Workshop (II)
Room: Coburg

ICAME 2010

Workshop: ***News, (new media), and corpora: from methodology to theory***

Convenor: Roberta Facchinetti (University of Verona, Italy)
Justus Liebig University, Giessen, May 26, 2010

14.30 ROBERTA FACCHINETTI – **Opening of the workshop**

14.45 PEDRO UREÑA GÓMEZ-MORENO (University of Granada, Spain)
Stance adverbs revisited: a corpus perspective on their use in e-news language

15.15 MARINA BONDI AND CORRADO SEIDENARI (University of Modena and Reggio Emilia, Italy)
Web-logs and news discourse: a corpus-driven analysis of evaluative phraseology in blog comments

15.45 ISABELL SCHULTHOFF (Universität Leipzig, Germany)
"*Capo* or *boss*? Constructing the Sicilian mafia in Italian newspapers: A corpus-based discourse analysis"

16.15 Coffee break

16.30 FEDERICO GASPARI (University of Bologna/Forlì, Italy)
A comparison of international news agency reports published online: Lexical bundles in the English-language output of *ANSA*, *Adnkronos*, *Reuters* and *UPI*

17.00 OLIVER MASON AND NICHOLAS GROOM (University of Birmingham, UK)
NLP meets CDA: investigating the linguistic representation of social actors in newspaper websites

17.30 STEFANIA SPINA (University of Perugia, Italy)
Methodological issues in a corpus of television news: structure, genres and annotation

18.00 ROBERTA FACCHINETTI – **Closing of the workshop**

Description Workshop II:

The focus of this pre-conference workshop will be on a set of topical issues pertaining to corpus-based studies on the language of news, including both printed and broadcast news.

So far, corpus studies in this context have focused mainly on written media and on the language of newspapers in particular. While not disregarding these studies, the workshop is intended to also address the interplay of different media in the actualization of news on television, radio and on the Internet. For example, we will also look at blogs, podcasts, vodcasts, and video sharing from a corpus-linguistic perspective.

Papers (20 mins + 10 mins discussion) are mainly focusing on methodological and theoretical issues concerning two lead questions:

- How can corpora and corpus-linguistic methods be applied to the study of news in old and new media, including the wide range of Internet-based communication?
- How do corpora and corpus-linguistic methods have to change to come to grips with this new multi-modal scenario?

Abstracts Workshop II:

Pedro Ureña Gómez-Moreno (*University of Granada*)

Stance adverbs revisited: a corpus perspective on their use in e-news language

Undoubtedly the media play an important role in transmitting information to form opinion and in shaping reality through their perspectives on the events being reported. Despite the fact that part of the media pursue a rigorous approach to reporting the news, these do not happen to adopt an utterly objective position towards the propositional content of their own discourse: rather, the opposite. While in general complying with maxims of impartiality and detachment, journalists depict the state of events couched in a more or less evaluative guise that they employ (un)deliberately for a number of reasons and discursive functions. Elaborating on a long tradition of prior research dealing with the expression of stance in different aspects of communication in general, and of language in particular, this study shows a corpus-based analysis of evaluative language within the domain of news, taking as a case in point the use of stance adverbs. For this purpose, I will draw on the analysis of a sample of e-news, which will be accessed through the WebCorp engine. The main objectives are to show the distribution of these adverbs in written journalistic discourse, and to analyse them with respect to the events being reported. Additionally, I will show the extent to which variation of usage in this case relies on linguistic context and register.

References

Bednarek, M. (2006). Evaluation in media discourse: Analysis of a newspaper corpus. London: Continuum.

Bell, A. (1991). The language of news media. Oxford: Blackwell.

Biber, D. and E. Finnegan. (1988). "Adverbial stance types in English". *Discourse Processes,* 11: 1-34.

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. (1999). Longman grammar of spoken and written English. London: Pearson Education Limited.

Biber, D. (2006). "Stance in spoken and written university registers". Journal of English for Academic Purposes, 5 (2) 97-116.

Bondi, M. (2008). "Emphatics in academic discourse: Integrating corpus and discourse tools in the study of cross-disciplinary variation". In A. Ädel, and R. Reppen. Corpora and discourse: The challenges of different settings. Amsterdam: Benjamins, 31-55.

Brand, C. (2008). Lexical processes in scientific discourse popularisation: A corpus-based study of the SARS coverage. Frankfurt am Main: Peter Lang.
Conboy, M. (2007). The language of the news. Routledge.
Conrad, S., and D. Biber. (2000). "Adverbial marking of stance in speech and writing". In S. Hunston, and G. Thompson (eds). Evaluation in text, Oxford: Oxford University Press. 56-73.
Facchinetti, R. (2003). "Pragmatics and sociological constraints on the functions of MAY in contemporary British English". In R. Facchinetti, M. Krug, and F. Palmer (eds.) Modality in Contemporary English. Berlin: Mouton de Gruyter, 301-327.
Fowler, R. (1991). Language in the news: Discourse and ideology in the press. London: Routledge.
Hidalgo, L. (2006). "The expression of writer stance by modal adjectives and adverbs in a comparable corpus of English and Spanish newspaper discourse". In A. M. Hornero, M. J. Luzón, and S. Murillo (eds.). Corpus Linguistics: Applications for the Study of English. Bern: Peter Lang. 125-140.
Hunston, S and G. Thompson. (2000). Evaluation in text: Authorial stance and the construction of discourse. Oxford: Oxford University Press.

~

Marina Bondi (*University of Modena and Reggio Emilia*)
Corrado Seidenari (*University of Modena and Reggio Emilia*)

Web-logs and news discourse: a corpus-driven analysis of evaluative phraseology in blogs.

The paper is part of a research project on evaluative meanings in Italian and English web-logs. Web-logs may be regarded as increasingly popular, virtual arenas where information (including news) is produced, shared and – crucially - commented on evaluatively.
Our focus here is on how bloggers evaluate the news in English and in Italian. Looking at business, politics and sport news on dedicated web-logs, we investigate the phraseology most frequently employed in comments on a piece of news.
The data for our study are taken from the Business, Politics and Sport sections of a blog corpus collecting texts dating from September 2008 to September 2009: in all 3,042,023 running words. The corpus is made up of 2 comparable sub-corpora: English blogs (1,459,575ws, 71 blog sources) and Italian blogs (1,582,448ws, 65 blog sources). Each sub-corpus is organized according to 6 macro-categories (Business, Entertainment, Science & Technology, Lifestyle, Politics, Sport) and 24 micro-categories derived from Technorati (http://technorati.com/blogs/directory/), one of the most popular blog directories on the WWW. For all sample texts, the beginning and the end of each post, comment thread and single comment is marked with the relevant tag.
Our methodology is adapted from Gledhill's (2000) study on Salient Grammatical Words (SGW). We start with an overview of the keyword lists extracted from the blog corpus, looking at potential markers of the nature and structure of news blogs. We then focus on markers of subjectivity, draw their collocational profile, identify evaluative uses in context, and, following Hunston (2008) and Groom (in press), explore the semantic sequences they are characterized by.
In the paper quantitative results are presented that confirm how deeply evaluative the language of web-logs can be, contributing evidence for them to be seen as virtual places for "you to have your say", rephrasing Baron (2008). Qualitative results are presented as well, documenting how blogs about business, politics and sport, even though relying on a common core evaluative phraseology, tend nonetheless to reveal a rather topic-specific evaluative phraseology. The conclusions explore the cultural implications of the analysis, looking at how parameters of evaluation are revealed in evaluative semantic sequences. This leads to emphasizing the central role of phraseology in constructing discourse and discursive identity.

~

Isabell Schulthoff (*University of Leipzig*)

Capo or *boss*? Constructing the Sicilian mafia in Italian newspapers: A corpus-based discourse analysis

On 11 April 2006, "il capo dei capi di Cosa nostra", the boss of bosses of the Sicilian Mafia, Bernardo Provenzano, was arrested after being on the run for more than 40 years. His capture and the end of the (predominant "bloody") leadership of the Corleonesi family that dominated Cosa Nostra provided the basis for a corpus-based discourse analysis of the representation of the Sicilian Mafia in three Italian newspapers: *Corriere della sera*, *La Repubblica* and *La Stampa* from May 1992 to April 2006 (including inter alia the articles about the the killing of the judges Giovanni Falcone and Paolo Borsellino, the arrest of the former no. 1 of Cosa Nostra, (Salvatore) Totò Riina in January 1993 and the capture of Bernardo Provenzano).
Using corpora and corpus linguistic techniques in discourse analysis has grown in recent years (Baker 2006, Baker forthcoming, ...). The present paper describes in particular the creation of the corpus and its sub-corpora and the problems associated with collecting the articles which refer to the Sicilian mafia from the different Internet archives of the respective newspapers. The corpus is subjected to a comparative analysis at a later point, by analysing the lexis of the different subcorpora and the corpus as a whole compared to an Italian reference corpus: CORIS / CODIS. Keywords, concordances and collocates (Stubbs 1996, 2001) are used in order to investigate the data by using WordSmith tools 5.0 (Scott 2008).

References:

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

Baker, Paul (forthcoming): "Representations of Islam in British broadsheet and tabloid newspapers 1999-2005", in: *Journal of Language and Politics*.

Scott, Mike (2008): *WordSmith Tool Version 5.0*. Liverpool: Lexical Analysis Software.

Stubbs, Michael (1996): *Text and Corpus Analysis*. Computer-assisted Studies of Language and Culture (= Language in Society 23). Oxford / Cambridge, Mass.: Blackwell.

Stubbs, Michael (2001): *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford / Cambridge, Mass.: Blackwell.

~

Federico Gaspari (*Università degli Studi di Bologna – sede di Forlì*)

A comparison of international news agency reports published online: Lexical bundles in the English-language output of *ANSA*, *Adnkronos*, *Reuters* and *UPI*

News agencies play a substantive role in determining what is reported by print, broadcast and web-based media all over the world (Van Dijk, 1988; Bell, 1991: 44ff.; Vuorinen, 1997; Read, 1999; Clausen, 2004; Richardson, 2007: 106ff.; Shrivastava, 2007; van Doorslaer, 2009). Stories carried by international news agencies are reproduced by several media outlets with varying degrees of adaptation, especially when reports are initially published on the Internet in major languages (Boyd-Barrett & Rantanen, 1998; Holland, 2006; Bielsa, 2007; Bielsa & Bassnett, 2009); English in particular exerts a strong influence on the (translation of) news that spreads globally (Hursti, 2001; Bassnett, 2005; Hajmohammadi, 2005; Kuo & Nakamura, 2005; Orengo, 2005; Schäffner, 2005; Conway & Bassnett, 2006; Lee, 2006; Kang, 2007).
The websites of the two leading Italian news agencies *ANSA* and *Adnkronos* currently offer constantly updated content in English, only part of which is devoted to domestic socio-political issues. This material results from the conflation of complex editorial processes involving rewriting of multiple sources, copy editing by a

combination of native and non-native speakers of English, cross-linguistic summarisation, and translation (often done by journalists), which in turn requires adaptation and the addition of background information for international readers.

The paper describes the steps taken in the construction of a corpus consisting of English-language reports published on the websites of four international news agencies: *ANSA* and *Adnkronos* (both based in Italy), plus *Reuters* and *United Press International (UPI)* – headquartered in London and Washington DC, respectively. Since their English-language output covers stories on a very wide range of topics originating anywhere in the world that are destined to a global Internet audience, the overall contents and potential readership of the texts circulated by these four news providers can be assumed to be broadly similar in principle.

The study focuses on lexical bundles as revealing indicators of the complex strategies at work in this heavily mediated news-making process (Biber et al., 1999; Stubbs, 2007). The lexical bundles found in the *ANSA* and *Adnkronos* online news reports are compared with those used by *Reuters* and *UPI*, in search for commonalities as well as divergences, and indeed peculiar usage patterns displayed by individual news agencies. While some lexical bundles occur with similar frequency in all four sources, the analysis looks more closely at those that are over- and under-represented across the sub-corpora; their discursive functions are discussed in more detail to account for the observed discrepancies.

References

Bassnett, S. (2005) "Bringing the News Back Home: Strategies of Acculturation and Foreignisation". *Language and Intercultural Communication* 5(2): 120-130.

Bell, A. (1991) *The Language of News Media*. Oxford: Blackwell.

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

Bielsa, E. (2007) "Translation in global news agencies". *Target* 19(1): 135-155.

Bielsa, E. & S. Bassnett (2009) *Translation in Global News*. London: Routledge.

Boyd-Barrett, O. & T. Rantanen (eds) (1998) *The Globalization of News*. London: Sage.

Clausen, L. (2004) "Localizing the global: 'domestication' processes in international news production". *Media, Culture & Society* 26(1): 25-44.

Conway, K. & S. Bassnett (eds) (2006) *Translation in Global News. Proceedings of the conference held at the University of Warwick, 23 June 2006.* Available at: http://www2.warwick.ac.uk/fac/arts/ctccs/research/tgn/events/tgn/translation-in-global-news-proceedings.pdf [last accessed 18th December 2009].

Hajmohammadi, A. (2005) "Translation evaluation in a news agency". *Perspectives: Studies in Translatology* 13(3): 215-224.

Holland, R. (2006) "Language(s) in the global news: Translation, audience design and discourse (mis)representation". *Target* 18(2): 229-259.

Hursti, K. (2001) "An insider's view on transformation and transfer in international news communication: an English-Finnish perspective". *Helsinki English Studies*, vol. 1. Available at: http://blogs.helsinki.fi/hes-eng/volumes/volume-1-special-issue-on-translation-studies/an-insiders-view-on-transformation-and-transfer-in-international-news-communication-an-english-finnish-perspective-kristian-hursti [last accessed 18th December 2009].

Kang, J.-H. (2007) "Recontextualization of News Discourse: A Case Study of Translation of News Discourse in North Korea". *The Translator* 13(2): 219-242.

Kuo, S.-H. & M. Nakamura (2005) "Translation or transformation? A case study of language and ideology in the Taiwanese press". *Discourse & Society* 16(3): 393-417.

Lee, C.-S. (2006) "Differences in News Translation between Broadcasting and Newspapers: A Case Study of Korean-English Translation". *Meta* 51(2): 317-327.

Orengo, A. (2005) "Localising News: Translation and the 'Global-national' Dichotomy". *Language and Intercultural Communication* 5(2):168-187.

Read, D. (1999) *The Power of News: the History of Reuters*. Second edition. Oxford: Oxford University Press.

Richardson, J.E. (2007). *Analysing Newspapers: An Approach from Critical Discourse Analysis*. Basingstoke: Palgrave MacMillan.

Schäffner, C. (2005) "Bringing a German Voice to English-speaking Readers: *Spiegel International*". *Language and Intercultural Communication* 5(2): 154-167.

Shrivastava, K.M. (2007) *News Agencies from Pigeon to Internet*. Elgin, IL: New Dawn Press.

Stubbs, M. (2007) "An Example of Frequent English Phraseology: Distribution, Structures and Functions". Facchinetti, R. (ed) *Corpus Linguistics Twenty-five Years On*. Amsterdam: Rodopi. 89-105.

Van Dijk, T.A. (1988) *News Analysis: Case Studies of International and National News in the Press*. Hillsdale, NJ: Lawrence Erlbaum.

van Doorslaer, L. (2009) "How language and (non-)translation impact on media newsrooms: the case of newspapers in Belgium". *Perspectives: Studies in Translatology* 17(2): 83-92.

Vuorinen, E. (1997) "News translation as gatekeeping". Snell-Hornby, M., Z. Jettmarová & K. Kaindl (eds) *Translation as intercultural communication: selected papers from the EST congress – Prague 1995*. Amsterdam: John Benjamins. 161-171.

~

Oliver Mason (*University of Birmingham*)
Nicholas Groom (*University of Birmingham*)

NLP meets CDA: investigating the linguistic representation of social actors in newspaper websites

Van Leeuwen (2008) describes a framework for analysing how human social actors (both individual and institutional) are represented in texts. In this presentation we report on an ongoing project in which we are attempting to apply this framework to a 200,000-word corpus of texts obtained from the websites of two UK newspapers, the left-wing *Guardian* and the right-wing *Daily Telegraph*.

The project described in our presentation has two main aims. Firstly, it aims to produce an analysis that will supplement van Leeuwen's theoretical framework with empirically-derived frequency information. Specifically, we want to find out which lexicogrammatical forms tend to be associated in news media texts with the different social actor roles identified by van Leeuwen, and in what proportions. This will allow researchers in the field of Critical Discourse Analysis to make statements about how common or uncommon certain phenomena are in media texts, thereby adding an empirical dimension to the critical analysis of media discourse.

Secondly, and more broadly, we aim to investigate the feasibility of supporting the analysis with existing techniques derived from computational linguistics. Identifying the representations of social actors is a time-consuming task, so the question arises as to how far this analysis can be automated. Using existing procedures from the field of natural language processing (such as named entity recognition, anaphora resolution, and a shallow parsing approach to identifying syntactic structures) we hope to speed up the analysis considerably, leaving the human analyst just with those aspects of representation that a computer program cannot handle. This would open up the topic for analysis on a much larger scale than has been possible hitherto, once again enhancing the robustness and validity of the claims that can be made within a general CDA framework. Newspaper articles could be analysed as soon as they were available on a newspaper's web site, with huge benefits to sociology, political science and media studies, as well as applied linguistics. Such a methodology would also be generalisable well beyond the confines of newspaper analysis, and would be particularly well suited to the critical analysis of blogs, MySpace pages, YouTube comments and other forms of electronic social media.

Reference:

Van Leeuwen, T. (2008): *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford: Oxford University Press.

~

Stefania Spina (*University for Foreigners Perugi*)

Methodological issues in a corpus of television news: structure, genres and annotation

This paper will describe some methodological issues concerning the constitution of the *Corpus di Italiano Televisivo,* 'Corpus of Italian Television Language' (Spina 2005), in the process of being compiled. The CIT is composed of transcriptions of television programs divided into five genres (current affairs, entertainment, advertising, sport, news). The data discussed in this paper come from the news subsection, which includes news programs transcribed in electronic form, and subsequently annotated and POS-tagged.

The aim of the corpus is to provide a balanced and consistent collection of data to perform research in the field of the language used on television. One of the most prominent features of the language of TV news is its great heterogeneity (Ennaji 1995; Spina 2009): from the macroscopic aspect of the mixing of different codes (news are a written text, with all its editorial activities of selection, deletion, and combination of different materials, but they are perceived by they final recipients as a spoken text), to its strong register variation (there is a constant tension between public/formal and private/informal registers that coexist and alternate in news: Fairclough 1995).

The process of annotation of the transcribed programs, however, has raised a number of methodological as well as theoretical issues on the structure of TV news (news are complex texts composed of linguistically differentiated parts) and on their status as a television genre (genre mixing is a more and more common practice in television news). The emergence of these issues from corpus annotation demonstrates that corpus-based methodologies can shed light on the linguistic specificity of TV news; it also proves, on the other hand, that new technological and methodological resources are needed to face the complexity of the transcription and annotation of television texts.

References

Ennaji, M. (1995). *A Syntactico-Semantic Study of the Language of News in Morocco*. International Journal of the Sociology of Language, 112, 97-111.

Fairclough, N. (1995). *Media Discourse*. London: Arnold

Spina, S. (2005). Il Corpus di Italiano Televisivo (Cit): struttura e annotazione. In E. Burr (Ed.), *Tradizione & Innovazione. Il parlato: teoria - corpora - linguistica dei corpora*. Proceedings of the VI SILFI Conference. Firenze: Franco Cesati: 413-426.

Spina S. (2009). Is Television Language a Spoken Variety? A Corpus-Based Study of Italian Television. Paper presented at the International Conference on Historical Linguistics (ICHL 2003, Copenhagen, 11-15 august 2003). in Rowe, Charley e Wyss, Eva (eds), *Language and New Media. Linguistic, Cultural and Technological Evolutions*, New York, Hampton Press, 211-242.

Corpus di Italiano televisivo: http://www.sspina.it/cit/cit.htm