Belladelli, Anna: The spread of *be going to* in American English. Is the spoken medium leading the change?

The spread of the semi-modal be going to as an expression of futurity has been attested in many recent corpus-based studies. More specifically, the dramatic increase of occurrences in recent written American English has been brought to light and measured also by means of the comparison between the Brown Corpus and the Frown Corpus. This comparison confirmed a general trend towards 'colloquialization', as defined by Leech (2004).

The present analysis – carried out on the same corpora and focused particularly on text categories and syntactic patterns – reveals a less clear-cut change; indeed, although the number of occurrences has undoubtedly increased in the Frown Corpus (compared to thirty years earlier), more than two thirds of them are found in quotes from direct speech and in fictional dialogue. By way of example, in the text category which seems to have undergone the most striking increase (namely Press reportage), be going to and gonna appear in the spoken context in more than 85 per cent of the cases.

Rather than testifying to a widespread increase in American English as a whole, the data put forward in the present study would rather record a more restricted spread in the use of be going to in the spoken medium; indeed, the occurrence of this semi-modal within quotes in non-fictional genres and in fictional dialogue – i.e. within transcriptions of oral utterances – gives evidence for its frequency in speech.

References
Berglund, Ylva (2005), *Expressions of Future in Present-day English: A Corpus-based Approach.* Uppsala: Acta Universitatis Upsaliensis. Available from: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-5794.
Leech, Geoffrey (2003), 'Modality on the move: the English modal auxiliaries 1961-1992'. In: *Modality in Contemporary English*, eds. Roberta Facchinetti, Manfred Krug and Frank R. Palmer. Berlin and New York: Mouton de Gruyter. 223-240.
Leech, Geoffrey (2004), 'Recent grammatical change in English: data, description, theory', in *Advances in Corpus Linguistics*, eds. Karin Aijmer and Bengt Altenberg. Amsterdam and New York: Rodopi. 61-81.
Leech, Geoffrey and Nicholas Smith (2006), 'Recent grammatical change in written English 1961-1992: some preliminary findings of a comparison of American with British English'. In: *The Changing Face of Corpus Linguistics*, eds. Antoinette Renouf and Andrew Kehoe. Amsterdam and New York: Rodopi. 185-204.
Mair, Christian (1997a), 'The spread of the *going-to*-future in written English: a corpus-based investigation into language change in progress'. In: *Language History and Linguistic Modelling. A Fest-schrift for Jacek Fisiak on his 60th Birthday*, eds. Raymond Hickey and Stanislaw Puppel. Berlin and New York: Mouton de Gruyter. 1537-1543.
Mair, Christian (1997b), 'Parallel corpora: a real-time approach to the study of language change in progress'. In: *Corpus-based Studies in English: Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17)*, ed. Magnus Ljung. Amsterdam and Atlanta: Rodopi. 195-209.

Berglund Prytz, Ylva & Lou Burnard: BNC XML: what's it all about?

The British National Corpus was first released over ten years ago, and has since become an important resource for English language research. The corpus has now been released in a new version – BNC XML Edition. This poster offers an over-view of the changes made to the corpus, and demonstrates how the end user can benefit from the improvements, added information, and new format.

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written. The latest edition is the *BNC XML Edition*, released in March 2007. With a few exceptions, the texts in the new edition are the same as in previous versions. The main differences between BNC XML Edition and older editions are:

- errors - known errors and inconsistencies have been corrected
- lemma information - added to each word to allow searches for lemma collocations with lemmas
- simplified part-of-speech information added - allowing searches for 'all verbs', 'any noun' etc.
- multi-word units: in response to popular demand, all items inside multi-word units have been assigned part-of-speech tags as well
- format - the new version is in XML, which means it is easier to use with different tools and also makes viewing the texts easier.

The new XML format makes the corpus usable with many software tools, including simple web browsers. The corpus is delivered with a copy of the XAIRA program. XAIRA is an enhanced version of SARA, originally produced for use with the BNC. In addition to the features included in SARA, you can use Xaira with BNC XML Edition to:

- search by tag only. For example 'all –ing-forms of verbs', 'preposition + that', etc.
- search subcorpora defined by existing text categories (genre, written/spoken, year of publication, target audience, etc.)
- define searchable subcorpora according to your own categorization
- display search result as graphs
- quickly see distribution across text categories (existing or user-defined)
- retrieve collocations based on words, lemmas, or part-of-speech tags

BNC webpage: http://www.natcorp.ox.ac.uk/

Cava, Amelia Maria: Explicit evaluation in scientific research article abstracts

At the present time, academic writing is no longer considered as an objective and anonymous kind of discourse. It is largely based on rhetorical activity involving interactions between writers and readers. In this perspective academics do not just offer their own points of view, but seek to negotiate a credible account of themselves and their work by claiming solidarity with readers, evaluating ideas and acknowledging alternative views. Therefore, a convincing discourse becomes central to building the entire text even in this genre.

The linguistic resources academics used to do this have been variously described as *evaluation* (Hunston / Thompson 2000), *appraisal* (Martin, 2000; White, 2003), *stance* (Biber / Finegan, 1989; Hyland, 1999) and *interpersonal metadiscourse* (Crismore, 1989; Hyland, 2000).

My intention is to investigate the explicit evaluative lexis used by authors in their studies in two scientific journals from biology and mathematics. Broadly speaking, scientific subject matters are usually considered as the 'most' objective discipline but this does not imply that scientists do not comment on or evaluate their own research. On the contrary, in research article abstracts stating results and findings imply evaluate those results and findings, writers need to convince readers that their research is worth reading. The analysis is basically quantitative and qualitative setting out to highlight certain lexicogrammatical features specifically related to evaluative language and the attitudinal lexis.

The investigation of 1035 published research article abstracts (about 20,000 words) by means of the concordancing software WordsmithTools is focused on seeking positive and negative adjectives, adverbs, adjuncts and verbs, or specific repetition and emphasis, which leads readers to believe in what they are reading. The investigated corpus is drawn from five years (2000 - 2005) worth of scientific abstracts printed in the following two journals: the *Mathematics and Computer in Simulation* and the *International Journal of Primatology*.

References

Del Lungo camiciotti & Tognini Bonelli (eds) 2004. *Academic Discorse – New Insigths into Evaluation*. Bern: Peter Lang.

Hyland, Ken 1996. 'Writing Without Conviction? Hedging in Science Research Articles'. *Applied Linguistics* 17/4.

Hyland, Ken 2002. 'Directives: Argument and Engagement in academic writing'. *Applied Linguistics* 23/2: 215 -239.

Hyland, Ken & Tse Polly 2004. 'Metadiscourse in Academic writing: a Reappraisal. *Applied Linguistics* 25/2: 156 – 177.

Hoey, Michael 1983. *On the surface of discourse*. London: George Allen & UNWIN.

Hunston, Susan & Thompson Geoff 2001. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.

Hyland, Ken 1998. *Hedging in scientific research articles*. Amsterdam/Philadelphia: J. Benjamins.

Scott, Mike 2004. *WordsmithTools 4.0*.

Sinclair, John 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Swales, John 1990. *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Thetela, Puleng 1997. 'Evaluated Entities and Parameters of Value in Academic Research Articles'. *English for Specific Purposes* 16/2: 101 -118.

Thompson Geoff & Ye Yiyun 1991. 'Evaluation in the Reporting Verbs Used in Academic Papers'. *Applied Linguistics* 12/ 4.

Thompson Geoff 1996. 'Voices in the Text: Discourse Perspectives on Language Reports'. *Applied Linguistics,* 17/4.

Thompson, Geoff 2001. 'Interaction in Academic writing: Learning to argue with the reader'. *Applied Linguistics,* 22/1: 51 - 78.

Chocano, Gema; Rocío Jiménez, Cristóbal Lozano, Amaya Mendikoetxea, Susana Murcia, Michael O'Donnell, Paul Rollinson & Iván I. Teomiro García: Exploring word order in learner corpora: The WOSLAC Project

This poster reports on work in progress under the framework of a research project investigating word order in Second Language Acquisition (WOSLAC), based on two written learner corpora: *Wri*CLE (L1 Spanish - L2 English) and CEDEL2 (L1 English - L2 Spanish). We will present (i) the motivation and objectives of the project, (ii) data collection, (iii) query software and (iv) data analysis.

The purpose of this three-year project is to determine the properties which constrain word order in the interlanguages of L2 learners of English (with L1 Spanish) and L2 learners of Spanish (with L1 English). Word order in English and Spanish differs significantly; in English word order is often said to be 'fixed', by lexico-syntactic properties, while Spanish allows for what is often referred to as 'free order', determined by syntactic-discoursive properties. An in-depth investigation into word order in advanced learners of L2 English and L2 Spanish will thus offer answers to questions regarding the relative difficulty of acquiring lexical-syntactic and syntactic-discoursive properties, as well as throwing some light on processes involving L1 transfer and the occurrence of constructions which cannot be attributed to the L1, nor to the target language.

Learner corpora are an invaluable tool to explore these issues. Data collection started over a year ago and our target is for *Wri*CLE and CEDEL2 to reach 1 million words by the end of the project. The corpora will be annotated using *UAM CorpusTool* (by Michael O'Donnell) which has been adapted for this study. We focus on structures presenting non canonical word order (left and right dislocation), as well as constructions involving special ordering of constituents: passives, clefts, extraposition and so on. The tool allows an analyst to select a text from the corpus, and annotate it in various ways. The analyst can highlight a segment (e.g., an *it*-cleft) and then assign features to that segment. The tool produces an XML-encoded version of the text file, including the features assigned to the segments. Because hand-annotation is slow, the tool will allow the analyst to associate lexico-syntactic patterns with each feature, allowing the tool to automatically detect instances of the pattern.

http://www.uam.es/woslac

Guo, Xiaotian: A two dimensional diagnose into learners' needs in using verbs

This paper reports on my research in Chinese learners written English in the area of verbs (Guo 2006). Contrastive interlanguage analysis (Granger 2002), was used in this two-dimensional perspective exploration. The two corpora compared are COLEC (the learner corpus) and LOCNESS (the NS corpus). The first dimension refers to the total number of verbs used by the learners and the NSs and the second dimension refers to the scope of use of individual verbs. In order to discover the first dimension, i.e. how many verbs are used by the learners and by the NSs, POS tagging and a special programme were jointly used. It was discovered that the learners use a much smaller amount of verbs (lemmas) than the NSs (569 vs. 890). It is commonly held that the more vocabulary learners have the more proficient their English should be. It follows that it seems more important to increase the number of learners' mental lexicons than to improve the existing vocabulary that is already in use by learners. In the second dimension, however, a detailed look at the collocates of TAKE shows that learners use a much narrower scope of collocates (both continuous like 'TAKE a view' and discontinuous 'TAKE … seriously'). Take the collocation 'TAKE a view' for example, whereas there are only two varieties in the learner English i.e. 'TAKE an attitude' and 'TAKE an idea', there are much more varieties of collocates in the NS English, i.e. 'TAKE a view', 'TAKE a stance', 'TAKE a side' , 'TAKE a position' and 'TAKE a stand'. I believe that the two-dimensional perspective into learner English can be used widely to diagnose learners' needs in using verbs. The multiple implications for language pedagogy are discussed including teaching material development, CALL software development, data-driven learning in the classroom and learner English dictionary compilation.

References
Granger S. (2002) A Bird's-eye View of Computer Learner Corpus Research. In Granger S., Hung J. and Petch-Tyson S. (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Language Learning and Language Teaching 6. Amsterdam & Philadelphia: Benjamins (245 pp.), pp. 3-33.
Guo X. (2006) *Verbs in the Written English of Chinese Learners: A Corpus-based omparison between Non-native Speakers and Native Speakers*. Unpublished PhD dissertation. The University of Birmingham.

Leedham, Maria: Corpora compared: lexical chunks in Chinese and British students' writing

There are now more than 80,000 Chinese students studying within Higher Education in the UK (Shen, 2006). While their written language often attains a high standard, there are observable differences from native speakers' language at the level of lexical chunks, defined by Wray (2002:9) as "a sequence… which is, or appears to be, prefabricated." In this poster presentation I describe my PhD work in progress on the differences in the use of lexical chunks between Chinese and British students in UK HE.

This study draws on the British Academic Written English corpus which comprises 3000 assignments at IIi level and above from undergraduate and Master's level students in a variety of disciplines in three UK universities. I have compiled two subcorpora of 200 assignments written by L1 (first language) Mandarin students and 200 by L1 English students from within the BAWE corpus. The two corpora are matched for year of study, age and discipline of study. Preliminary research using WordSmith Tools suggests that the L1 Mandarin students' writing contains a substantial number of lexical chunks that are translations from Mandarin chunks.

References

British Academic Written English project.
http://www2.warwick.ac.uk/fac/soc/celte/research/bawe/overview/

Leedham, M.E. 2006 'Do I Speak Better? : A longitudinal study of lexical chunking in the spoken language of two Japanese students'. In The East Asian Learner, May 2006, Available at: http://www.brookes.ac.uk/schools/education/eal/eal-2-2/vol2_2_leedham.pdf

Shen, W. 2006 Memorandum to Select Committee on Home Affairs. Downloaded from http://www.publications.parliament.uk/pa/cm200506/cmselect/cmhaff/775/775awe31.htm on 8.11.06

Wordsmith Tools version 4.0, developed by Mike Scott and OUP, downloaded from www.lexically.net//wordsmith/purchasing.htm

Wray, A. 2002 *Formulaic Language and the Lexicon*, CUP

Lindmark, Kerstin: Influence of English on Swedish in beginning and experienced translators

To be able to study phenomena in translated texts that make them seem different from original texts in the same language, I have compiled a corpus of translations made by translation students during their first two semesters of the translator training programme (Lindmark 2006). One of the features that cause problems for neophyte translators is prepositions. To broaden the scope of the investigation and to get an idea about the importance of experience, translations of made by the students are compared with translations made by more experienced translators (cf. Englund in press), using the English-Swedish Parallel Corpus. A pilot study of a number of phrases containing prepositions, for example *protection for*, *level(s) of*, and *movement of* shows that the experienced translators tend to choose solutions containing compounds for the *of* phrases to a larger degree than the student translators, who tend to use a preposition. Regarding the *for* phrases, the experienced translators tend use the nearest Swedish equivalent preposition and use *för*, whereas the students use other prepositions to a larger extent. In the present poster, these and further results are exposed and quantified, and it is investigated whether there is any correlation between how proficient the translation students are at English (DiaLang test) and how they translate phrases containing prepositions. Further, the poster tests the hypothesis that they will stay (too) close to the English structure initially, then unjustifiably turn away from it, and finally reach a level of skill where they can determine when it is best to follow the English structure or to deviate from it, respectively.

References

DiaLang test: http://www.dialang.org

Englund-Dimitrova, B. In press. Segmentation of the writing process in translation: experts vs. novices; to be published in: Sullivan, K.P.H. and Lindgren, E. Computer key-stroke logging: Methods and Applications. Amsterdam: Elsevier.

English-Swedish Parallel Corpus: http://www.englund.lu.se/research/corpus/corpus/espc.html Visited 28 November 2006

Lindmark, K. Source language influence on target language – a corpus for studying variation and development, submitted for publication.

Pedrola, Monica: European Parliamentary debates. A corpus-based analysis of hedging strategies

The present paper presents the work currently being carried out within the Contact project at the Catholic University of Milan, Italy, as part of the national eColingua project which is investigating the use of parallel and comparable corpora of English and Italian texts in developing research and teaching methodologies in foreign language acquisition and translator training. The main object of investigation is mediated English, i.e. language which has undergone some kind of elaboration. The working hypothesis behind the research is that translator-mediated language is characterised by such universal features as *explicitation, simplification, normalisation and levelling out* (Baker, 1995, 1996; Laviosa, 1998). The new element in this line of research lies in verifying whether and to what extent the above hypothesis may be influenced by the changing nature of English as a lingua franca (Mauranen, 2003; Seidlholfer, 2005).

In order to test this hypothesis monolingual and bilingual corpora are being compiled. They are composed of original English texts, original Italian texts and their English translations, translated English texts from other European languages.

The present paper focuses in particular on European Parliamentary debates, which are available online in their original form and in a variety of mediated forms. As a result of the process of mediation through translation that they undergo, these texts are instances of what Schäffner and Adab (1997) define as *hybrid texts*. The analysis wishes to compare the frequency and variety of use of selected *hedging strategies* (Markkanen & Schröder, 1997; Hyland, 1998) in the various corpora, in order to explore their role as carriers of hybridisation phenomena and to investigate the effect of these hybridisation processes on native English.

References

Baker, M. 1995. "Corpora in Translation Studies: an overview and some suggestions for future research", in Target 7/2, pp. 223-243.

Baker, M. 1996. "Corpus-based translation studies: the challenges that lie ahead", in Somers, H. (ed) *Terminology LSP and Translation*, John Benjamins, Amsterdam/Philadelphia, pp. 176-186.

Hyland, K. 1998. *Hedging in scientific research articles*, John Benjamins, Amsterdam/Philadelphia.

Laviosa, S. 1998. "The English comparable corpus: a resource and a methodology", in Bowker, L. et al. (eds) Unity in Diversity: current trends in translation studies, St. Jerome, Manchester, pp. 101-112.

Markkanen, R. and Schröder, H. (1997) (eds) *Hedging and discourse : approaches to the analysis of a pragmatic phenomenon in academic texts,* W. de Gruyter, Berlin/New York.

Mauranen, A. 2003. "The Corpus of English as Lingua Franca in Academic Settings", TESOL Quarterly, 37: 3, pp. 513-527.

Schäffner, C. e Adab, B. 1995. "Translation as intercultural communication – contact as conflict, in translation as intercultural communication", selected papers from EST Congress (Prague 1995), M. Snell-Hornby, Z. Jettmarovà and K. Kaindl (eds) John Benjamins, Amsterdam/Philadelphia, pp. 325-337.

Seidlhofer, B. 2005. "English as a lingua franca", *ELT Journal,* Oxford: Oct 2005. Vol. 59, Iss. 4; p. 339-341

Säily, Tanja & Jukka Suomela: Comparing type counts: The case of women, men and *-ity* in early English letters

No matter how carefully a corpus is designed, the data will always be sparse and unevenly distributed with respect to some categories of interest. While some measurements obtained from subcorpora of varying sizes can be directly normalised by, e.g., word count, there are many measurements that do not grow linearly with the amount of data.

We study the productivity of the noun-forming suffixes *-ness* and *-ity* in 17th-century English letters, using material from the *Corpus of Early English Correspondence*. Combining morphology and historical sociolinguistics, we wish to compare the numbers of different types of *-ness* and *-ity* used by different sociolinguistic groups. For example, we believe that *-ity*, a 'learned' and etymologically foreign suffix, is in this material less productive with women than with men, as 17th-century women received far less education than men.

Unfortunately, only about a quarter of the 17th-century material in the corpus was written by women. This makes comparison difficult: while the number of types does grow with the number of tokens, the growth is generally not linear (Herdan 1966: 76), so we cannot directly normalise the type counts. To overcome this problem, we apply techniques from permutation testing (cf. Good 2005).

We formulate a hypothesis that gender is significant in the case of *-ity*. To test such hypotheses, we use a computer program that chooses a large number of random permutations of parts of the corpus and computes type accumulation curves with confidence intervals for these permutations. By these curves, we can determine, e.g., whether the subcorpus that consists of women's texts contains fewer types of *-ity* than almost all randomly chosen subcorpora of approximately the same size; this turned out to be the case.

References

The Corpus of Early English Correspondence (2006) Compiled by the *Sociolinguistics and Language History* project team (Nevalainen, T., H. Raumolin-Brunberg, J. Keränen, M. Nevala, A. Nurmi, M. Palander-Collin) at the Department of English, University of Helsinki. A parsed version of the corpus will be available through the Oxford Text Archive and ICAME.

Good, P. (2005) *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. 3rd edition. Springer Series in Statistics. Berlin: Springer-Verlag.

Herdan, G. (1966) *The Advanced Theory of Language as Choice and Chance*. Berlin: Springer-Verlag.

Sift, Patricia: Directive speech acts in instructional texts, 1350-1750

This study traces the realisation of directive speech acts in text types of religious and non-religious instruction. The point of departure is a corpus-based study of directives in Late Middle English and Early Modern English sermons (1350-1750). In these sermons, the directives addressed by the preachers to their audiences are realised in a number of ways, which can be grouped according to syntactic patterns. These are, for instance, performatives (*I admonish you to abstain*), imperatives (*Abstain*), hortatives (*Abstain we* or *Let us abstain*), or modal constructions (*You must abstain*, *We should abstain*, *A true Christian ought to abstain*). Within these structures, the speech-act verbs are of importance (*I admonish you* vs. *I suggest to you*, which is much weaker) as well as the verbs designating the action expected by the audience (physical vs. cognitive: *Kneel down and pray* vs. *Listen and consider*). What is more, the physical and cognitive 'actions' required by the audience can be immediate (to be performed at the time of utterance, as in the previous examples) or they can refer to future, principle-based actions (*Come to Church every Sunday*, *Forever remember your duties*).
For the period under investigation, the following can be observed:

− within Late Middle English and Early Modern English (and their sub-periods), the occurrence of directives with varying syntactic structures, propositional content and illocutionary force seems to depend largely on the social class of the audience addressed
− across time, these patterns vary, with an increasing tendency towards first-person structures (signalling solidarity with the audience and positive face-work) and verbs of cognition requiring immediate action (guiding the audience through the sermon text)

A next step is to expand the study of directives to other instructional text types. On the basis of the *Helsinki Corpus* as well as complementary data, we will see whether the above findings are unique to sermons or whether parallels can be found in the realisation of directives in other text types of both religious and worldly instruction.

Smith, Adam; Deanna Wong; Pam Peters & Canzhong Wu: Trends in electronic document structure

This poster presents the comprehensive results of a research project into the structure and segmentation of online documents, introduced at ICAME 2001. The study is based on EDOC, a 200,000 word corpus of electronic documents from the year 2000/1, in the contrasting genres of information delivery and instruction. The project involves innovative use of markup to annotate visual and layout features of electronic texts (etexts), and the subsequent development of software to quantify these features.

The ICE-based markup applied to written texts was enlarged with an extended tag set, to capture more aspects of layout such as lists and different heading levels. The EDOC markup was converted to XML and validated to support software that allows text features such as paragraph length, number of lists and average length of list item to be calculated.

Software especially developed for the project allowed us to compare the results for the whole EDOC corpus (100 texts) with those for the 8 pilot study texts analysed in Peters and Smith (2003), and to revisit the earlier hypotheses in relation to the greatly enlarged set of informational and instructional etexts. Much stronger evidence was found of the increased segmentation of texts mounted on the web, with the average size of textual units clearly lower than those of the pilot study, in both generic categories. For example, average sentence length came down from 19.4 to 17.9 words for instructional e-texts, and there were only 3.6 sentences per paragraph in informational texts over the whole corpus as compared to 4.6 in the pilot study. Average length of list items reduced in both categories, from 21.6 to 12.7 words for informational and from 14.6 to 13.9 for instructional.

While much has been written by Nielsen (2000) and others, on the structural demands and possibilities of web-based documents, less attention has been paid to what is happening to written units such as the sentence and the paragraph which are further evolving from patterns established in the print medium. The compilation of EDOC at an early stage in the development of the internet has given us a benchmark which will support comparative studies on how structural and segmental features of etexts are adapting to the medium.

References

Bray, T., J. Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau (eds) (2006) *Extensible Markup Language (XML) 1.0* (Fourth Edition). http://www.w3.org/TR/2006/REC-xml-20060816/

Haussamen, B. (1994) "The future of the English sentence", *Visible Language*, 28(1):4-25

Nielsen, J (2000) *Designing Web Usability*. Indianapolis: New Riders

Peters, P. and A. Smith (2003) "Textual structure and segmentation in online documents" in S.Granger and S. Petch-Tyson (eds) *Extending the scope of corpus-based research.* Amsterdam, New York: Rodopi. 71-85

Spano, Marcella: Introducing MedInTCorpus: research perspectives

The MedIntCorpus is a corpus for academic research purposes currently developed at the Catholic University of the Sacred Heart of Milan as part of the National eColingua Project. The research aims to illustrate from a synchronic and diachronic perspective the reciprocal influence between English and other languages on a lexico-grammar, textual and pragmatic level. The present paper focuses on the preliminary considerations and decisions underlying the planning and compilation of the MedInTCorpus, Medical International Corpus or Medicine in Translation Corpus, which can be described as a comparable synchronic corpus of English made up of four sections comprising medical articles, namely clinical investigations, case reports and studies, published by American, European, British and Italian specialized journals. The research based on this corpus will also include a translational section of Italian – English parallel texts and a comparable diachronic corpus.

      The first part of the work is dedicated to the mapping of the structure of the corpus and the description of the criteria involved in the selection of the texts to include in the corpus. The second part concerns the general frame of research in international English and the way in which the MedInT corpus is embedded in it. The last part is dedicated to the description of the hypothesis I intend to validate with my research. I argue that features such as explicitation, normalization, levelling out and simplification, considered as typical traits of translation are, more specifically, characteristics of non-native language production.

References

Baker, M., (1996), Corpus-based Translation Studies: The Challenges that Lie Ahead, in Harold Somers (ed) Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager, Amsterdam & Philadelphia: John Benjamins

Baker, M., (1995), Corpora in Translation Studies: An Overview and Some Suggestions for Future Research, Target 7(2), pp. 223-243

Baker, M., Corpus Linguistics and Translation Studies, Implication and Applications, in Baker et als (eds), (1993), Text and Technology: In Honour of John Sinclair, Amsterdam & Philadelphia: John Benjamins

Jenkins, J., (2006), Current perspectives on teaching World Englishes and English as a Lingua Franca, TESOL Quarterly 40(1), pp.157-181

Jenkins, J., (2005), Respecting diversity, promoting intelligibility: The challenge for English as an International Language, Perspectives, Journal of TESOL-Italy XXII(1), pp.11-21

Laviosa, S., (2002), Corpus-based translation studies : theory, findings, applications, Amsterdam ; New York, NY : Rodopi

Laviosa, S., (1998), The English Comparable Corpus: A Resource and a Methodology, in L. Bowker; M. Cronin; D. Kenny; J. Pearson (eds.), Unity in Diversity? Current Trends in Translation Studies. Manchester: St. Jerome

Laviosa, S., (1997), How Comparable Can 'Comparable Corpora' Be?, Target 9(2), pp. 289-319

Mauranen, A., (2005), English as Lingua Franca: An Unknown Language?, In G. Cortese and A. Duszak (eds), Identity, Community, Discourse, Bern: Peter Lang

Seidlhofer, B., (2005), English as a lingua franca, ELT Journal, 59(2), 339-341

Seidlhofer, B., (2000), Mind the gap: English as a mother tongue vs English as a lingua franca, Views, 9(1), pp. 51-68

Taavitsainen, Irma; Päivi Pahta, Turo Hiltunen, Ville Marttila, Maura Ratia & Jukka Tyrkkö: The Corpus of Early Modern English Medical Texts (1500-1700)

The poster reports on the diachronic three-part Corpus of Early English Medical Writing (1375-1800), compiled by the Scientific Thought-styles project team in Helsinki. Our presentation focuses on the second part, the corpus of Early Modern English Medical Texts (EMEMT), which is approaching completion.[1] This electronic tool will be useful for a wide variety of studies on linguistic developments. It also provides material for more accurate assessments of the special language of medicine, and eventually, for a genre-based language history. The selection of material in EMEMT takes the challenge of representing the reality of medical writing in 1500-1700 as faithfully as possible.

The development of the medical register with widening genres from the Late Medieval to the Late Modern period is reflected in the selection of texts and the categorization of the corpus. The poster illustrates how the sociohistorical background changes, how new genres come into being and old ones continue or become transformed and adapted to new uses. The first scientific journals are founded and postal services develop, offering new avenues for the exchange and sharing of information. Forms of communication change with the new media, reflecting the widening discourse communities. The category of learned writing shows innovation developing new genres, while the tradition of remedybooks and *materia medica* continue much in the old vein relying on old patterns of thought. The split into academic and "popular" becomes more pronounced.

[1] The first part, *Middle English Medical Texts*, came out in December 2005. The third, *Late Modern English Medical Texts* (1700-1800) is under work as well.

References

Middle English Medical Texts. 2005. Compiled by Irma Taavitsainen, Päivi Pahta & Martti Mäkinen with software by Raymond Hickey. Amsterdam and Philadelphia: John Benjamins.

Taavitsainen, Irma & Päivi Pahta. 1997. "The Corpus of Early English Medical Writing". *ICAME Journal* 21.71–78.

Wong, Deanna: Backchannel types and prosody in Australian English telephone dialogues

Backchannels (Yngve, 1970) function to support and enhance the verbal exchange Their forms span a continuum from the minimal *mm* to complex grammatical constructions (for examples, see Holmes & Stubbe, 1997; Tottie, 1991). There is also evidence that suggests the intonational structure of some backchannel forms determines both their meaning and function in the on-going conversation (Gardner, 2001). The range of forms and their varying prosodic features means that backchannels prove to be something of an enigma, sitting outside of mainstream linguistic theory. To further complicate matters, the dynamic nature of conversation often results in backchannels occurring simultaneously with the speaker's talk, yet they rarely interrupt it.

Whilst interest in backchannels in conversation has been constant since the mid-twentieth century (for an early discussion see Fries, 1952), the literature is weakened by the reliance on axiomatic definitions that assume a small repertoire, and a corresponding failure to establish this assumption empirically. Further, there have been no attempts to examine backchannels as coherent sequences of spoken utterances that span a speaker's turn.

The following presentation describes research designed to redress this lack, using corpus mark-up to objectively identify backchannels and establish the relationship between their structure and their prosodic features. Rather than taking single instances out of context, it examines backchannels as part of a sequence of non-interruptive utterances that occur within a speaker's turn. The data used in this investigation were taken from annotated orthographic transcriptions and digital recordings of telephone dialogues from the Australian component of the International Corpus of English (ICE). In order to produce a visual representation of the often complex interaction between the types of backchannels and prosodic systems, the data will be annotated using the ELAN annotation tool (Max-Planck-Institute for Psycholinguistics, 2002). The overall aim of the research is to establish the underlying principles that govern listening in telephone interaction.

References
Fries, C. C. (1952). *The structure of English : an introduction to the construction of English sentences*. New York: Harcourt Brace.

Gardner, R. (2001). *When listeners talk: response tokens and listener stance.* Amsterdam: John Benjamins Publishing Co.

Holmes, J., & Stubbe, M. (1997). Good listeners: Gender differences in New Zealand conversation. *Women and Language, 20*(2), 7-14.

Max-Planck-Institute for Psycholinguistics. (2002). EUDICO Linguistic Annotator (ELAN) (Version 2.6.3). Nijmegen, The Netherlands: Max-Planck-Institute for Psycholinguistics.

Tottie, G. (1991). Conversation style in British and American English: The case of backchannels. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 254-271). London: Longman.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting Chicago Linguistics Society* (pp. 567-577). Chicago: Chicago Linguistics Society.