Tracking Immigration Discourse through Time:
A Semantic Vector Space Approach to Discourse Analysis

Kris Heylen, Thomas Wielfaert, Dirk Speelman
QLVL, KU Leuven (University of Leuven, Belgium)

Theoretical Background

The focus of Cognitive Sociolinguistics (Geeraerts, Kristiansen, Peirsman 2010) on the social grounding of language has made it the locus for a rapprochement between Cognitive Linguistics and Critical Discourse Analysis, and thus spawned a number of cognitively inspired studies on the construal and framing of politics and identity in discourse (Dirven, Polzenhagen & Wolf 2007, Hart 2011). However, whereas Cognitive Sociolinguistics also stresses the need for a usage-based, quantitative methodology, studies in Critical Discourse Analysis still mainly rely on qualitative analysis and relatively small data sets. Although the complexity of semantic and pragmatic discourse phenomena necessarily involves a fair amount of interpretative analysis, we will argue that quantitative methods can help to strengthen the empirical base of discourse studies. In previous work (Peirsman, Heylen & Geeraerts 2010), we have shown how Semantic Vector Spaces can serve as an explorative tool to investigate in large data collections which wording is used to construe attitudes towards religions and how these change over time. In this paper we extend this work and use Semantic Vector Spaces to identify construal patterns in immigration discourse and we analyze the construal on the more fine grained level of specific utterances.

Case Study, data and methods.

As a case study, we look at immigration discourse in Belgium and the Netherlands in the period that both countries experienced a break-through of political parties with a strong anti-immigration platform (Vlaams Blok in Belgium in 1999 and Lijst Pim Fortuyn in the Netherlands in 2002). In a 1.3 million word corpus of Dutch and Belgian newspapers from 1999 to 2005, we collect the occurrences of lexemes referring to the concept IMMIGRANT in Dutch: *allochtoon, vreemdeling, (im)migrant, buitenlander, nieuwkomer*. We use Semantic Vector Spaces both on a type-level and a token-level (Turney & Pantel 2010, Heylen,Speelman & Geeraerts 2012) to structure the meaning relations between the lexemes and between their occurrences based on the similarity of their contextual usage. Multi Dimensional Scaling allows us to represent these contextual similarities in a 2-dimensional semantic space. Using an interactive visualization, we can then analyze the different construal patterns and their contextual realizations, the differences between the Netherlands and Belgium and between different newspapers, as well as the change in the construal patterns from 1999 to 2005.

Results

In both countries, the contextual usage of the lexemes shows an initial focus on three topics: (1) the effect of immigration problems on the local political system (election results, voting rights); (2) racism and discrimination as an ethical issue; (3) immigrants and crime. Although these topics remain in the discourse, there is a clear shifts to two new topics towards the middle of the decade: (1) a more practical discussion of how to increase the socio-economic opportunities of immigrants; (2) a focus on new immigration (as opposed to the immigrant population from the wave of immigration in the 50s, 60s and 70s). In general, this shift occurs later in the Netherlands then in Belgium, and more clearly in the high-brow newspapers then in the popular press

References

Dirven, R., Polzenhagen, F., & Wolf, H.-G. (2007). Cognitive Linguistics, Ideology, and Critical Discourse Analysis. In: D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics*,. 1222–1241, Oxford University Press.

Geeraerts, D., Kristiansen, G., & Peirsman, Y. (2010). *Advances in Cognitive Sociolinguistics.* Berlin/New York, Mouton de Gruyter.

Hart, C. (Ed.) (2011). *Critical discourse studies in context and cognition.* Amsterdam: Benjamins.

Heylen, K., Speelman, D., & Geeraerts, D. (2012). Looking at word meaning. An interactive visualization of Semantic Vector Spaces for Dutch synsets. In: *Proceedings of the EACL-2012 joint workshop of LINGVIS & UNCLH: Visualization of Language Patters and Uncovering Language History from Multilingual Resources*, 16–24.

Peirsman, Y., Heylen, K., & Geeraerts, D. (2010). Applying word space models to sociolinguistics. Religion names before and after 9/11. In: D. Geeraerts, G. Kristiansen & Y. Peirsman (Eds.),. *Advances in Cognitive Sociolinguistics*, 111–137.

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research* 37(1), 141–188.