Shorter notices

Corpus of 19th-Century Scottish Correspondence
(in preparation)

Marina Dossena (Associate Professor of English at the University of Bergamo, Italy) has recently begun the compilation of a Corpus of Nineteenth-Century Scottish Correspondence (19CSC). The corpus will include both private and non-private letters, written by both men and women of varying ages. Special attention will be given to business correspondence and to letters written by emigrants.

Contact address: marina.dossena@unibg.it

XXIII International Conference on English Language Research on Computerized Corpora of Modern and Medieval English (ICAME 23)

Arken Conference Centre, near Gothenburg, 22–26 May, 2002

Roberta Facchinetti
Università di Verona

Despite there being no previous agreement, Munich airport turned out to be the meeting point for a great number of delegates flocking to Gothenburg for the 23rd annual ICAME conference. The boarding gate was our environment for a pre-tasting of the conference, each one of us enquiring about the state of the art of others' researches and projects. The exchange of views and discussions continued on the plane and even in the taxis leading us to the conference venue, *Arken Conference Centre*, a luxurious hotel situated 20 minutes from Gothenburg right in front of the *archipelago*. There we were, "far from the madding

crowd" of the bustling city and … closer to Volvo, Sweden's most famous car company.

On our arrival, our excellent hostess Karin Aijmer and her super-efficient team were scurrying about the hotel, welcoming the participants, re-arranging the timetable, and allotting rooms to the convenors. Many of us strolled along the narrow path outside the hotel, down to the pier, where several boats were docked; we were all waiting for the starting whistle by Karin, ready to sail in compact formation towards this year's goal: "the theory and use of English language corpora": three, two, one ... off we went with the welcome party, for the 23rd ICAME Conference.

During the conference, six plenary speeches were delivered respectively by John Sinclair, Michael Hoey, Michael Halliday, John DuBois, Helge Dyvik, and Geoffrey Leech. Each of them complemented the others in a most fruitful way. **John Sinclair** qualified the scope of corpus-driven linguistics with reference to the complex relationship existing between scholars, theories, and data. Indeed, since a corpus confers greater status on the patterns of data found within it, a new picture of language is bound to emerge, thus forcing scholars and theories to relate to this empirical information in a completely different way. **Michael Hoey** focused on the role of corpus linguistics by shedding light on a few textual matters, with specific reference to thematization, textual interpretation, paragraphing, text domain and genre, and colligational prosody. On his turn, **Michael Halliday** started form the tenet that the foundations of grammar – and of grammaticalization – lie in the spoken language; then he recalled the many issues faced when compiling a spoken corpus in the attempt of (re)presenting and interpreting spoken data that will not impose on speech the design features which obscure its essential semogenic powers. Still with reference to speech, **John DuBois** highlighted the importance of understanding what natural discourse is for anybody embarking on creating and exploiting spoken language databases, while **Helge Dyvik** moved to the role of parallel corpora for the retrieval of semantic properties of expressions; he exemplified his remarks with instances from the project "From Parallel Corpus to Wordnet", which is being carried out at the University of Bergen. Finally, **Geoffrey Leech** provided corpus-based evidence to some grammatical changes in British and American English, including the present progressive tense, semi-modals, verb and negative contractions, adjectives, nouns, genitives, acronyms, and preposition stranding in relative clauses.

These and many other topics were dealt with in a total of about 80 presentations, distributed among full papers, works-in-progress, and posters; special attention was given to the following:

1. compiling and using new corpora;
2. bilingual and multilingual corpora;
3. language acquisition and teaching;
4. discourse and text;
5. diachronic corpora;
6. corpus data, text type, and grammar.

The sessions pertaining to the COMPILATION AND USE OF NEW CORPORA and to BILINGUAL AND MULTILINGUAL CORPORA were by far the most exploited fields of discussion. The former focused on a wide variety of specific themes, from the extraction of linguistic information from the web by means of Webcorp – both for research (Antoinette Renouf, Andrew Kehoe & David Mezquiriz) and for teaching purposes (Natalie Kübler) – to the analysis of web personal ads in South-East Asia (Peter Tan, Vincent B.Y. Ooi, Andy K. L. Chiang), the identification of word functions for automated indexing of a corpus (David Roberts & Clive Souter), the evaluation of a pragmatic model of speech act assignment in spoken language corpora (Amanda Schiffrin & Clive Souter), the development of spoken vocabulary in a corpus of children's English (Clive Souter), and the illustration of the resources offered by the Oxford Text Archive (Ylva Berglund, & Martin Wynne). As for corpus-building, Vivian de Klerk illustrated the state of the art in developing a corpus of spoken Xhosa English, while Fiona Douglas focused on SCOTS, the Scottish Corpus of Texts and Speech project; finally, John Kirk and Jeffrey Kallen suggested a corpus-based model for the standardization of National Varieties of English.

In turn, the BILINGUAL AND MULTILINGUAL CORPORA at the conference dealt with a number of languages contrasting with English: Swedish (Bengt Altenberg, Anna-Lena Fredriksson, & Åke Viberg), Chinese (Tony Mc Enery & Zhongua Xiao), Spanish (Anna-Brita Stenström & Annette Myre Jörgensen), Brasilian (Stella Tagnin), German (Josef Schmied), and Norwegian (Stig Johansson & Janne Engen, Lene Nordrum). In this context, theoretical discussion thrived on the use of corpus evidence in translation and contrastive description (Elena Tognini Bonelli), on the study of translation-specific lexicogrammar (Per-Ola Nilsson), and on the development of specialized multilingual thematic dictionaries (the present writer).

The sessions pertaining to LANGUAGE ACQUISITION AND TEACHING were largely covered by the highly professional Swedish group who focused on Swedish learners of English: Margareta Westergren Axelsson on the Uppsala Student English corpus, Monika Hägglund on English verb-particle constructions, Andreas Eriksson on tense and aspect, and Mia Boström Aronsson on

extraposition. Other scholars focused on a plethora of subject-specific studies: new applications of child language corpora (Norbert Schlüter), identification of missing and redundant words in an EFL corpus (Sylviane Granger), assessment of learner acquisition in EFL teaching based on new information and communication technologies (Fanny Meunier), illustration of a model for quantitative EFL learner corpus research (Przemysaw Kaszubski), classroom interaction (He Anping), "putting verbs" (Caroline David), metadiscourse (Annelie Ädel), demonstrative reference as a cohesive device (Rumiana Blagoeva), and finally movie narratives in the Giessen-Long Beach Chaplin Corpus of semi-controlled native speaker, ESL and EFL speaker interactions (Anja Janoschka, Andreas H. Jucker, Simone Müller, & Sara Smith).

With reference to Discourse and Text, papers ranged from text typology (Kaoru Takahashi) to modality (Angela Downing, Maria Teresa Prat Zagrebelsky), hyperbolic forms (Claudia Claridge), and syntactic issues like adverbials in *it*-cleft constructions (Hilde Hasselgård), the grammar of "or" (Mira Ariel), indefiniteness (Mikko Laitinen), word-order variation (Rolf Kreyer) and the sociopragmatics of *let's* (Bernard de Clerck). More register-related researches focused on political discourse (Jonathan Charteris-Black), the teaching of apology strategies (Tomoko Kaneko), patterns of repetition in different registers (Natalia Gvishiani), professional oral reading (Jacqueline Monschau), and variation in newspaper corpora idioms (David Minugh).

The sessions pertaining to Diachronic corpora were largely concentrated on pragmatics, with a historical study of English directives (Thomas Kohnen), of the "ers", "ums" and "ahs" in Early Modern English dialogues (Jonathan Culpeper & Merja Kytö), and of face-works in Early English sermons (Patricia Sift). Discussion was also on the use of relativizers in the 19th century speech-related and written genres (Christine Johansson), on the tagging and parsing of the Corpus of Early English Correspondence (Arja Nurmi & Ann Taylor), and finally on the illustration of the corpus of Early English Medical Writing 1375–1750 (Irma Taavitsainen & Päivi Pahta).

Finally, with reference to corpus data, text type, and grammar, a good number of papers dealt with nouns and nominalization, with specific reference to measure nouns (Lieselotte Brems), nominal *that*-clauses (Ilka Mindt), the "support" function of general nouns (Michaela Mahlberg), the lexical and discourse environments of fact-nominalizations (Kristin Davidse), and esophoric reference (Peter Willemse). Others touched on a wide range of grammatical and syntactic issues: the structure of the BNC World Edition for the study of manner adverbs (Junsaku Nakamura & Tomoiji Tabata), cognitive grammar (Joybrato Mukherjee), the reflexive *yourself* (Göran Kjellmer), the mapping of the English

possessive (Gregory Garretson), the use of pronouns in nonassertive contexts (Nelleke Oostdijk), terms of comparison (Shunji Yamazaki), relativizers after *same* in Present-day English (Gunnel Tottie & Hans Martin Lehmann), agreement phenomena in the Longman American Spoken Corpus and the British National Corpus (Hans Martin Lehmann), and finally subjective evaluation and perception verb complementation (Aurelia Usoniene).

SOFTWARE DEMONSTRATIONS focused on the ICLE cd-rom (Sylviane Granger & Fanny Meunier), on BNCweb 2.0, the webface interface to the BNC World Edition developed by the Zurich team (Sebastian Hoffman, Hans Martin Lehmann & Peter Schneider), on the COLT cd-rom (Knut Hofland), and finally on the English-Norwegian Parallel Corpus and the Oslo Multilingual Corpus (Stig Johansson).

This highly stimulating scientific programme was further enriched by the equally intriguing social events the conference was interspersed with, including a lunch at *Gothia Towers* and the luxurious conference dinner at *Nya Elfsborgs Fästning*, a fortress almost as old as the city of Gothenburg that it defends. Excellently organized, like everything else after all, was the guided tour around Gothenburg; on a cosy bus sheltering us from a rainy day, we learnt that Gothenburg is Sweden's second largest city after the capital *Stockholm*; it has a canal system that rings the town centre, while on terra firma trams whiz through the streets. We saw the opera house and the concert hall; we passed by *Nordstan*, the biggest shopping centre in Sweden, admired the picturesque row of little 18th century houses around a cobbled courtyard at *Kronhusbodarna*, and headed for Gothenburg's famous *Kungsportsavenyn* known as *Avenyn* (the Avenue), with a vigorous street life boasting restaurants, trendy shops, and brasseries.

Yet what intrigued us most was our excursion to the *Terra Nova* shipyard where *Ostindiefararen* is being built. This is a full-size replica of *Göteborg*, an 18th century vessel owned by the Swedish East India Company. The boat is being brought into line with modern technology with the aim of sailing her to China and from there to everywhere in the world. The frame is now in place and the hull has taken its full form; with its 56,000 hand-wrought nails, 4,000 cubic metres of wood, and 50 tons of iron … this ship is indeed a *corpus* in all intents and purposes and might easily epitomize our 23rd ICAME conference. Indeed, our group of over one hundred people landed at the Arken (*the ark*, a ship, again) Conference Centre and in our diachronic and synchronic studies, thus uniting past and present, we shared the daunting, fascinating task of building and exploiting corpora of authentic speech, which can then be boarded like vessels to explore further or anew the manifold lands of English and other languages.