

Putting Meaning on the Map: Integration of Geographic and Semantic Variation in Multivariate Models of Language Use

Natalia Levshina

ABSTRACT

This paper provides an overview of recent developments in (dia)lectally enriched lexical and constructional semantics. These developments reflect the quantitative turn that has been taken by dialectology and linguistics in general, as well as the active integration and cross-fertilization of different linguistic disciplines. This latter development has manifested itself vividly in multifactorial models, which reveal an interplay of various semantic, pragmatic, geographical, social and other factors in language use. The paper focuses on the onomasiological (from meaning to form) and semasiological (from form to meaning) studies that take geographic variation into account. It also discusses the achievements of aggregate lexical (dia)lectometry based on the onomasiological profile approach.

1. Recent trends in studies of dialectal variation

Like most linguistic disciplines, contemporary dialectology has recently been undergoing two major changes. On the one hand, it has taken a rigorously quantitative turn. On the other hand, we are witnessing a rapid integration of dialectology with neighbouring disciplines.

The quantitative turn has manifested itself in the move from ‘traditional’ dialectology to dialectometry (see Szmrecsanyi, this vol.), which measures linguistic distances between geographic locations. These distances are computed by aggregating the overlap across many individual variables (e.g. regional phonetic or lexical variants) in the speech of informants in different locations, as performed in a pioneering study by Séguy (1971). However, the differences between geographic variants are often a matter of degree. For instance, [zen], a variant of Dutch *zijn* ‘to

be', is more similar to [zɛ'n] than to [zɪnt], and all three variants are more similar to one another than to [bɪn] (see the data in Heeringa & Nerbonne 2001). To take such continua into account, dialectometrists compute acoustic/articulatory distances between phonemes or Levenshtein distances between the geographic variants of lexemes (e.g. Nerbonne et al. 1996). Next, the distances between the locations are added up across the words or phonemes, and can then be represented visually and analysed. One popular method is a cluster analysis of the locations, which allows dialect regions to be identified (Figure 1).



Figure 1. Six dialect groups in the Dutch-speaking area (Prokić et al. 2012). The clustering is based on the differences and similarities in the pronunciations of 562 words in 613 locations.

This method is currently being developed in several directions. First, in most dialectometric studies, which are based on data aggregation, one actually loses the information about the individual variables that enter the analysis. Several methods have been developed to overcome this drawback, e.g. Prokić 2007; Grieve et al. 2011; Prokić et al. 2012. For instance, Grieve's approach, which is based on spatial correlation measures and factor analysis, allows different dimensions of dialectal variation to be represented, retaining information about the initial linguistic variables, whereas the approach adopted by Prokić et al. (2012) identifies

the regional shibboleths (distinctive variants) of Dutch and German dialect areas. Also pertinent is the question of how to incorporate the social characteristics of individual speakers into dialectometric models, as achieved in traditional dialectology. A possible solution has been offered by Wieling (2012), who uses mixed-effect regression modelling, including a few demographic variables as fixed effects.



Figure 2. Dutch dialect areas based on the pronunciation of the word *vrijdag* “Friday” (Prokić et al. 2012). The word *vrijdag* is the most distinctive in the Frisian area (the dark region on the map).

On the other hand, the boundaries between dialectology and neighbouring disciplines are also becoming more flexible. The most natural and fruitful connection is probably between dialectology, typology and areal linguistics (cf. this volume). Provided that sufficient amounts of data are available, quantitative dialectometric models can be easily extended beyond one language. For instance, Gooskens & Heeringa (2004) consider the relationships between several geographic varieties of Frisian, on the one hand, and other Germanic languages, on the other. This enables them to model the dialect continuum in Frisian and also to interpret the distances between the Frisian varieties and other Germanic languages in

terms of genetic relationships and language contact. Another interdisciplinary example is perceptual dialectology, which integrates dialectology, social psychology, psycholinguistics and sociolinguistics. Basically, this is ‘folk dialectology’, which deals with the beliefs, attitudes, and stereotypes that ordinary speakers have about dialects (e.g. Preston 1999; Berthele 2010).

This article will focus on yet another recent development. Although there has been substantial progress in the dialectal studies of phonetic, and some morphological and syntactic alternations, which did not involve (obvious) semantic differences between the geographic variants, it is clear that substantial cross-dialectal variation in the meaning of words and constructions exists. For instance, in Pskov dialects, the noun *pivo* denotes any kind of beverage, whereas in contemporary literary Russian it means ‘beer’ (Lukjanova 2010). The dialectal variant reflects a more ancient meaning, which has become specialized. Conversely, the same concept can be denoted in different dialects by various words and constructions. An example is the famous difference in the way of naming a doughnut with a hole in Moscow (*pončik*) and St Petersburg (*pyška*). Note that the word *pončik* can be heard in St Petersburg as well, because the word also refers to round fried pies, sometimes with fillings (see Figure 3).

pončik (Moscow)/*pyška* (St Petersburg)



pončik (both cities)



Figure 3. Semantic variation of *pončik* and *pyška* in Moscow and Saint Petersburg.

Such semantic differences are often subtle and probabilistic. They involve varying degrees of entrenchment and salience of semantic features and lexical units, rather than crisp and clear distinctions (cf. Geeraerts et al. 1994). The question then is how to disentangle the semantic and geo-

graphic factors in a robust empirical model of language use. Several solutions exist, depending on the specific task and perspective of the study. They can be classified into three main groups, which are outlined below (examples are provided in the following sections):

1. Onomasiological studies of near-synonymous words and constructions. That is, the speaker's choice between different naming opportunities is modelled, establishing the role of the aforementioned factors in predicting the use of an expression in a given context. The weight of these factors is compared across the lects with the help of logistic regression or other statistical techniques. Such analyses normally represent integrative quantitative models of language variation, which include geographic, social, semantic and pragmatic factors. This is why the approach is often referred to as multifactorial grammar. These studies are based on comparable corpora, which contain balanced samples from different language varieties, commonly referred to as "lects"—an umbrella term for any geographic and social language varieties, including idiolects.
2. Semasiological studies, which focus on variation in the use of a word or construction in different lects. These analyses differ in the number and nature of the variables which serve as a *tertium comparationis*. Some of them are based on collocational analysis, like the so-called distinctive collexeme analysis developed by Gries & Stefanowitsch (2004), which compares the slot fillers of a construction in different lects. Other methods are multifactorial and involve large sets of heterogeneous variables.
3. Aggregate lectometric studies, which measure cross-lectal lexical differences across the concepts that belong to a specific lexical field. One of the best-known approaches is based on so-called onomasiological profiles, which consist of the lexical terms that designate a specific concept, and their relative frequencies in two or more language varieties.

The methods mentioned above are based on several fundamental assumptions. First, corpus data are seen as a reliable source of evidence about our knowledge of a language because they represent non-elicited usage in natural settings. Second, they are based on the distributional hypothesis, which says that natural language semantics can be modelled

with the help of contextual features found in large amounts of text data. This idea is expressed in John Firth's (1957: 11) famous maxim "You shall know a word by the company that it keeps." An early implementation of the distributional approach in lexical semantics is the distributional analysis of verbal semantics by Apresjan (1966).

The structure of the remaining part of the paper is as follows. The next section provides examples of lectally enriched onomasiological studies of specific words and constructions. Section 3 describes some quantitative models of semasiological variation of words and constructions in different lects. Section 4 offers an overview of studies based on aggregation of cross-lectal differences in the lexical choices. Finally, Section 5 provides a summary and suggests some paths for further investigation.

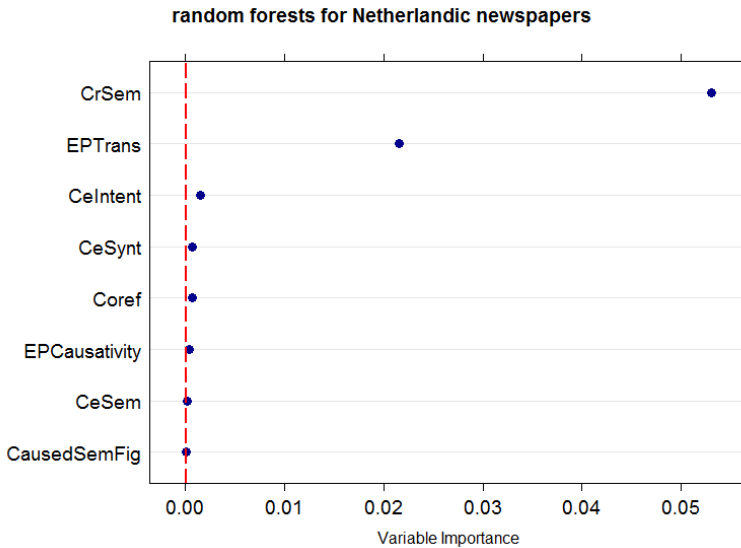
2. *Modelling the speaker's choice: onomasiological multifactorial grammar*

The multifactorial onomasiological studies that model the speaker's choice between two or more constructions deal mainly with the national varieties of English and Dutch. They usually focus on abstract syntactic constructions. A well-known example is a series of papers on the English dative alternation by Joan Bresnan and her colleagues (Bresnan et al. 2007; Bresnan & Hay 2008; Grimm & Bresnan 2009; Bresnan & Ford 2010). The alternation involves two variants: the double object construction, as in *Mary gives John the book*, and the prepositional *to*-dative, as in *Mary gives the book to John*. In Bresnan et al. 2007, the authors show that the use of one or other of the constructions can be predicted with a high degree of precision by contextual factors. The multivariate analyses (logistic regression) provide clear evidence of a 'harmonic alignment' of distinct but related features. Namely, discourse givenness, animacy, definiteness, pronominality and the relative length of the postverbal participants (the recipient and the theme) are aligned with the position that follows the verb first (the recipient in the double object construction, and the theme in the prepositional dative construction), whereas non-givenness, inanimacy, etc. are observed in the final position in both constructions.

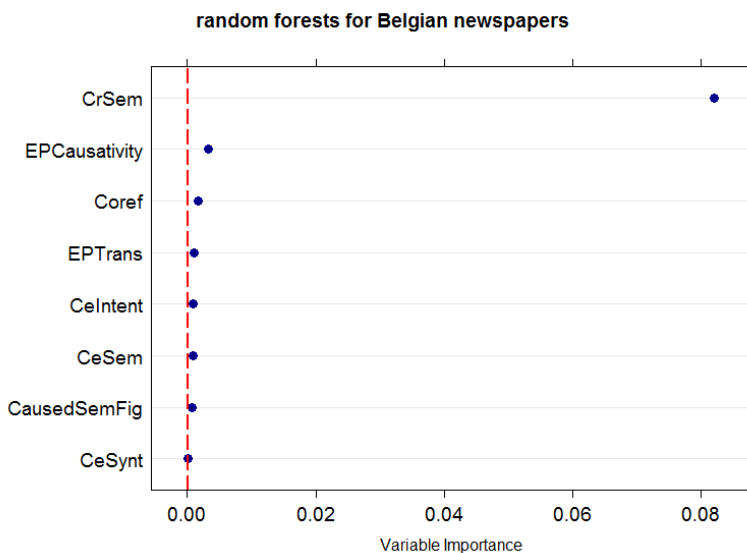
The follow-up studies of Bresnan et al. (2007) show that, although the main division of pragmatic labour between the constructions holds across many national varieties of English, there are subtle yet significant differences in the effect size of different variables. This suggests that the speakers of different varieties may be more or less sensitive to particular

contextual cues. For instance, a comparison of size effects for the dative alternation with the verb *give* in American and New Zealand English (Bresnan & Hey 2008) demonstrated that New Zealand speakers are more sensitive to animacy than Americans, and that this difference is a result of a gradual language change.

Other examples of slightly different constraints between lects have been observed in the use of the *'s-* vs. *of-*genitive in several varieties and registers of English (Szmrecsanyi 2010), in the presence or absence of presentative *er* “there” in Belgian and Netherlandic Dutch (Grondelaers et al. 2002), and in the semantics of the Dutch analytic causatives with auxiliaries *doen* “do” and *laten* “let” in Netherlandic and Belgian (Flemish) newspapers (Levshina 2011). Figure 4 shows the effects of semantic and syntactic variables in predicting the choice between the auxiliaries for the Netherlandic and Belgian newspaper data. The effect sizes were measured with the help of random forests (cf. Tagliamonte & Baayen 2012). The further the point from the origin, the more important the variable.



(A)



(B)

Figure 4. Different effect sizes of contextual variables in the choice between two Dutch causative verbs in the newspaper register. A: Netherlandic data; B: Belgian (Flemish) data.

The plots show that the effect of transitivity on the effected predicate *EPTrans* (e.g. *ontwerpen* ‘design’ in *Hij liet zijn huis ontwerpen* ‘He had his house designed’) plays a much stronger role in the choice between the two constructions in the Netherlandic newspapers than it does in the Flemish ones. Closer inspection, however, shows that this is a result of the high frequency of the collocations of *laten* with the transitive verbs *weten* ‘know’, *zien* ‘see’ and *horen* ‘hear’ in the Netherlandic newspapers. This demonstrates the important role played by lexically specific exemplars in constructional variation. Such lexical differences may arise due to many possible factors, for instance, the different degrees of importance of particular referents and their characteristics in language communities, or the different degrees of “routinization” of particular linguistic expressions, which makes them more or less readily available to the speakers in the process of communication. The inventory and relative weight of such factors in language use are still to be established.

Quantitative, lectally enriched models of lexical variation are, unfortunately, quite rare in contemporary usage-based linguistics. One of the few examples is Glynn's (2009) study of the near-synonyms *annoy*, *bother* and *hassle* in British and American English. He does not, however, find substantial geographic differences in the division of semantic labour between the verbs.

3. *Modelling lectal differences in the semantic structure of words and constructions*

3.1. *Lectally oriented distinctive collexeme analysis*

The distinctive collexeme analysis developed by Gries & Stefanowitsch (2004) is based on analysis of corresponding constructional slot fillers in near-synonymous constructions (e.g. the double object dative vs. the prepositional dative) in different lectal variants of the same constructions. It has been applied by Wulff et al. (2007) to model the semantic differences between the British and American variants of the *V-into-V* causative, as in *He tricked me into employing him*. The authors compare thousands of verb pairs that fill in the first and the second verb slots, and employ a statistical test (Fisher's exact test) to identify the verbs that are the most typical of one or other lect. For instance, many verbs distinctive of the construction's British (more precisely, those that occupy the first verbal slot, e.g. *trick* in the example above) denote negative emotions (e.g. *She terrified me into doing it*) and threatening (*He blackmailed me into doing it*). These semantic classes are not typical of the American collexemes. On the other hand, the distinctive American verbs often refer to communication (e.g. *She talked me into doing it*) with a total lack of such distinctive collexemes in the corresponding slot of the British construction. Wulff et al. hypothesize that these and other differences may reflect the varying degrees of entrenchment of specific semantic frames in the two cultures, hence the title of their paper, "Brutal Brits and Persuasive Americans."

A similar analysis was carried out in Levshina et al. 2009, which compared the Netherlandic and Belgian variants of the causative construction with *doen* mentioned above. The analyses of the constructional slots revealed that the Netherlandic variant seems to be restricted to the so-called affective causation, i.e. situations when a stimulus produces a mental reaction, e.g. *Zijn kapsel doet me denken aan een vogelnest* 'His

hairstyle reminds me of (lit. makes me think) of a bird's nest'. Since Netherlandic Dutch is traditionally considered to be the leader in various processes of language change, and Belgian Dutch is believed to be more archaic, the more limited semantic repertoire of the Netherlandic *doen* ties in well with previous observations of the ongoing qualitative and quantitative shrinking of the auxiliary *doen* in Dutch (e.g. Duinhoven 1994; Verhagen 2000).

3.2. *Multivariate comparisons of the semantics of lectal variants*

The multivariate data of the type described in Section 2 can also be used to compare the semantic differences between different lectal variants of the same word or construction. One of the examples is Glynn's (*forthcoming*) analysis of the British and American differences in the polysemy of *annoy*. As in the aforementioned study of the verbs *annoy*, *bother* and *hassle*, he uses a multiple correspondence analysis to explore the relationships between various semantic features and the national varieties, but this time he focuses only on one specific verb, *annoy*. The analyses show that, although the central sense ('anger') is shared by both varieties, there are some lectal differences on the semantic periphery of the verb. More specifically, the American variant often denotes 'hurt', which is associated with serious topics, e.g. anxiety, personal relationship troubles or family concerns. The British variant is more often used in the sense 'tire/interrupt', often in a humorous way, being also associated with such features as familiarity, interruptions and expenditure of energy.

Multivariate methods, such as cluster analysis and multidimensional scaling, also allow one to compare the general semantic organization of lectal variants. For instance, in an analysis of corpus data coded for 35 semantic, morphological, syntactic and other variables, Levshina (2012) found that the Netherlandic causative *doen* (see above) is not only less frequent and less semantically diverse than its Belgian Dutch counterpart, but also has a different structure with a tight cluster that corresponds to affective causation, and only sporadic uses of the other senses. This is shown in Figure 5, which represents the individual exemplars of the construction in the common semantic space created with the help of a multidimensional scaling analysis. Note that the closer the exemplars are on the map, the more semantic features they share. The results suggest that the general schema of the Netherlandic causative has become weaker

than that of the Belgian variant. This finding is in line with the process of *doen* gradually becoming obsolete.

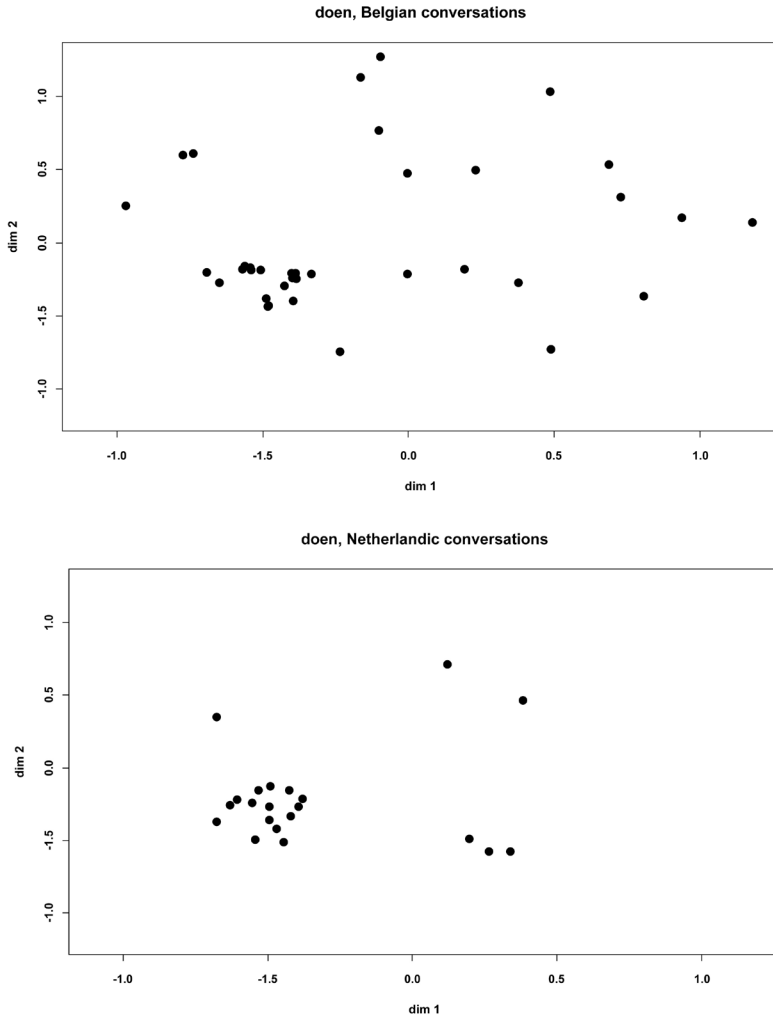


Figure 5. The semantic space of the causative *doen* in Dutch with the Belgian (on top) and Netherlandic (below) exemplars (spoken data). The Belgian causative auxiliary demonstrates greater semantic variation and productivity than its Netherlandic counterpart, which is almost exclusively represented by the expression *doen denken aan* ‘remind, make think of’ (the dense cluster on the left).

4. *Aggregate lexical lectometry: onomasiological profiles*

The onomasiological profiles approach was developed in Geeraerts et al. 1999 and tested on two lexical fields—clothing and soccer terms. The data represented several lectal dimensions: geographic (two Belgian vs. two Netherlandic locations), temporal (1950, 1970 and 1990) and register-related (magazines and newspapers, on the one hand, and shop window labels [for clothes], on the other). An onomasiological profile is formed by different naming choices for a concept available to speakers of a certain language. For instance, the most common names for a dress in Dutch are *jurk*, *kleed* and *japon*. Next, the relative frequencies of use of each lexical variant are calculated in each lect. Such profiles are created for every concept (e.g. dress, shirt, offside) in each lexical field, and the aggregate uniformity indices between the lects are computed. A comparison of the uniformity measures revealed a strong lexical convergence between the Belgian and Netherlandic varieties, as the uniformity measures steadily increased from 1950 to 1990. The method also made it possible to measure the distances between different strata within each geographic variety. For instance, it has been shown that the uniformity index between the Belgian magazines (the standard variety) and shop-window data (a substandard stratum) is lower than that shown between the corresponding Dutch sub-varieties. This suggests a difference in the structure of the dialect repertoires in the Netherlands and the Dutch-speaking part of Belgium (cf. Auer 2005), which relates to the belated standardization of Belgian Dutch. A more advanced quantitative approach has recently been proposed by Ruetten & Speelman (2014) who use the data from Geeraerts et al. (1994). Their approach is based on Individual Differences Scaling, which is a form of Multidimensional Scaling. It allows for aggregate differences to be captured but also keeps open the possibility of investigating the behaviour of individual variables, similar to the recent dialectometric studies that were discussed at the beginning of this paper.

The study by Geeraerts et al. (1999) has been replicated for European and Brazilian Portuguese by Soares da Silva (2010). The uniformity indices show that the varieties are diverging in the vocabulary of clothing. The results also indicate a greater difference between the Brazilian standard and substandard strata than between those in European Portuguese.

5. *Summary and perspectives*

The studies cited above demonstrate that semantic and geographic variation can be integrated in fine-grained and robust models of language use. Although most cross-lectal differences in the use of constructions and words are quite subtle, they are nonetheless very revealing, often providing signals of ongoing processes of language change, or suggesting interesting lectal differences in the entrenchment of specific semantic frames.

So far, the approaches presented here have only been applied to a few languages. Most languages (as well as words and constructions) are still waiting their turn. Of course, the methods mentioned above are highly labour-intensive because they require large representative data sets and advanced statistical techniques. However, these approaches are an important step towards a more complete and realistic model of language variation.

The future directions of research, in my view, mostly relate to “catching up” with the achievements of quantitative dialectology and dialectometry in phonetics and morphology. Unfortunately, the study of lexical and syntactic variation in non-elicited texts requires very large data samples, which should be collected for many locations and individuals. On the other hand, it is crucial that a much larger repertoire of constructions and lexical fields is explored, in order to be able to see the general tendencies. To achieve these goals, we should use computational linguistic tools to automate the semantic annotation process. Finally, the studies should go beyond one language and take into account areal effects and genetic relationships. All this will make lexical and constructional dialectology a more challenging, but also more fruitful enterprise in the future.

References

- Apresjan, Juri, 1966, “Analyse distributionnelle des significations et champs sémantiques structurés,” *Langages* 1, pp. 44–74.
- Auer, Peter, 2005, “Europe’s Sociolinguistic Unity, or: A Typology of European Dialect/Standard Constellations,” *Perspectives on Variation: Sociolinguistic, Historical, Comparative*, eds. N. Delbecq, J. van der Auwera & D. Geeraerts, Berlin, pp. 7–42.
- Berthele, Raphael, 2010, “Investigation into the Folk’s Mental Models of Linguistic Varieties,” *Advances in Cognitive Sociolinguistics*, eds.

- D. Geeraerts, G. Kristiansen & Y. Peirsman, Berlin & New York, pp. 265–90.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen, 2007, “Predicting the Dative Alternation,” *Cognitive Foundations of Interpretation*, eds. G. Boume, I. Krämer & J. Zwarts, Amsterdam, pp. 69–94.
- Bresnan, Joan and Marilyn Ford, 2010, “Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English,” *Language* 86 (1): 168–213.
- Bresnan, Joan & Jennifer Hay, 2008, “Gradient Grammar: An Effect of Animacy on the Syntax of give in New Zealand and American English,” *Lingua* 118 (2), pp. 245–59.
- Duinhoven, Anton M., 1994, “Het hulpwerkwoord doen heeft afgedaan,” *Forum der Letteren* 35 (2), pp. 110–31.
- Firth, John R., 1957, “A Synopsis of Linguistic Theory 1930–1955,” *Studies in Linguistic Analysis*, ed. J.R. Firth, Oxford, pp. 1–32.
- Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema, 1994, *The Structure of Lexical Variation. Meaning, Naming, and Context*, Berlin & New York.
- Geeraerts, Dirk, Stefan Grondelaers & Dirk Spielman, 1999, *Convergentie en divergentie in de Nederlandse woordenschat*, Amsterdam.
- Geeraerts, Dirk, Gitte Kristiansen & Yves Peirsman, eds. 2010, *Advances in Cognitive Sociolinguistics*, Berlin & New York.
- Glynn, Dylan, 2009, “Synonymy, Lexical Fields, and Grammatical Constructions. A Study in Usage-Based Cognitive Semantics,” *Cognitive Foundations of Linguistic Usage Patterns: Empirical Studies*, eds. H.-J. Schmid & S. Handl, Berlin & New York, pp. 89–117.
- Glynn, Dylan, *forthcoming*, “Sociolinguistic Cognitive Semantics: A quantitative Study of Dialect Effects on Polysemy,” *Review of Cognitive Linguistics*.
- Gooskens, Charlotte & Wilbert Heeringa, 2004, “The Position of Frisian in the Germanic Language Area,” *On the Boundaries of Phonology and Phonetics, To Honour Dr. Tjeerd de Graaf*, eds. D. Gilbers, M. Schreuder & N. Knevel, Groningen, pp. 61–87.
- Gries, Stefan Th. & Anatol Stefanowitsch, 2004, “Extending Collostructional Analysis: A Corpus-Based Perspectives on ‘Alternations,’” *International Journal of Corpus Linguistics* 9 (1), pp. 97–129.

- Grieve, Jack, Dirk Speelman & Dirk Geeraerts, 2011, "A Statistical Method for the Identification and Aggregation of Regional Linguistic Variation," *Language Variation and Change* 23, pp. 93–221.
- Grimm, Scott & Joan Bresnan, 2009, "Spatiotemporal Variation in the Dative Alternation: a Study of Four Corpora of British and American English," Paper Presented at Conference Grammar & Corpora 2009, September 2009, Mannheim.
- Grondelaers, Stefan, Dirk Speelman & Dirk Geeraerts, 2002, "Regressing on *er*: Statistical Analysis of Texts and Language Variation," *Proceedings of the 6th International Conference on the Statistical Analysis of Textual Data*, eds. A. Morin & P. Sébillot, Rennes, pp. 335–46.
- Heeringa, Wilbert & John Nerbonne, 2001, "Dialect Areas and Dialect Continua," *Language Variation and Change*, vol. 13, eds. D. Sankoff, W. Labov & A. Kroch, New York, pp. 375–400.
- Levshina, Natalia, 2011, *Doe wat je niet laten kan: A Usage-Based Analysis of Dutch Causative Constructions*, PhD thesis, University of Leuven.
- Levshina, Natalia, 2012, "Comparing Constructions: A Usage-Based Analysis of the Causative Construction with *doen* in Netherlandic and Belgian Dutch," *Constructions and Frames* 4 (1), pp. 76–101.
- Levshina, Natalia, Dirk Geeraerts & Dirk Speelman, 2011, "Changing the World vs. Changing the Mind: Distinctive Collexeme Analysis of the Causative Construction with *doen* in Belgian and Netherlandic Dutch," *Language Variation—European Perspectives III: Selected Papers from the 5th International Conference on Language Variation in Europe, Copenhagen, June 2009*, eds. F. Gregersen, J. Parrot & P. Quist, Amsterdam, pp. 111–23.
- Lukjanova, Svetlana V., 2010, "Naimenovanija napitkov v narodnoj reči (na materiale pskovskix govorov)," *Vestnik Pskovskogo gosudarsvennogo pedagogičeskogo instituta*, Pskov, pp. 33–37.
- Nerbonne, John, Wilbert Heeringa, Eric van den Hout, Peter van de Kooi, Simone Otten & Willem van de Vis, 1996, "Phonetic Distance between Dutch Dialects," *CLIN VI: Proceedings of the Sixth CLIN Meeting*, eds. G. Durieux, W. Daelemans & S. Gillis, Antwerp: pp. 185–202.
- Preston, Dennis R., ed. 1999, *Handbook of Perceptual Dialectology*, Amsterdam.

- Prokić, Jelena, 2007, "Identifying Linguistic Structure in a Quantitative Analysis of Dialect Pronunciation," *Proceedings of the ACL 2007 Student Research Workshop*, Prague, pp. 61–66.
- Prokić, Jelena, Çağrı Çöltekin & John Nerbonne, 2012, "Detecting Shibboleths," *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, Avignon, pp. 72–80.
- Ruette, Tom & Dirk Speelman, 2014, "Transparent Aggregation of Variables with Individual Differences Scaling," *Literary and Linguistic Computing* 29 (1), 89–106.
- Séguy, Jean, 1971, "La relation entre la distance spatiale et la distance lexicale," *Revue de Linguistique Romane* 35 (138), 335–57.
- Soares da Silva, Augusto, 2010, "Measuring and Parametrizing Lexical Convergence and Divergence between European and Brazilian Portuguese," *Advances in Cognitive Sociolinguistics*, eds. D. Geeraerts, G. Kristiansen & Y. Peirsman, Berlin & New York, pp. 41–84.
- Szmrecsanyi, Benedikt, 2010, "The English Genitive Alternation in a Cognitive Sociolinguistics Perspective," *Advances in Cognitive Sociolinguistics*, eds. D. Geeraerts, G. Kristiansen & Y. Peirsman, Berlin & New York, pp. 141–66.
- Tagliamonte, Sali A. & R. Harald Baayen, 2012, "Models, Forests and Trees of York English: Was/Were Variation as a Case Study for Statistical Practice," *Language Variation and Change* 24 (2), pp. 135–78.
- Verhagen, Arie, 2000, "Interpreting Usage: Construing the History of Dutch Causal verbs," *Usage Based Models of Language*, eds. M. Barlow & S. Kemmer, Stanford, pp. 261–86.
- Wieling, Martijn, 2012, *A Quantitative Approach to Social and Geographical Dialect Variation*, PhD thesis, Rijksuniversiteit Groningen.
- Wulff, Stefanie, Anatol Stefanowitsch & Stefan Th. Gries, 2007, "Brutal Brits and Persuasive Americans: Variety-Specific Meaning Construction in the *Into*-Causative," *Aspects of Meaning Construction*, eds. G. Radden, K.-M. Köpcke, T. Berg & P. Siemund, Amsterdam & Philadelphia, pp. 265–81.