

Lexical Variation and Choice: A usage-based perspective on (semi-)synonymy

Eline Zenner (FWO Flanders), Dirk Speelman & Dirk Geeraerts
University of Leuven

Lexical variation is an underdeveloped topic in sociolinguistics and corpus linguistics: studies taking a “contextualized, pragmatic conception of onomasiology, which focuses on the actual choices made for a particular name as a designation of a particular referent” are rare (Grondelaers & Geeraerts 2003). Specifically, the issue of what features determine the choice a user makes from a set of (semi-)synonyms needs to be tackled more systematically.

Our research project contributes to this issue by examining the choice between English person reference nouns (*designer, lover*) and their Dutch alternative (*ontwerper, minnaar*) in Dutch newspapers. In a first study, we examined the influence of several features on the preference for 120 of these English nouns compared to their Dutch alternatives (based on a dataset of 1.5 million tokens for 100 concepts; Zenner *et al.* 2010). The selection of alternatives is restricted to pure synonyms. Though this “restrictive” selection criterion allows us to compare the behaviour a large number of concepts, the issue of semantic nuances between near-synonyms cannot be ignored: a comprehensive approach, which identifies these nuances and addresses their interplay with external, contextual features in determining lexical choice, is an indispensable addition (Geeraerts 2010).

In this paper, we incorporate this approach by performing an in-depth analysis of the near-synonyms *junk(ie)* and *verslaafde*. Our analysis relies on 11 000 tokens extracted from a Dutch newspaper corpus, which consists of over 1 billion words (from 1999 to 2005) and represents the two main national varieties of Dutch (Belgian Dutch and Netherlandic Dutch). The size of our dataset allows us to systematically investigate the influence of both semantic nuances and contextual features on the choice between both lexemes. Firstly, we identify the semantic content of the lexemes by, amongst others, determining the object of the addiction for each token (*drugs, fashion, rock’n’roll ...*). Next, all tokens are coded for a set of contextual features that may influence lexical choice in these contexts. Taking a broad definition of context, we include lexical features such as regional variation (Belgian vs. Netherlandic Dutch) and register variation (popular vs. quality newspapers). Taking a more narrow definition of context, we include the topic of the text, collocational patterns etc. We also take stylistic variation into account (e.g. avoidance of repetition).

Using logistic regression analysis, we determine the interplay of the semantic nuances and the contextual features in influencing lexical choice. We compare the results to those from the first method and highlight the benefits and drawbacks of both approaches. Also, we pay special attention to regional differences in (a) the semasiological range of the different lexical items and (b) the importance of the textual and conceptual features in lexical choice, by performing the regression analysis separately for Belgian and Netherlandic Dutch.

Overall, this project aims at adding to the debate on corpus-based studies of lexical variation, by showing the importance of establishing the relation between form, meaning and function and the interplay of this relation with external contextual features. Secondly, we hope to broaden the structural scope of existing borrowing research and add to current developments in loanword typology (Haspelmath & Tadmor 2009). Finally, we raise the question of which methodological steps need to be taken to allow for large-scale investigations of semantic and pragmatic nuances in near-synonyms.

Geeraerts, D., S. Grondelaers & P. Bakema. 1994. *The Structure of Lexical Variation. Meaning, Naming and Context*. Berlin/New York: Mouton de Gruyter.

Geeraerts, D. 2010. Lexical Variation in Space. In Auer, P. & J.E. Schmidt (eds.), *Language and Space: Theories and Methods*, 821-837. Berlin/New York: Mouton de Gruyter.

Grondelaers, S. & D. Geeraerts. 2003. Towards a pragmatic model of cognitive onomasiology. In Cuyckens, H., R. Dirven & J. Taeldeman (eds.), *Cognitive Approaches to Lexical Semantics*, 67-92. Berlin/New York: Mouton de Gruyter.

Haspelmath, M. & U. Tadmor (eds.). 2009. *Loanwords in the World’s Languages: A Comparative Handbook*. Berlin/New York: Mouton de Gruyter.

Zenner, E., D. Geeraerts & D. Speelman. 2010. The usage-based onomasiology of anglicisms: Shifting the focus from types to tokens. Paper at *Identifying & Describing Lexical Borrowings*, Belgium, March 2010.