Register, sociolect and individual style:
Bridging the gap in Corpus-assisted Discourse Studies

Vaclav Brezina & Michael Barlow
Lancaster University & University of Auckland

In the analysis of discourse using language corpora, two distinct traditions can be identified: 1) corpus-based sociolinguistic studies and 2) studies of register variation. The former focus on the exploration of different linguistic features in relation to social variables such as the speaker's gender, age and socio-economic background (cf. Baker 2010; Watt 2007; Macaulay 2002). The latter follow the seminal work of Biber (1991 [1988]) and look into functional variation across different registers (cf. Biber and Conrad 2009). The present study aims to bridge the gap between the two traditions: it focuses on the interaction between register, social and individual variation in large language corpora and explores the consequences from a cognitive perspective. It is based on four spoken language corpora: *BNC 64* (1.5 million words), *Counselling sessions corpus* (2 million words), *Committee meetings corpus* (245,000 words), and a corpus of *White House Press Conferences* (3.85 million words). The corpora represent four different spoken situations: informal speech, counselling sessions, education committee meetings and White House press conferences.

   In particular, the study offers an insight into the use of a number of lexico-grammatical features (nominalizations, personal pronouns, epistemic markers, etc.) in the four spoken contexts. It explores the register-based, sociolectal as well as the individual variation using multidimensional factorial analysis. The results suggest that in the same dataset, register, social and individual variation can be observed along the same general dimension: involved, narrative & personalised vs. informational, descriptive & detached. As expected, more formal spoken registers cluster closer to the informational end (see Figure 1). In addition, male speakers consistently prefer a less involved style than female speakers in all the analysed spoken registers. Finally, the analysis shows that individual speakers themselves are fairly consistent in the use of the key lexico-grammatical features as shown by different samples from the same speaker that cluster close to each other.

   The study aims to contribute to the theoretical understanding of lexico-grammatical variation in speech in different social and situational contexts. The factors extracted in the factorial analysis explain a large part of this variation. Unlike previous studies, the present research also systematically examines individual variation, which is usually discarded as noise in the data (cf. Johnstone 1996).
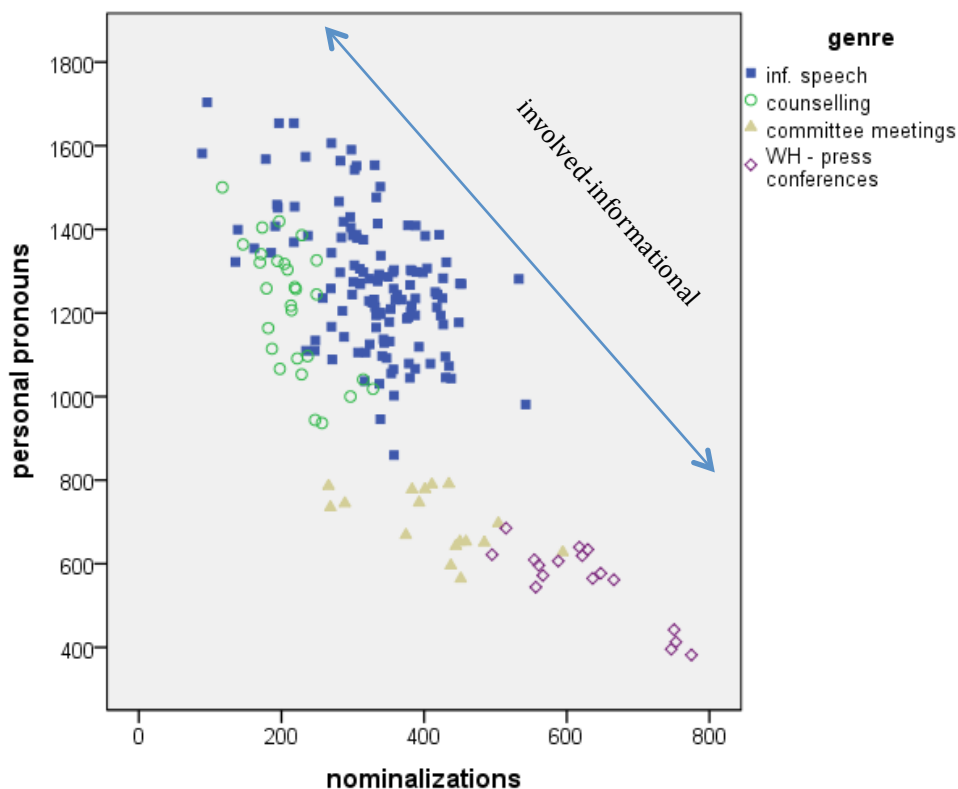

Figure 1: Personal pronouns and nominalisations in the four corpora