writing*. Cambridge: Cambridge University Press.

———(1989) 'A typology of English texts,' *Linguistics*, 27: 3-43.

———(1995) *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.

Biber, D. and E. Finegan (1986) 'An initial typology of English text types,' in J. Aarts and W. Meijs (eds.) *Corpus Linguistics II*. Amesterdam: Rodopi. 19-46.

———(1989) 'Drift and the Evolution of English Style: A History of Three Genres,' *Language*, 65. 3: 487-517

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

Burrows, J. F. (1987) *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press.

———(1988) 'Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style,' *Literary and Linguistic Computing,* 2: 61-70.

———(1992) 'Computers and the Study of Literature,' in C. S. Butler (ed.) *Computers and Written Texts*. Oxford: Blackwell. 167-204.

Nakamura, J. (1991) 'The Relationships among Genres in the LOB Corpus Based upon the Distribution of Grammatical Tags,' *JACET Bulletin*, 22: 55-74.

———(1992) 'The comparison of the Brown and the LOB corpora based upon the distribution of grammatical tags,' *Journal of Foreign Language and Literature,* The College of General Education, The University of Tokushima, 3: 43-58.

———(1993a) 'Quantitative comparison of modals in the Brown and the LOB corpora,' *ICAME Journal*, 17: 29-48.

———(1993b) 'Statistical Methods and Large Corpora: A New Tool for De-scribing Text Types,' in M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins. 293-312.

———(1994) 'Extended Hayashi's Quantification Method Type III and its Application in Corpus Linguistics,' *Journal of Language and Literature,* The Faculty of Integrated Arts and Sciences, The University of Tokushima, 1: 141-192.

———(1995) 'Text Typology and Corpus: A Critical Review of Biber's Methodololgy,' *English Corpus Studies*, 2: 75-90.

———(1998) 'Kensakushita Detao Bunsekisuru (Analysis of the Data Retrieved from a Corpus),' in T. Saito, J. Nakamura and I. Akano (eds.) *Eigo Kopasu Gengogaku*: *Kiso to Jissen (English Corpus Linguistics: Its Principles and Applications)*. Tokyo: Kenkyusha Shuppan. 77-102. (In Japanese).

———(2002) 'A Galaxy of Words: Structures Based upon distributions of Verbs, Nouns and Adjectives in the LOB Corpus,' in T. Saito, J. Nakamura and S. Yamazaki (eds.) *English corpus Linguistics in Japan.* Amsterdam: Rodopi. 19-42.

Nakamura, J. and J. M. Sinclair (1995) 'The World of Woman in the Bank of English: Internal criteria for the classification of corpora,' *Literary and Linguistic Computing*, 10. 2: 99-110.

Nakamura, J. and Tomoji Tabata (2003). The Structure of the BNC World Edition Based upon the Distribution of –*ly* Manner Adverbs: Cross-Examination by Means of Principal Component Analysis and Quantification of Contingency Table. In Tahenryokaiseki wo Motiita Genre/ Style-bunnseki no Houhouron no Hikakukenkyu. (pp. 1-47). Heisei13 - Heisei14 Kagakukenkyuhi Hojyokin Kenkyu Seika Houkokusyo. No. 13610597.

Sinclair, J. (1990) *COLLINS COBUILD ENGLISH GRAMMAR*. London: Collins.

Quirk, R., *et al*. (1885) *A COMPREHENSIVE GRAMMAR OF THE ENGLISH LANGUAGE*. London: Longman.

Tabata, T. (1994) 'Dickens's Narrative Style: A Statistical Approach to Chronological Variation,' *RISSH,* 30: 165-182.

———(1995) 'Narrative Style and the Frequencies of Very Common Words: A Corpus-based Approach to Dickens's First-person and Third-person Narratives,' *English Corpus Studies,* 2: 91-109.

———(1998) 'Kopasu ni motozuku Buntaironkenkyu (Stylistic studies based on electronic corpora),' in T. Saito, J. Nakamura and I. Akano (eds.) *Eigo Kopasu Gengogaku*: *Kiso to Jissen (English Corpus Linguistics: Its Principles and Applications)*. Tokyo: Kenkyusha Shuppan. 167-187. (In Japanese).

———(2002) 'Word-Class Distribution and the Stylistic Variation in Dickens: Correspondence Analysis of His Works,' in T. Saito, J. Nakamura and S. Yamasaki (eds.) *English Corpus Linguistics in Japan.* Amsterdam: Rodopi, 165-182.

Newsome, Georgie: A corpus-based study of invariant tags in three varieties of English

Discourse markers are a feature of everyday conversation – they signal attitudes and beliefs to their interlocutors beyond the base utterance. One particular type of discourse marker is the invariant tag (InT), for example New Zealand and Canadian *eh*. Previous studies of InTs have clearly described InT uses in one language variety (e.g. Berland, 1997, on London teenage talk; Stubbe and Holmes, 1995, on NZ English), on sociolinguistic features (e.g. Stubbe and Holmes, 1995) and on single markers (e.g. Meyerhoff, 1994; Avis, 1972; Love, 1973; Gibson, 1977; Gold, 2005 on *eh*). However, the class of InTs has not yet been fully described, and the variety of approaches taken (corpus- and survey-based) does not easily allow for cross-varietal or cross-linguistic comparison.

This study investigates InTs in three varieties of English from a corpus-based approach. It lists the InTs available in NZ, British and Indian English through their occurrences in their respective International Corpus of English corpora, and compares usages across the varieties. The description offers a clearer overview of the InT class for descriptive grammars, as well as more explicit definitions and usage guides for EFL/ESL pedagogy. An unambiguous description of several InTs and their meanings will also allow more thorough comparison in studies of other English varieties.Finally, the results offer another viewpoint on the issue of representativeness in corpora with respect to regional versus national varieties of the Englishes.

References
Avis, W. (1972). So *eh*? Is Canadian, eh?. *Canadian Journal of Linguistics* 17: 89-105.
Berland, U. (1997). Invariant tags: pragmatic functions of *innit, okay, right* and *yeah* in London teenage conversations. Unpublished MA thesis, University of Bergen.
Gold, E. (2005). Canadian *Eh?:* A survey of contemporary use. Annual Conference of the Canadian Linguistic Association, University of Manitoba, Winnipeg.
Love, T. (1973). An examination of *eh* as question particle. Honours thesis, University of Alberta.
Meyerhoff, M. (1994). 'Sounds pretty ethnic, eh?': a pragmatic particle in New Zealand English. *Language in Society* 23: 367-388.
Stubbe, M. and Holmes, J. (1995). *You know*, *eh* and other exasperating 'expressions': an analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language and Communication* 15: 63-88.

Nurmi, Arja: *May seem odd to you* — **The social history of modal auxiliary** *may*

This paper aims at charting the history of the modal auxiliary *may* between 1400–1800. The material consists of the Corpora of Early English Correspondence (CEEC), altogether 5.2 million words of personal correspondence covering that time period.

May shows a declining trend during the Late Middle and Early Modern English periods in the Helsinki Corpus (1350–1420 and 1640–1710), with an apparent specialisation of the verb to favour some genres more than others (Dury 2002). In order to get a fuller picture of the development, it is necessary to create a more detailed timeline of it, with no gaps in the material. Since there is clear variation between genres, it is also advantageous to undertake the task with a single-genre corpus, such as the CEEC.

Facchinetti (2003) provides some sociolinguistic data on the use of *may* in present-day British English. There appears to be no difference in usage between males and females, or between age groups. Speaker education, however, does appear to be significant, with university-educated speakers favouring *may* more than others. This trend seems to tie in with the genre distribution attested already in the Helsinki Corpus, where the more learned genres (handbooks, educational treatises, philosophy and science) lead in the use of *may*.

The Corpora of Early English Correspondence are designed for testing social variables such as gender, social status, education and social mobility. In addition, they provide an opportunity to observe register variation in the form of the relationships between letter writers and recipients. This paper will present a detailed timeline of the development of *may* over four centuries, showing the social embedding of the auxiliary over the time studied.

References

Dury, Richard (2002) "May". In Gotti, Maurizio et al. *Variation in Central Modals. A Repertoire of Forms and Types of Usage in Middle English and Early Modern English.* Bern: Peter Lang. 83–128.

Facchinetti, Roberta (2003) "Pragmatic and sociological constraints on the functions of *may* in contemporary British English". In Facchinetti, Roberta et al. (eds) *Modality in Contemporary English*. Berlin: Mouton de Gruyter. 301–327.

O'Donnell, Michael: The UAM CorpusTool

This paper will introduce a new corpus annotation tool called the *UAM CorpusTool*. The basic concept in the tool is a 'project', which operates over a corpus (a collection of plain text files). The user then specifies a number of annotation 'layers', each one annotating the texts in a different way. For instance, a project might annotate each document in terms of:

- whole-document characteristics (register, writer, text-type);
- the generic structure of the text;
- annotation of clauses or sentences in terms of syntactic or semantic structure;
- annotation of NPs for type, recoverability, etc.

The tool can be set to automatically segment each document into sentences, or leave it to the human to select segments. In any case, segment boundaries are easily adjusted by dragging. Once a segment is selected, the user can edit the annotations on that segment in an easy manner.

Because each text file may have multiple layers of annotation, CorpusTool uses stand-off annotation: the original text file is untouched, and each annotation file (in xml format) simply indicates the char range of the segment in the text file.

Once a corpus is annotated, the user can search *across* layers, for instance, asking for "recoverable-np in passive-clause in editorial" only shows segments tagged with recoverable-np which occur within segments defined at other layers.

The tool also offers information on each text, such as word and sentence counts, lexical density, pronominal usage, etc. The user can also perform contrastive studies, nominating subsets of the corpus to be contrasted, e.g., "clause in male-author" vs. "clause in female-writer" to produce a gender study.

The tool is free, available from the web, and works on Windows, Linux/Unix or Macintosh.

O'Donnell, Matt; Michael Hoey, Mike Scott & Michaela Mahlberg: 'Where did you say that?' The positional features of words as a characteristic of meaning

The analysis of the frequency, rank and 'keyness' of words in text is foundational to a corpus-based analysis of meaning and serves to direct the concordance and collocational investigations that follow (Scott and Tribble 2006). Considerable attention has been focused on developing methods and theory relating to the patterns of meaning created by words in their immediate context in terms of collocation, colligation, semantic preference/association and semantic prosody (Sinclair 2004; Stubbs 2001; Mahlberg 2005; Hoey 2005). Limited exploration, however, has been carried out on the effect of *where* words occur in their respective texts. Hoey (2005) refers to this factor as 'textual colligation'. To test the hypothesis that words/items are primed to occur in or to avoid certain textual positions, we study the keywords found in text-initial (TI) position in a corpus of newspaper stories classified as 'home news'. We define TI position as the first non-headline/lead sentence of each article and have created a subcorpus (TISC) consisting of just these sentences. A reference corpus (NISC), consisting of all remaining (non-initial) sentences in each article is used to calculate a list of candidate keywords (Scott 1997) that by frequency could be described as primed for text initial position. Words are selected at random from these lists and their collocational, colligational, semantic and pragmatic patterns identified in both TISC and NISC. For example, '*announced*' was found to be primed for TI position with a semantic association for TIME words in R1 ('*announced yesterday*', '*announced last night*', etc.) in both TISC and NISC, but is one of the dominate usages in TISC but not in NISC. Similarly, '*announced*' was found to occur in sentence final position far more frequently in TISC than NISC and with a passive formation, '*it was announced* TIME'. The paper discusses both the methods and initial results of this analysis.

References
Hoey, M. (2005), *Lexical Priming: A new theory of words and language* (London: Routledge)
Mahlberg, M. (2005, *English General Nouns: A corpus theoretical approach* (Amsterdam: John Benjamins)
Scott, M. (1997), 'PC Analysis of keywords—and key key words', *System*, 25(1), 1-13
Scott, M. and C. Tribble (2006), *Textual Patterns: Key words and corpus analysis in language education* (Amsterdam: John Benjamins)
Sinclair, J. (2004), *Trust the Text: Language, corpus and discourse* (London: Routledge)
Stubbs, M. (2001), *Words and Phrases: Corpus studies of lexical semantics* (Oxford: Blackwell)

Oncins-Martínez, José L.: What corpora can tell us about Anglicisms in contemporary Spanish

The study of Anglicisms in contemporary Spanish has given rise to a number of works in the last few decades, e.g., Pratt (1980), Lorenzo (1996) or Prado (2001). However, since large corpora of Spanish have not been available for linguists and lexicographers until very recently (Davies & Face 2006), works like the aforementioned, and the study of Anglicisms in general, have not benefited from this useful tool.

The aim of this paper is to show how study of this pervasive linguistic phenomenon can be enriched by corpus data. It reports on research carried out on different aspects of two types of Anglicism —phraseological and semantic— in contemporary Spanish. The study makes use of two large corpora of Spanish: the CORDE* and the CREA**, available at www.rae.es.

The first part of the paper consists in a comparative analysis of the frequency of phraseological Anglicisms in CORDE and CREA. The comparison allows us to verify and assess the growing presence of a number of phraseological units in Spanish in the last two or three decades. These Anglicisms are also quantitatively analysed in terms of their distribution across text-types and genres, and their presence in European vs. American Spanish.

The second part focuses on semantic Anglicisms, e.g. *ignorar* ('not to pay attention to sb/sthg') < Eng. *ignore*. A number of examples will show that new senses from English can modify the linguistic environment and behaviour of the Spanish cognates in different ways, for example, by altering the collocational and colligational patterns of the Spanish forms (now Sp. *ignorar* can take a human as object); or by changing their semantic prosody (Louw 1993; Stubbs 2001), like *dramáticamente* (< Eng. *dramatically*).

Results are compared with data from the BNC to assess the extent to which these new usages are "anglicising" contemporary Spanish.

*CORDE (Corpus Diacrónico del Español/ Diachronic Corpus of Spanish)
**CREA (Corpus del Español Actual/ Corpus of Contemporary Spanish)

References

Davies, M. and T.L. Face (2006) "Vocabulary Coverage in Spanish Textbooks: How Respresentative Is It?", *Proceedings of the 9th Hispanic Linguistics Symposium*, eds. N. Sagarra and A. J. Toribio, Somerville, MA: Cascadilla Proceedings Project, 132-43.

Lorenzo, E. (1996) *Anglicismos hispánicos*. Madrid: Gredos.

Louw, B. "Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies", *Text and Technology*, eds. M. Baker *et al*. Amsterdam: Benjamins, 157-76.

Prado, M. (2001) *Diccionario de falsos amigos. Inglés-español.* Madrid: Gredos.

Pratt, C. (1980) *El anglicismo en el español contemporáneo*. Madrid: Gredos.

Stubbs, M. (2001) *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Oostdijk, Nelleke; Louis ten Bosch & Bert Cranen: New ways of analysing language variation: Bayesian modelling

Individual speakers and writers differ in terms of their language skills, while they also knowingly or unknowingly adapt their language use to fit the occasion and to obtain a particular communicative effect. Apparently, language is an inherently flexible system permitting variation.

Language variation can be seen to manifest itself in many different ways. While the particular choice of words is a common tell-tale characteristic of language variation, there is also variation in sentence length and complexity, word order, etc. Such variation does not occur randomly. Rather, the variation appears to be systematic such that specific linguistic variants (words and phrases, but also larger syntactic structures) are more likely to occur in a particular context than other variants.

From research into language variation carried out over the years, a picture is beginning to emerge which shows that language variation involves a great many factors (linguistic and extra-linguistic) which in some complex way interact, giving rise to the preference of one variant over another in a particular context (Quirk 1985; Quirk et al. 1985; Biber 1988; Biber et al. 1999). Previous research has been quite successful in identifying various individual factors that are at play but when it comes to the way these factors interact, we find that with the analytic means used so far this interaction has largely eluded us. This then is the reason for exploring new methods and techniques.

A technique which looks very promising is Bayesian modelling (Heckerman 1995). This approach makes it possible to model the interaction between the different factors that affect structural properties of language use. In the present paper we present the results of a study which aimed to investigate the placement of adverbials in the clause using Bayesian modelling.

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

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

Heckerman, D. 1995. A Tutorial on Learning with Bayesian Networks, *Technical Report MSR-TR-95-06*, Microsoft Research.

Quirk, R. 1995. The Survey of English Usage and Adverbial Realisations. In R. Quirk. 1995. *Grammatical and Lexical Variation in English*. 120-128. London: Longman.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Harlow: Longman.

Peters, Pam: The Mandative in spoken varieties of English

The survival of the mandative subjunctive in varieties other than British during C20, especially AmEng (Overgaard 1995), may have contributed to its apparent revival in BrEng towards the end of the century. Yet evidence for use of the mandative has come almost entirely from the written medium (Hundt 1998; Peters 2006), and from the more formal registers and institutional writing in the standard corpora (Brown, LOB, ACE, WC-NZ).

This paper focuses on the occurrences of the mandative in the speech-inclusive registers of writing (newspaper reporting, fiction), in the spoken components of several ICE corpora, and a new Australian corpus of talkback radio, to show its relatively limited role in spoken discourse. In spoken data its occurrences go with dialogue between unequal dyads, such as courtroom judgements as well as advice provided by "experts" on talkback radio. But the instances of the "should" paraphrase are also relatively low in spoken discourse, even the kinds of dialogue which are more hospitable to the mandative. Other kinds of paraphrase are readily available, e.g. the to-infinitive, depending on the complementation patterns associated with particular verbs: compare "asked that she give" with "asked her to give". Thus alternatives to the mandative seem to be selected wherever the controlling verb permits (Biber et al., 1999), except where the communicative context itself sanctions the mandative. The paper includes analysis of adjectives (e.g. "necessary") and nouns (e.g. "requirement") associated with mandative constructions, to show the complementary range of alternatives supported by them.

Taking into account this wider range of alternatives to the mandative subjunctive, some regional variation still presents itself in that the mandative is less often found in spoken BrEng than the other varieties. The data show why recent grammars (Biber et al. 1999; Huddleston and Pullum 2002) prefer to treat the mandative construction as a variant formulation of the content clause, rather than an independent verbal system. This is justified by its very limited occurrences in spoken discourse, although its use with a variety of controlling verbs, nouns and adjectives in institutional writing still supports the classical treatment.

References

Biber D. et al. (1999) Longman Grammar of Spoken and Written English London, Longman

Huddleston R. and Pullum G. (2002) Cambridge Grammar of the English Language Cambridge, Cambridge University Press

Hundt M. (1998) "It is important that this study (should) be based on the analysis of parallel corpora: on the use of the mandative subjunctive in four major varieties of English. In H. Lindquist et al. eds The Major Varieties of English. Vaxjo University

Overgaard (1995) The Mandative Subjunctive in American and British English in the Twentieth Century. Uppsala, Acta Universitatis Upsaliensis.

Peters (2006) "Description and Prescription in English usage". In Bas Aarts and A McMahon eds Handbook of English Linguistics. Oxford, Blackwells

Prado-Alonso, José Carlos: Noun phrase full inversion in scientific discourse

Over the past few years, *full-verb* inversion —constructions in which the subject follows the entire verb phrase in a declarative clause, as in "On the near corner was *Herb's Gas Station*" or "Upstairs was *a bedroom and a bathroom*"— has been the subject of extensive research (cf. Birner 1996; Dorgeloh 1997; Chen 2003; Kreyer 2004), the focus of each individual study varying according to the nature and goals of the specific theoretical framework adopted.

On the basis of the kind of phrasal category occurring as clause-initial constituent, five different types of *full* inversion have been traditionally distinguished in the literature on the topic: *noun phrase*, *adverb phrase*, *verb phrase*, *adjective phrase*, and *prepositional phrase full* inversion. Broadly speaking analyses of *full* inversion carried within the generative (cf. Bresnan and Kanerva 1992; Schachter 1992) and functional paradigms (cf. Birner 1996; Dorgeloh 1997; Chen 2003; Kreyer 2004) have focused mainly on inversions following a locative constituent, which subsumes "spatial locations, path, and directions, and their extension to some temporal and abstract locative domains" (Bresnan 1994: 75). In this sense, they neglect the analysis of those *full* inversion types, namely *noun phrase* and *adjective phrase full* inversion, which do not contain such clause-initial constituents. In order to bridge this gap, the present paper offers a corpus-based analysis of *noun phrase* inversion in Present-day English scientific texts taken from the FLOB and the FROWN corpora. It is argued that the distribution of the construction in this textual category is related to the degree of abstractness of its texts, and that *noun phrase* inversion performs a text-structuring function which gives coherence to the discourse.

References

Birner, Betty. J. 1996. *The Discourse Function of Inversion in English*. New York: Garland.

Bresnan, Joan. 1994. "Locative Inversion and the Architecture of Universal Grammar". *Language* 70, 72-131.

Bresnan, Joan and Kanerva, Joni M. 1992. "The Thematic Hierarchy and Locative Inversion in UG: A Reply to Schachter's Comments" In: Tim Stowell and Eric Wehrli (eds), *Syntax and Semantics*. *Vol. 26*: *Syntax and the Lexicon*. New York: Academic Press, 111-125.

Chen, Rong. 2003. *English Inversion: A Ground-Before-Figure Construction*. Berlin: Mouton de Gruyter.

Dorgeloh, Heidrun. 1997. *Inversion in Modern English: Form and Function*. Amsterdam: John Benjamins.

Kreyer, Rolf. 2004. *Inversion in Modern Written English. Syntactic Complexity, Information Status and the Creative Writer*. Unpublished PhD Dissertation, University of Bonn.

Schachter, Paul. 1992. "Comments on Bresnan and Kanerva's Locative Inversion in Chichewa: a case study of factorization in grammar". In: Tim Stowell and Eric Wehrli (eds), *Syntax and Semantics*. *Vol. 26: Syntax and the Lexicon*, New York: Academic Press, 103-110.

Rissanen, Matti: Varying evidence from various corpora: the grammaticalisation of *in case*

For most of the 1990s, *The Helsinki Corpus of English Texts* was the only structured long-diachrony general corpus of English, including one thousand years of English from the earliest extant texts to the 18th century. In recent years, a number of important new multi-genre corpora and databases covering the earlier periods of English have seen daylight. The most important are *The Dictionary of Old English Corpus*, *The Middle English Compendium*, and the expanded parsed versions of the Middle and Early Modern English parts of *The Helsinki Corpus*, prepared at the University of Pennsylvania (*PPCME2* and *PPCEME*). *The Oxford English Dictionary Online* is also useful for diachronic studies covering a long period of time.

The purpose of my paper is to compare the types of information that can be obtained from these electronic resources and discuss their usefulness for various research topics. I will focus on the grammaticalisation of the conditional connective (preposition and subordinator) *in case*. The noun *cas(e)* was borrowed from French in Early Middle English, and the phrase *in case* developed into a connective in the fourteenth century. At the earliest stages of grammaticalisation, a number of variant prepositional phrases formed with *case* are recorded.

In Late Middle and Early Modern English the genre distribution of *in case (that/of)* is of importance. The phrase is favoured in officialese, i.e., in statutes, documents and official letters.

It is obvious that when tracking the paths of grammaticalisation of connectives, the best results can be obtained by combining the information provided by the various corpora and databases referred to above. The results obtained from them should, however, be supplemented with evidence from special corpora, such as *The Corpus of Early English Correspondence* and *The Corpus of Early English Medical Writing*.

References

Hopper, Paul J. and Elizabeth Closs Traugott 2003 [1993]. *Grammaticalization.* Cambridge: Cambridge University Press.

Kortmann, Bernd 1997. *Adverbial subordination: A typology and history of adverbial subordinators based on European languages.* Berlin and New York: Mouton de Gruyter.

Rissanen, Matti 2000. "Standardisation and the language of early statutes". In: Laura Wright (ed.), *The development of Standard English: Theories, descriptions, conflicts.* Cambridge: Cambridge University Press. 117-130.

Rissanen, Matti 2003. "On the development and grammaticalisation of borrowed conditional subordinators in Middle English." *Studies in medieval English language and literature* published by the Japan Society for Medieval English Studies 18: 1-19.

Traugott, Elizabeth Closs and Richard B. Dasher 2002. *Regularity in semantic change.* Cambridge: C.U.P.

Römer, Ute: Corpora and academic discourse analysis: Providing insights into the lexicalgrammar of linguistic book reviews

The past few years (and ICAME conferences) have seen an increasing interest in studies based on new kinds of specialised corpora that capture an ever-growing range of text types, including academic genres, e.g. research articles or book reviews from particular disciplines. Now that more and larger collections of such specialised texts are becoming available, corpus researchers are able to investigate different community discourses and provide accurate descriptions of their lexicalgrammars.

By means of presenting selected results from a large-scale study on expressions of evaluative meaning in academic writing, the present paper will analyse in what ways lexical items and grammatical structures are interrelated in the language of book reviews, a so far largely under-researched genre. It will be demonstrated how new generation corpus tools (*Collocate*, Barlow 2004 and *ConcGram*, Greaves 2005) can be used to identify meaningful items and typical patterns in a 3.5-million word corpus of linguistic book reviews. A comparison of selected lexicalgrammatical patterns in this specialised corpus and in BNC_written as a general reference corpus will then provide further insights into the special characteristics of book review language and indicate which phrasal units authors of linguistic book reviews preferably use in order to create particular meanings.

References

Barlow, Michael. 2004. *Collocate 1.0: Locating collocations and terminology.* Houston, TX: Athelstan.

Greaves, Chris. 2005. *ConcGram Concordancer with ConcGram Analysis*. HongKong: HKUST.

Rudanko, Juhani: Exploring the spread of the *to -ing* pattern: the case of *accustomed* and *unaccustomed*

Consider sentence (1):

> 1. Familiar with crime, he was unaccustomed to be charged with it (OED, 1891, Farrar, *Darkn. and Dawn*)

In (1) the matrix adjective *unaccustomed* is found in a subject control structure and selects a *to* infinitive complement. Such complements have been common with accustomed and unaccustomed, but today the adjectives are also found with *to -ing* complements, (*un)accustomed to being charged*.

The paper provides illustrations of sentential complements of *accustomed* and *unaccustomed* in subject control structures from recent centuries, shedding light on the incidence of *to* infinitive and *to -ing* complements. The spread of the *to -ing* pattern is placed in the general framework of the Great Complement Shift (Rohdenburg 2006, Vosberg 2006), and building on recent work on *accustomed*, the paper examines factors that may have promoted or impeded grammatical change. In addition to exploring the potential impact of extraction in this connection, the paper also considers the nature of the lower predicate and investigates the question of whether the semantic role of the lower subject is a factor in such change. The evidence considered comes from large electronic corpora from the eighteenth and nineteenth centuries, including the extended version of the Corpus of Late Modern English and the new CEN datebase, both developed at the University of Leuven. For present-day English, the BNC and the full Bank of English Corpus provide the indispensable databases for examining the pattern, with the latter also permitting a comparison of current British and American English usage.

References

Rohdenburg, G. 2006. "The Role of Functional Constraints in the Evolution of the English Complementation System," in C. Dalton-Puffer et al. eds., *Syntax, Style and Grammatical Norms,* Bern: Peter Lang: 143-166.

Vosberg. U. 2006. *Die Grosse Komplementverschiebung.* Tuebingen: Narr.

Reconciling corpus evidence and classroom practice: Discourse presentation as a case in point

A growing body of applied corpus-linguistic research suggests that the representation of certain lexico-grammatical features in EFL textbooks is at variance with their use in native English (e.g. Mindt 1996, 1997; Römer 2005). One crucial *discourse* area in which the gap is particularly wide is speech, thought, emotion and sound presentation, an area which I will refer to as discourse presentation. The aim of the talk is to explore that gap and to address implications for EFL teaching.

Based on corpus-linguistic analyses of BNC data, I will discuss the use of selected features distinctive of discourse presentation in conversation: the preference for direct over indirect mode (cf. McIntyre et al. 2004), the tendency to present not only speech and thought but also emotion and sound, and the reporting verbs BE+*like* (e.g. Andersen 2001) and GO.

Analysis of selected textbooks shows that textbook representation is overwhelmingly concerned with speech but hardly thought presentation (let alone emotion and sound presentation), indirect and narratized mode but not direct mode, and reporting verbs typical of writing but not speech.

The implications of teaching authentic discourse presentation, just as teaching authentic conversation generally, are far-reaching. I will focus on what seems the most central (and controversial) implication: Standard English (SE), which has traditionally been the major model in EFL for both writing *and speech* (Quirk et al. 1985: 7), needs to be reduced to "a 'core' variety" (Bex 1993: 261) underlying teaching writing; teaching speaking, conversely, should be based on conversational grammar. I will argue that a 'register approach' is best suited to accommodate such a redefinition of SE, shifting the emphasis from a monolithic view of language to a register-sensitive view (Conrad 2000) which recognizes the functional diversity of language use (Stubbs 1993: 11).

References
Andersen, G. (2001) *Pragmatic Markers and Sociolinguistic Variation. A relevance-theoretic approach to the language of adolescents*. Amsterdam/Philadelphia: John Benjamins
Bex, T. (1993) 'Standards of English in Europe.' *Multilingua* 12-3: 249-264
Conrad, S. (2000). 'Will corpus linguistics revolutionize grammar teaching in the 21st century?' TESOL Quarterly 34: 3, 548-560.
McCarthy, M. J. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press
McIntyre, D., C. Bellard-Thomson, J. Heywood, T. McEnery, E. Semino and M. Short (2004) 'Investigating the presentation of speech, writing and thought in spoken British English: A corpus-based approach.' *ICAME* 28: 49-76
Mindt, D. (1996) 'English corpus linguistics and the foreign language teaching syllabus.' In: J. Thomas and M. Short (eds.) *Using Corpora for Language Research*, London: Longman, pp. 232-247.
Mindt, D. (1997) 'Corpora and the Teaching of English in Germany.' In: A: Wichmann, S. Fligelstone, T. McEnery, and G. Knowles. *Teaching and Language Corpora*. Harlow: Longman, pp. 41-50
Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. (1985) *A Comprehensive Grammar of the English Language*. London: Longman
Stubbs, M. (1993) 'British Traditions in Text Analysis – From Firth to Sinclair.' In: Baker, M., G. Francis, and E. Tognini-Bonelli. (eds.). *Text and Technology. In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, pp. 1-33
Römer, U. (2005) *Progressives, Patterns, Pedagogy. A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam/Philadelphia: John Benjamins.
Rühlemann, C. (forthcoming) *Conversation in Context. A Corpus-driven Approach*. London: Continuum

Rühlemann, C. (under consideration) 'A register approach to teaching conversation: Farewell to Standard English?'

Sand, Andrea: Patterns and language contact: Multiword units in the New Englishes

The present paper is a pilot study for a systematic investigation of multiword units in the New Englishes. As has been shown in previous studies, these varieties display a large number of differences to international Standard English. On the one hand, we encounter innovations, such as *suck teeth* referring to a paralinguistic expression of disapproval in Jamaica (cf. Patrick/Figueroa 2002), or loan translations from indigenous languages, such as Kenyan English *bad eye* 'the power to bewitch people' (cf. Skandera 2003: 181). More interesting from a comparative point of view, however, are those multiword units which are derived from international standard English, but have been altered in the New Englishes (cf. Platt et al. 1984: 87-115, Simo Bobda 1998: 8-9 or Crystal 2003: 163-164).

It is especially striking that several of the expressions described in previous studies are not restricted to one particular variety but appear common across several varieties, such as *with regards to* instead of *with regard to, in light of* instead of *in the light of* or *get a baby* instead of *have a baby* (cf. also Sand 1999, Skandera 2003).

To shed light on the kind and frequency of such changes, several large web-based newspaper corpora will be used to complement the smaller ICE-subcorpora. The general aim is to determine whether the interface of lexicon and grammar shows parallels similar to those found in the area of morphosyntax (e.g. Sand 2004).

References

Crystal, David. 2003. *English as a Global Language*. 2nd ed. edition. Cambridge: CUP.

Patrick, Peter und Esther Figueroa. 2002.'*Kiss-teeth*,' *American Speech* 77(4):383-97.

Platt, John T., Heidi Weber, und Mian Lian Ho. 1984. *The New Englishes*. London: Routledge.

Sand, Andrea. 1999. *Linguistic Variation in Jamaica: A Corpus-based Study of Radio and Newspaper Usage*. Tübingen: Narr.

Sand, Andrea. 2004. 'Shared morpho-syntactic features in contact varieties of English: article use', *World Englishes* 23: 281-298.

Simo Bobda, Augustin. 1998. *The Indigenization of English in Cameroon and New Englishisms*. Essen: LAUD.

Skandera, Paul. 2003. *Drawing a Map of Africa: Idiom in Kenyan English*. Tübingen: Narr.

Schilk, Marco: Correlations between collocational preferences and verb-complementational patterns in Indian English and British English

In research into Indian English, the largest second-language variety of English in terms of numbers of speakers, it has been shown that significant differences exist on the level of verb-complementational patterns when compared with British English, the historical input variety. For example, ditransitive verbs such as *give* are more frequently used monotransitively in Indian English than in British English. (cf. Mukherjee & Hoffmann 2006). However, the underlying reasons for such differences are yet largely unclear. In other corpus-based studies, different collocational preferences have been found in Indian English and British English, e.g. for *chaste* (cf. Schilk 2006). In the present paper, the two aspects of lexicogrammatical variation between the two varieties will be combined.

In this paper, I will start off from some pilot studies based on the Indian and British components of the International Corpus of English (ICE) which indicate that collocational preferences and verb-complementational patterns do correlate: for example, there is a tendency to use a monotransitive complementation pattern for the verb *give* when it collocates with the noun *idea* in the direct object position in Indian English, whereas in British English a ditransitive complementation pattern is preferred. However, due to the small size of the ICE-corpora, collocational profiles of verbs are relatively restricted. Thus, in a second step, a large web-derived corpus of Indian English will be used to describe the collocational preferences of verbs such as *give* and *offer* on a more solid basis. Finally, different collocational preferences of verbs will related to differences at the level of the complementation of the verbs under scrutiny.

References
Mukherjee, J. & S. Hoffmann (2006): "Describing verb-complementational profiles of New Englishes: a pilot study of Indian English", *English World-Wide* 27(2): 147-173.
Schilk, M. (2006): "Collocations in Indian English: a corpus-based sample analysis", *Anglia* 124(2): 276-316.

Springer, Philip: *However, …* **A corpus based analysis of advanced Dutch learners' use of connectives in writing**

The CEF states that a writer at C2 (*Mastery*) level "Can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points" (council of Europe 2001: 61). In order to achieve this level of discourse competence, advanced learner writers need a good command of the following lexico-grammatical discourse devices: focusing devices, backgrounding/clause combining devices, and text-organizing cohesive devices.

My current research compares the use of these devices in advanced learner English and native speaker academic texts. For this, two corpora, a corpus of written work by Dutch students of English and a corpus of English native speaker articles from linguistics and literature journals, were compiled and manually annotated for the discourse devices mentioned above. The outcome of this analysis will be used to determine the need for adjustment to teaching material for written skills in the light of the C1 and C2 levels of the Common European Framework.

This paper focuses on one of the lexico-grammatical devices that the corpora were annotated for, the use of connectives by Dutch advanced learners of English. Findings related to the overuse and underuse and positional features of connectives in the English writing of Dutch students will be presented.

Reference

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

Tagliamonte, Sali A.: Corpora from the virtual world: Teenagers, instant messaging and language change

Instant Messaging (IM) is a form of computer mediated communication (Baron 2003:399) in which messages are typed back and forth in real time, as in (1).

(1)   [024] do u still talk to him???
        [003] rarely
        [024] u want to msg him and see if he replies@!!!!
        [003] k

In North America, 80% of 16-19 year olds spend three hours or more a day engaged in this type of interaction with their peers — a goldmine of interactive, written language from the sector of the population most involved in linguistic change — teenagers.

In this paper, I demonstrate how a 1.2 million word corpus of IM was collected and analyzed along with a 'mirror' corpus comprising one-on-one conversations with a sub-set of the same speakers, the Interview Corpus (IV).

First, I present an analysis of IM forms, including acronyms, e.g. *brb, ttyl, gtg*; abbreviations, e.g. *bf, cya, nvm*, symbolic uses, e.g. *u, c, r* and emotional forms, e.g. *arg, haha, meh*. Second, I present an analysis of a feature currently innovating among adolescents, intensifier *so*, as in (2).

(2)   [013] omg closer is **so** good

Preliminary analyses reveal thousands of IM forms; however their proportion is extremely low. Comparative analysis of the intensifier system in IM and IV also suggests that IM is conservative. The overall rate of intensification is lower in IM (24%) than in IV (36%) and IM contains more formal intensifiers, i.e. *very*. Yet the central intensifiers are consistent across data sets with females leading in the use of *so*, the in-coming intensifier.

These findings challenge the deleterious perceptions of IM and suggest that they have been over-blown in the media. Instead, IM provides an illuminating reflection of the dynamic ongoing, normal processes of linguistic change that are currently underway in the English language (see also Herring 2003).

References

Baron, Naomi S. (2004). See you online: Gender issues in college student use of instant messaging. Journal of Language and Social Psychology. 23(4): 397-423.

Herring, Susan C. (2003). Gender and power in on-line communication. In The handbook of language and gender. Janet Holmes & Miriam Meyerhoff (Ed.), Malden: Blackwell. 202-228.

Tavecchio, Lotte: English and Dutch contrasted: discourse, grammatical and rhetorical differences in sentence structure

With English being the international language of communication, many Dutch texts that are aimed at an international readership are now produced in English. The quality of these English texts and the extent to which the English used is influenced by the writer's native language is a much discussed topic by both linguists and language professionals. This raises questions about whether there are indeed essential structural differences between such closely related languages as English and Dutch, and, if so, what the main differences are on a grammatical, rhetorical and discourse level.

To gain insight into the differences in sentence structure between English and Dutch, a self-compiled 300,000 word corpus consisting of four different genres has been annotated on both a discourse and a grammatical level. Sentences have not only been segmented into discourse units, where punctuation marks have been considered to play an important role in marking unit boundaries, but the syntactic realisation of these units has also been annotated. This makes it possible to uncover recurrent sentencing patterns in both languages, to relate such patterns to rhetorical design and to investigate the extent to which different patterns may be related to syntactic differences in the two language systems involved.

This paper will present the main results of this corpus study. It will show how sentences are typically structured in English and Dutch and what the main differences are between these two languages in this respect. In addition, it will consider what the influence is of genre on sentence structure.

Van Bogaert, Julie: A reassessment of the syntactic classification of pragmatic markers: An account of the positions of *I think* and *you know*

In the study of pragmatic markers like *I think* and *you know,* traditionally, a distinction is made between three canonical syntactic positions: initial, medial and final (e.g. Urmson 1952; Thompson & Mulac 1991). Neat and straightforward as this division may seem, when faced with the complexities of real (spoken) language data, one comes to the conclusion that it is not fine-grained enough to adequately describe the syntactic behaviour of pragmatic markers and to account for their functional properties. Therefore, I will propose in this paper a refinement of the basic three-way system in terms of, on the one hand, the syntactic level at which a particular marker occurs and on the other hand, in terms of what that marker holds within its scope. For example, the canonical tripartition would group together utterances (1) and (2) as medial uses, failing to account for the fact that in (1), *I think* is used at the level of the clause, occurring in between the clause constituents Subject and Finite, while in (2) the marker has been inserted within a phrase, separating a preposition and its complement.

> (1) And and she *I think* had been involved in <,> uhm sort of political activity in the fifties
> (2) Father McDade d'you you remember in *I think* lecture three uh Rabbi Sacks said at one point faith is not measured by acts of worship alone

Also, the canonical tripartition takes it for granted that markers have scope over the entire clause (complex) in which they occur. I would like to argue, however, that markers can be used with local scope, applying to one particular clause constituent or phrase constituent.

In this paper, I will present a syntactic description of and comparison between spoken ICE-GB data of *I think* (1733 occurrences) and *you know* (1086 occurrences)*,* the two most frequently used cognitive clauses functioning as pragmatic markers in the corpus, on the basis of a classificatory system taking into account the notions of syntactic level and scope. The data show that *you know* is used with local scope – i.e. singling out a specific clause constituent or phrase constituent - more often than *I think* (34.7% vs. 5.4%)*.* The markers' syntactic preferences are believed to reflect their functional specificities.

References

ICE-GB: International Corpus of English – the British Component

Thompson, S.A. & A. Mulac (1991) "A quantitative perspective on the grammaticalization of epistemic parentheticals in English", in E.C. Traugott & B. Heine (eds.) *Approaches to Grammaticalization.* Amsterdam: Benjamins, 313-339.

Urmson, J.O. (1952) "Parenthetical verbs", *Mind* 61*:* 480-496.

Van linden, An; Kristin Davidse & Mieke Rosselle: The clausal complementation of *(it's) important*

In the description of clausal complements controlled by adjectives, most reference grammars make only a formal distinction between *that*-clauses and *to*-clauses (e.g., Quirk et al. 1985: 1220–31; Biber et al. 1999: 671–83; 716–22). However, if these complements are looked at from a functional angle, the picture becomes more complex. In particular, the *that*-clauses may realize either (i) a *proposition*, a piece of information about whose validity speaker and hearer can argue (Halliday 1994: 70), e.g.,

>   (1) [I]t's **important** that the NEC is now dominated by members of the Shadow Cabinet. (CB)

or (ii) a *desired action*, an action assessed as desirable by the speaker or another entity, e.g.,

>   (2) It is **important** that the woman be the one who is in charge of the entire process. (CB)
>   (3) It is **important** that the information we collect is as accurate as possible. (CB)

To-complements always seem to realize desired actions, e.g.,

>   (4) The Cowboys believe it is **important** to have licensed premises at a central location in addition to their headquarters. (CB)

In this paper we want to chart the formal and functional distribution of clausal complements with the adjective *important*, drawing on data from the diachronic *Corpus of Late Modern English Texts* (CLMET) (De Smet 2005) and the synchronic *Cobuild Corpus* (CB). How do formal features such as *to*-infinitive, subjunctive, indicative or modal VP relate to the distinction between proposition and desired action? What are the relative frequencies in historical and present-day data of these two functional patterns? What rhetorical, illocutionary and attitudinal values (McGregor 1997: 239–44) do they express with regard to their complements? Do they form two distinct constructional schemata, or have they influenced each other?

   Of particular interest here is the mixed pattern in which a *to*-complement with a mental predicate (*note, remember, realise*, etc.) is followed by a proposition, e.g.,

>   (5) It is **important** to realise that in these times of fast change it can be dangerous to let things drift. (CB)

Whilst the desirability of the mental act is certainly an element of the semantics of (5), the matrix plus mental *to*-complement seems to have conventionalised into a larger constructional unit with its own semantics and pragmatics, and the pattern as a whole functionally bears more resemblance to the pattern with proposition in (1). The diachronic data even suggest that the latter developed from the mixed pattern in the course of the 19th century.

References
Biber, Douglas, Stig Johansson, Geoffrey Leech, and Randolph Quirk. 1999. *Longman grammar of spoken and written English*. London: Longman.
De Smet, Hendrik. 2005. A corpus of Late Modern English texts. *Icame Journal 29:* 69–82.
Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. Second Edition. London: Longman.
McGregor, W. 1997. *Semiotic Grammar*. Oxford: Clarendon.
Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and David Crystal. 1985. *A comprehensive grammar of the English language*. London: Longman.

Van Olmen, Daniël: *Boy, dontcha you play with me*. The emergence of a prohibitive marker in English?

It is a curious fact of the English language that *not* may but need not cliticize to *do* in subjectless prohibitives but is obligatorily reduced to *n't* in structures with subjects like *don't you do that*!. In this paper, I briefly consider recent formalist accounts of this peculiarity – such as Potsdam (1998) and Rupp (1999)– and conclude that they are not adequate as the debate amongst these scholars already shows. The phonological explanation offered by Akmaijan (1984) and substantiated by Bergs (2005) is examined in more detail, in particular the claim that the pragmatically most salient element, i.e. the additional subject, is always the sole intonational centre before the main verb. This feature of the proposal is argued to be problematic in the light of prosodic corpus data and of imperative *dontcha*, which Bergs – strangely– accepts and Akmaijan –consistently but wrongly– dismisses.

In my opinion, the non-expandability of the prohibitive with subject is best understood from the perspective of the verifiable diachronic development *do not you – don't you – dontcha* in Early Modern, Late Modern and Present-Day English (i.e. a process of cliticization to the element in sentence-initial position) as a case of grammaticalization. The advantage of this approach is that the historical occurrence of *do not you* does not need to be considered as a pragmatic or poetic override of phonological principles (Bergs 2005). Similarly to the fairly common claim that *don't* is a particle, it is contended here that *dontcha* is a budding marker of the prohibitive, which would make English typologically 'ordinary' (van der Auwera 2006). Contemporary internet data such as the title example suggest that it is indeed no longer readily analyzable for all speakers of English. Finally, the historical non-occurrence of *do you not*, the alternative to *don't you/dontcha* in questions, is explained functionally (à la Rissanen 1999).

References

Akmajian, A. (1984), Sentence types and the form-function fit. *Natural Language and Linguistic Theory* 2, 1-23.

Bergs, A. (2005), *Don't you kill the messenger!* The structure and development of the prohibitive. Manuscript. University of Düsseldorf.

Potsdam, E. (1998), *Syntactic Issues in the English Imperative*. New York: Garland Publishing.

Rissanen, M. (1999), *Isn't it?* or *is it not?* On the order of postverbal subject and negative particle in the history of English. In: I. Tieken-Boon van Ostade et al. (eds.), *Negation in the History of English*. Berlin: Mouton de Gruyter, 189-205.

Rupp, L. (1999), *Aspects of the Syntax of English Imperatives*. Unpublished Ph.D. dissertation. University of Essex.

van der Auwera, J. (2006), Why languages prefer prohibitives? *Journal of Foreign Languages* 1, 1-25.

van Rooy, Bertus & Lize Terblanche: A multi-dimensional/multi-functional analysis of a learner corpus

This paper reports on a multi-dimensional/multi-functional (MD/MF) analysis (Biber, 1988; Biber, Conrad & Reppen, 1998) of the Tswana Learner English Corpus (TLEC), together with the Louvain Corpus of Native English Essays (LOCNESS). In total, 383 argumentative essays from the TLE and 188 essays from LOCNESS are analysed. Using pair-wise t-tests, 53 of the 67 linguistic features analysed (Biber, 1988) are shown to differ significantly between the two corpora, affirming that these features capture significant differences between argumentative writing produced by native and non-native speaking students.

An MD/MF-analysis is conducted in an attempt to understand differences between learner and native student writing, examining the relationships between linguistic features Factor Analysis, similar to what Reynolds (2005) has done. Our objective is to discover dimensions that may tell differences between native and non-native student writing, rather than register differences.

A very strong first factor with up to 20 features can be extracted, which can be interpreted as features of **grammatical complexity** more characteristic of native student writing, e.g. preposition phrases, total nouns and nominalisations, attributive and predicative adjectives, and more complex verb phrases, such as past tense, perfect aspect and infinitives. This factor does not indicate a register difference, but a difference in grammatical competence and complexity, thus confirming widespread conclusions of second language writing research (e.g. Silva, 1993; Hinkel, 2002).

The second factor can be interpreted as **dense information presentation** (word length, type/token-ratio) versus a style that seems to be more **immediate** (personal pronouns, primary verbs 'do' and 'be' and present tense) and more concerned with **overt explanation** at the same time (modals, conditional and causal subordination). The denser style is more characteristic of LOCNESS, while the TLE is more colloquial in style, while lengthier attempts are made to explain things. This factor may be indicative of differences in rhetorical style, as suggested by De Klerk and Gough (2002).

References

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

Biber, D, Conrad, S & Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

De Klerk, V. & Gough, D. 2002. Black South African English. In Mesthrie, R. (ed.) *Language in South Africa*. Ed. Cambridge: Cambridge University Press, pp. 356-378.

Hinkel, E. 2002. *Second Language Writers' Text: Linguistic and Rhetorical Features*. Mahwah: Lawrence Erlbaum.

Reynolds, D.W. 2005. Linguistic correlates of second language literacy development: evidence from middle-grade learner essays. *Journal of Second Language Writing*, 14: 19-45.

Silva, T. 1993. Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27: 657-677.

Virtanen, Tuija: Discourse linguistics meets corpus linguistics: Theoretical and methodological issues in the troubled relationship

Discourse linguistics and corpus linguistics have an uneasy relationship because of their inherent ontological and epistemological differences. Yet, it is a steady relationship going well back in corpus-linguistic time, and one that both parties are highly motivated to keep up despite its many hazards and challenges. Singling out five complementary dimensions of discourse, understood here in a broad sense, this paper shows that not all of them will be equally accessible to users of corpus methods. Two fundamental aspects of discourse are identified as particularly challenging to corpus-linguistic enquiry, i.e. the distinction between a product and a process oriented approach, and the status of the primary notion of context. This raises the issue of authenticity, suggesting a need to rethink what we mean by the notion. The important methodological distinction between a corpus-based and a corpus-driven approach to discourse serves to highlight key issues in the joint history of discourse linguistics and corpus linguistics. The paper rounds off by a discussion of the added value to be gained by a combination of discourse linguistic and corpus linguistic approaches and methods: both parties can complement one another in constructive ways to uncover new aspects of discourse that will suggest reconsideration of our present understanding and disclose tacit assumptions.

1st

Wichmann, Anne & Nicole Dehe: The prosody of adverbials and discourse markers: the case of epistemic comment clauses

This study focuses on the function and prosodic realisation of a subset of short interpolations of the form pronoun (*I*)-verb, i.e., so-called comment clauses. They represent a recurring feature of conversational discourse and have been regarded as parentheticals at least since Jespersen (1937). They include the expressions *I believe, I assume,* and *I think*, which have been treated as epistemic adverbials (e.g., Aijmer 1997, Thompson & Mulac 1991), as well as combinations such as *I mean* and *you know,* which are treated as discourse markers (e.g., Murphy 1993). For some short parenthetical clauses it has been argued that their prosodic realisation is directly related to aspects of their interpretation (Reinhart 1983), but no phonetic evidence is given in this work

The comment clauses discussed here were retrieved from the spoken section of ICE GB. We analysed 311 utterances containing 314 comment clauses in sentence-medial or sentence final position (including: *I think*: 202, *I suppose*: 69) Of these, 153 sound files were clear enough to be analysed instrumentally to supplement the auditory analysis.

Our findings suggest a threefold classification. Firstly, we identify cases where the interpolation is, as predicted in prosodic theory, given its own intonational domain which may or may not be complemented by material from outside the comment clause (prosodic separation); secondly, there are cases where, despite the syntactic disjuncture, the clause is integrated melodically (if not temporally) into a larger intonation phrase (prosodic integration). Finally, we identify cases where the clause is melodically integrated within a hesitant planning phase. Our prosodic analysis supports the interpretation of Comment Clauses as epistemic adverbials but also suggests evidence of further grammaticalisation (or pragmaticalisation) towards the expression of interpersonal rather than epistemic meaning.

We will conclude with a conspectus of what is known about the prosody of adverbials and discourse markers, and attempt to show the systematic patterns that emerge from a wide variety of studies.

References

Aijmer, Karin. 1997. *I think* - an English modal particle. In Toril Swan & Olaf Jansen Westvik (eds.) *Modality in Germanic Languages - Historical and Comparative Perspective*.Berlin/ New York: Mouton de Gruyter. 1-47.

Jespersen, Otto. 1937. *Analytic Syntax*. London: Allen & Unwin.

Murphy, Lynne M. 1993. "Discourse markers and sentential syntax." In *Studies in the Linguistic Sciences* 23 (1): 163-167.

Reinhart, Tanya. 1983. "Point of view - The use of parentheticals." In *Essays on Deixis*, Gisa Rauh (ed.), Tübingen: Gunter Narr. 169-194.

Thompson, Sandra A. & Anthony Mulac 1991. "A quantitative perspective on the grammaticization of epistemic parentheticals in English." In Traugott, Elizabeth C. & Heine, Bernd, *Approaches to Grammaticalization, Vol. 2: Types of grammatical markers* (Typological Studies in Language 19: 2), pp. 313-327. Amsterdam/ Philadelphia: John Benjamins.

Willemse, Peter; Kristin Davidse & Lieselotte Brems: The development of extended type noun uses: a comparison between English *sort / kind* and French *sorte / espèce*

In this paper, we will compare the synchronic layering (Hopper 1991) of so-called 'type nouns' in English and French on the basis of corpus analysis. It has been established for English that, starting from their original lexical 'subcategorization' sense, particularly *sort* and *kind* have developed many other uses internal as well as external to NP-structure due to grammaticalization and subjectification (Aijmer 2002, Denison 2002). In this paper, we will investigate whether French *sorte* and *espèce* manifest similar layering of uses, and are comparable as to degree of grammaticalization. The quantified comparison will be based on exhaustive extractions on *sort* and *kind* from the Times subcorpus of the COBUILD corpus and on random samples of 400 tokens for *sorte* and *espèce* from *Frantexte*. (The spoken corpora COLT and *corpus de français parlé* were also consulted but did not yield enough data for sound comparison.)

The main uses that we will consider in this comparison are head, postdeterminer, modifier, emphasizer, qualifier and discourse particle, which were discussed for English type nouns in De Smedt, Brems, Davidse (in press). We will investigate whether all these uses are also found with *sorte* and *espèce* and in what relative proportions. Particular attention will go to distinct developments such as the pejorative use, e.g. *espèce de porc*. This is a highly grammaticalized use, typically construed without article, which may have developed from the qualifying use, e.g. *cette espèce de géant*, whereby expressions such as *un espèce de crétin*, in which the masculine article *un* preceding feminine *espèce* shows that the type noun is dominated by the second noun, also have to be considered. The main theoretical issue that we will seek to elucidate by this comparative inventory is what role is played by language-specific structural properties of the grammaticalizing environment. As a working hypothesis we will assume that the more extensive recursive premodifier structure of the English NP has enabled a greater diversity of grammaticalized uses resulting from the shift from head to modifier, i.e. type noun used as <u>head</u> + *of* + second noun → type noun+*of/a* used as <u>modifier</u>. Arguably, French has compensated for this by forming for instance discourse markers such as *en quelque sorte* (in contrast to *sort of/sorta*) by the conventionalisation of sequences with type nouns.

References

Aijmer, K. 2002. *English Discourse Particles: Evidence from a Corpus*. (Studies in Corpus Linguistics X.) Amsterdam: Benjamins.

Denison, D. 2002. History of the *Sort of* Construction Family. Paper presented at the Second International Conference on Construction Grammar (Helsinki, 6-8 September 2002).

L. De Smedt, L. Brems, K. Davidse (in press) NP-internal functions and extended uses of the 'type' nouns *kind*, *sort*, and *type*: towards a comprehensive, corpus-based description. In R. Facchinetti (ed), *Corpus linguistics 25 years on*. Amsterdam: Rodopi.

Hopper, P. 1991. On some principles of grammaticalization. In E. Traugott & B. Heine (eds) *Approaches to Grammaticalization*. Vol. I, 17-35. Amsterdam: Benjamins.