Parsing. Psychological, Computational, and Theoretical Perspectives. Cambridge: CUP, 26-66.

Ferreira, F. (1991). "Effects of length and syntactic complexity on initiation times for prepared dentences." In: *Journal of Memory and Language* 30:2, 210–233.

Gibson, E. (1998). "Linguistic complexity: Locality of syntactic dependencies." In: *Cognition* 68:1, 1-76.

Sampson, Geoffrey (2001). *Empirical Linguistics.* London: Continuum.

Taavitsainen, Irma (2001). "Changing conventions of writing: The dynamics of genres, text types, and text traditions." In: *EJES* 5:2, 139-150.

Wasow, Thomas (1997). "Remarks on grammatical weight." In: *Language Variation and Change* 9:1, 81-105.

Rema Rossini Favretti / Fabio Tamburini
rema.rossini@unibo.it
University of Bologna
Exploring register features through statistical analysis and interpretation of corpus evidence
06.03.2009, 13.00–13.30 Uhr, Raum 41/111

In this paper we intend to explore the extent to which statistical analysis and interpretation of corpus evidence can contribute to the analysis of register variation. Encapsulating Firth's tradition, statistical association is considered as closely linked to co-occurrence frequency. Collocations are focused and investigated as surface evidence that is extracted from a corpus through automatic procedures – positional and frequency-based – and the application of probabilistic criteria. In a previous work (Rossini 2008) it was argued that collocations as habitual co-occurrences of words can be considered to reflect experiences which occur repetitively in a speech community. It was suggested that the statistical analysis of the habitual co-occurrences of a word can make it possible to capture elements of frame/s evoked by a word. Collocates that are most characteristic for the word chosen as the node have been analysed as providing clues the word's meaning and contextual form. Can they be considered as clues to register features? In the analysis reference is made to the text data contained in CORIS/CODIS, a 120-million word synchronic corpus of written language representing a wide range of varieties in written Italian – press, fiction, academic, legal and administrative prose, ephemera and miscellanea – which, through tools specifically developed, can be singularly or comparatively investigated. The results are illustrated in a graph where an automatic procedure recursively generates and compares collocation sets using the log-likelihood measure and presents them in a 3D-graph.

Bartsch, S. (2004). *Structural and Functional Properties of Collocations in English. A corpus study of lexical and pragmatic constraints on lexical co-occurrence.* Tübingen: Narr.

CORIS/CODIS corpus: http://corpora.dslo.unibo.it/coris_eng.html.

Evert, S. (2005). *The Statistics of Word Cooccurrences. Word pairs and collocations.* PhD Dissertation, IMS, University of Stuttgart.

Fillmore, C. J. / Baker, C. F. / Sato, H. (2004). "FrameNet as a 'Net'." In: *Proc. of LREC 2004, 5,* 1561-1564.