Detecting gender-preferential patterns of linguistic features in face-to-face communication

Jens Fauth and Hans-Jörg Schmid, University of Bayreuth, Germany

Gender-preferential language is a research area with a long tradition. Most of the previous investigations, however, included linguistic features that are not suited for large-scale automatic frequency counts in computerized corpora, because they are semantic or pragmatic in nature rather than grammatical, and/or difficult to operationalise (e.g. *references to emotions*). The present paper sets out to discover previously unknown gender-related linguistic features that are retrievable automatically from corpora.

Although the present study chooses a grammatical approach similar to Biber's (1988), it differs in that it does not only look for features that have already been suspected to bear a significant relation to gender but also for features that were not determined *a priori* in this respect. All in all 45 linguistic features from various grammatical domains were taken into consideration. Their occurrences were retrieved from the subcorpus *direct conversation* of *ICE-GB* and counted. Only conversations between same-sex participants were investigated because this constellation is known to enhance gender differences (for a complementary study on mixed-sex conversations see the abstract for this conference by Schmid and Fauth). In a second step the Pearson Correlation Coefficient was used to determine significant correlations between gender and linguistic features. The results were then used to predict the gender of the speakers of a conversation by computing a score for each text based on the correlating features. Basic descriptive statistics such as *mean, standard deviation* and *z-scores* were used for this purpose.

The results show that four linguistic features are correlated with female gender: third person personal pronouns, indefinite pronouns, predicative adjectives and intensive adverbs. Male gender is correlated with three different features: the definite article, nominalizations and NP postmodifications realized by *of*-PPs. These features are interpreted with regard to their communicative function and the extent to which their use reflects gender-specific topic choices. The methodology allows us to predict the gender of speakers in same-sex conversations with a probability of 88.10% for female texts and 85.80% for male texts.

Jens Fauth
Chair of English Linguistics
University of Bayreuth
D-95440 Bayreuth
E-mail: jot.fauth@gmx.net
Fax: +49 921 553627

Professor Hans-Jörg Schmid
Chair of English Linguistics
University of Bayreuth
D-95440 Bayreuth
Hans-Joerg.Schmid@uni-bayreuth.de
Fax: +49 921 553627