

Dirk Geeraerts (Leuven)

## A lectometric definition of lexical destandardization

**Abstract:** This paper argues that a lectometric approach may shed light on the distinction between destandardization and demotization, a pair of concepts that plays a key role in ongoing discussions about contemporary trends in standard languages. Instead of a binary distinction, the paper proposes three different types of destandardization, defined as quantitatively measurable changes in a stratigraphic language continuum. The three types are illustrated on the basis of a case study describing changes in the vocabulary of Dutch in The Netherlands and Flanders between 1990 and 2010.

Recent theory formation about the evolution of standard languages in Europe has been dominated by two ideas: on the one hand, Auer's typology of dialect and standard language constellations (Auer 2005, 2011); on the other, the notions of 'destandardization' and 'demotization' introduced by the SLICE (Standard Language Ideology in Contemporary Europe) network. Disregarding many subtypes and variations painstakingly described by Auer, the former involves the idea that the languages of Europe tend to follow a long-term evolution from exoglossic diglossia in the medieval period to endoglossic diglossia in Early Modern times, followed by an evolution to a diaglossic situation – a fully fleshed out stratigraphic spectrum between standard language and base dialects – in the Modern period. In some cases, dialect loss in the contemporary period may further lead to a shrinking of the spectrum. In the framework developed by the SLICE network (Coupland/Kristiansen 2011; Kristiansen 2016), contemporary changes at the top of the stratigraphic spectrum are considered. Specifically, an increasing, 'post-modern' tolerance for variation is supposed to take shape in two different forms: either as 'demotization' or as 'destandardization'. Demotization (a terminological reference to the *demotisierung* introduced by Mattheier 1997) involves cases in which more variation enters into the standard language but in which the standard language ideal as such is not affected: the 'best language' becomes more open to variation, but the normative concept of a best language as such is not weakened. Destandardization by contrast involves changes through which established standard languages lose their exclusive status as 'best language' and a broader range of speech varieties is accepted within the public sphere. The distinction between destandardization and demotization has triggered a lot of debate, not least so because they were not introduced with a clear operational definition.

The present paper, then, develops the idea that a lectometric approach may help to see clear in the process of destandardization and its mechanisms. (Unless otherwise stated, we will use *destandardization* in a broad sense as referring to any type of loss of linguistic standardization, and not in the specific reading in which it contrasts with *demotization*.) In general, lectometry uses quantitative measures to establish the relative similarity (or distance) between different lects, where lects are thought of as collections of linguistic features that can vary along any extra-linguistic contextual dimension in the broadest sense possible. Thus, in dialectometry (e.g. Goebel 2006; Heeringa 2004; Nerbonne/Kleweg 2003; Séguy 1971; Szmrecsanyi 2013), stylometry (e.g. Grieve 2007; Luyckx/Daelemans 2011), and some forms of language perception research (e.g. Gooskens/Heeringa 2004; Van Bezooijen/Heeringa 2006; Speelman/Impe/Geeraerts 2014), distances between lects are studied along the geographical, discursive, and subjective axis respectively. The present study belongs to a fourth subfield of lectometry, viz. sociolectometry, which quantifies aggregate language variation in relation to sociolinguistic variation. Specifically, the paper takes a lexical lectometric approach. Developing the framework set out in Geeraerts/Grondelaers/Bakema (1994), a pioneering example of such a lexical sociolectometric approach is Geeraerts/Grondelaers/Speelman (1999), which examines lexical variables to measure the relation between the two main national varieties of Dutch. Expanding on this early work in sociolectometry, Speelman/Grondelaers/Geeraerts (2003), Ruetten et al. (2014) and Ruetten/Ehret/Szmrecsanyi (2016) provide methodological elaborations by introducing techniques such as cluster analysis, multi-dimensional scaling and Semantic Vector Space models. Descriptive elaborations involve a.o. contact linguistics (Zenner/Speelman/Geeraerts 2012) and the interaction of semantic and lexical variation (Speelman/Geeraerts 2009).

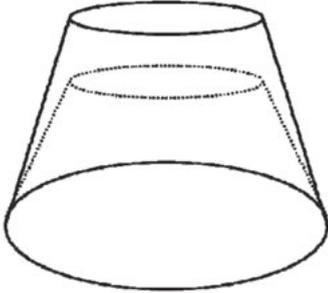
To see how a lectometric approach may shed light on the issue of destandardization, the conical representation of stratigraphic spectra used by Auer (2005) provides a fruitful starting-point. The conical visualization assumes an essentially two-dimensional structure of variation. The vertical dimension represents a hierarchical ordering along a situational dimension: the higher a situation is situated in the stratificational cone, the more standard language use will be expected. Informative media language, for instance, whether written or spoken, will generally be expected to conform to the standard language norm, regardless of how internally varied that norm may be. Casual conversations in an informal context, by contrast, will generally come with less outspoken expectations w.r.t. standard language use. The horizontal dimension, conversely, may primarily be thought of in terms of geographic variation: to the extent that dialect differences exist, they will show up more readily in situations with less stringent standard language expectations. But the geographical dimension would obviously not be the only

one to be considered alongside the situational dimension; at least social features (such as the speaker characteristics of sociolinguistics) and thematic differences (as for instance in Language for Special Purposes) would need to be added to get a more complete picture of the variation. The conical representation, in other words, is a simplified model of a multidimensional variational structure, but precisely as a simplified model, it can help us to think analytically about the dynamics of standardization – always keeping in mind that more complicated approaches may need to be introduced later to accommodate the multidimensionality of variation.

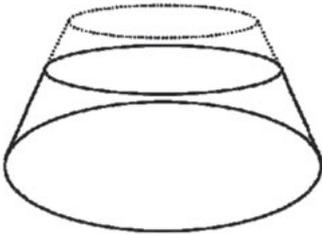
So when could we talk about an increasing or decreasing standardization, assuming a basic conical structure with two layers? Three structural changes need to be distinguished. In the first place, standardization could take the form of a decrease in the distance between the two levels, where ‘distance’ metaphorically refers to the degree of similarity between the language use found in the two situations. A case in point would be the processes of dialect loss leading to a compression of the stratigraphic spectrum, as mentioned in Auer’s typology. Figure 1 graphically represents the opposite development, i.e. a process of decreasing stratigraphic standardization. The dotted line represents the original situation, while the solid line depicts a situation in which the upper level has moved further away from the base level (with both levels in themselves maintaining their original degree of variation, as represented by the surface of the ellipse).

In the second place, a decrease in the distance between the two levels might still be considered a form of destandardization if the movement between the two levels is rather from the top level to the bottom level rather than the other way around. When not just the degree of rapprochement but also the direction of the process is taken into account, a distinction can be made between the type of standardization that fits the traditional (some would say ‘ideological’) conception of standardization, and developments in the other direction. In the former, the features of the hierarchically superior level trickle down towards the inferior one, as when colloquial language use loses any remaining dialect features and becomes an RP-like polite speech. In the latter, what used to be informal language percolates into the formal, upper-level situations, thus bringing qualitative change in the substance of the standard language. This second process is a type of destandardization to the extent that the old standard norm gives way to a new one that is influenced by the initial informal, colloquial, less valued forms of language use. One could therefore say, perhaps, that it is a qualitative type of destandardization in comparison with the quantitative destandardization that occurs when the stratificational distance between the two levels grows. But we will see presently that this qualitative destandardization can also be defined quantitatively, and moreover, there are other candidates for a terminological

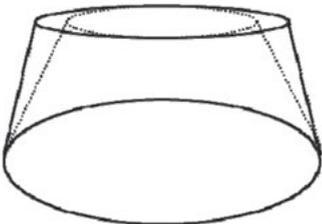
identification of this second process: ‘informalization’ e.g. could be considered as one of the likely terminological options. Figure 2 graphically represents a process of qualitative destandardization.



**Fig. 1:** Graphical representation of increasing stratigraphic distance



**Fig. 2:** Graphical representation of qualitative destandardization



**Fig. 3:** Graphical representation of internal destandardization

In the third place, destandardization may take the form of increasing variation within the highest level, regardless of whether this growing variation correlates with changes in the relationship with regard to the other level. Again, various terminological alternatives can be considered for identifying such a process: it might be called 'internal destandardization', but 'dehomogenization' or 'heterogenization' could also be considered. Figure 3 graphically represents a process of internal destandardization.

Before we proceed to a quantitative, lectometric definition of the three developments, four remarks have to be made. First, it can be shown that the three types of change identified here are mathematically independent, in the sense that all possible combinations of developments (growing or diminishing distances between layers, movement from top to bottom or from bottom to top, increasing or decreasing variation within the top level) could in principle occur. In practice, one will not expect all possible configurations to occur with equal frequency, or even to occur at all, but that would of course be a major topic for a research programme to be built on the essentials described here.

Second, the three processes have been described in terms of a simplified model with two levels, but it will be clear that the developments may be complex when a more elaborate stratificational structure is considered. Specifically, there is no reason to assume that developments within a multilevel structure will be uniform across all levels. A growing together of the standard language at the top level and colloquial speech does not necessarily imply (at a third level) dialect loss, even though Auer suggests that it is the normal historical evolution in the modern era. Again, the question is an empirical one.

Third, the behavioral changes classified above need to be supplemented with attitudinal data to get a complete picture of the situation. For instance, even if colloquialisms regularly penetrate in formal situations, they could still be considered inappropriate: norms may persist in the mind even if their force in actual behavior weakens. The remainder of this paper will be exclusively devoted to the behavioral perspective, though. Given that the attitudinal, and by extension ideological, parameter is a crucial feature of the SLICE approach, the present paper's restriction to the behavioral dimension limits its claims: it is primarily meant to bring some more analytical and descriptive clarity to the behavioral side of (de)standardization processes, but it does not intend to provide a full conceptual or methodological framework for the study of (de)standardization in all its aspects.

Fourth, how does the destandardization/demotization framework relate to the three processes that can be distinguished on systematic grounds? Demotization, with its emphasis on the relaxation of existing standard norms, is probably best conceived of in terms of the third process: more variation enters into the

standard language, but the position of the standard with regard to other levels of language use remains roughly the same. The notion of destandardization, on the other hand, seems to relate primarily to the second process in combination with the first, to the extent that higher level language use grows closer to lower level language use, though not in the bottom-up way that is expected by traditional standard language ideologies but rather in a top-down way: hierarchical differences are leveled out, but they are leveled out in favor of an initially subordinate level rather than the other way around. It remains to be seen whether this attempt at reinterpretation of the destandardization/demotization conceptual pair does full justice to the framework (to repeat, the attitudinal dimension is still lacking), but at the same time, the exercise illustrates the importance of a multidimensional approach. If demotization indeed involves leveling plus a top-down directionality, both dimensions need to be included in the description.

We may now provide lectometric definitions of the three types of destandardization that we distinguished. Let  $x_i$  be any lexical item from a set of synonyms expressing concept  $Z$ , and  $Y_1$  and  $Y_2$  two datasets subject to lectometrical analysis, then the degree of uniformity in the lexical expression of  $Z$  between  $Y_1$  and  $Y_2$  is captured by (1). Essentially, this formula defines the uniformity between  $Y_1$  and  $Y_2$  for  $Z$  as the degree of overlap in the naming events for  $Z$  in both datasets. (In corpus-based research, ‘datasets’ would typically be subcorpora that can be considered representative for specific lects. But relevant datasets could obviously also have a different origin, deriving from experimental research or surveys.) For a set of  $n$  concepts, the aggregate uniformity between  $Y_1$  and  $Y_2$  is calculated by (2) as a straightforward average, or by (3) as a weighted average, in which the weighting factor  $G$  is the relative frequency of any concept  $Z_i$  in  $Y_1$  and  $Y_2$  taken together. Formula (4) provides a measure for the internal uniformity of lexical usage in a given dataset, based on the assumption that a language situation could be considered more uniform to the extent that there are less competing forms for expressing a given concept, and to the extent that dominant forms exist within that set of alternatives. In (5) and (6), the internal uniformity measure is aggregated over a set of  $n$  concepts, respectively without and with weighting. (For further background and applications relating to this suite of formulae, see Geeraerts/Grondelaers/Bakema 1999.)

$$(1) \quad U_Z(Y_1, Y_2) = \sum_{i=1}^n \min(F_{Z, Y_1}(x_i), F_{Z, Y_2}(x_i))$$

$$(2) \quad U(Y_1, Y_2) = \frac{1}{n} \sum_{i=1}^n U_{Z_i}(Y_1, Y_2)$$

$$(3) \quad U'(Y_1, Y_2) = \sum_{i=1}^n U_{Z_i}(Y_1, Y_2) \cdot G_{Z_i}(Y_1 \cup Y_2)$$

$$(4) \quad I_Z(Y) = \sum_{i=1}^n F_{Z, Y}(x_i)^2$$

$$(5) \quad I(Y) = \frac{1}{n} \sum_{i=1}^n I_{z_i}(Y)$$

$$(6) \quad I'(Y) = \sum_{i=1}^n I_{z_i}(Y) \cdot G_{z_i}(Y)$$

The definitions of the three types of destandardization now follow in a straightforward fashion. We consider four lects, differentiated by stratificational position and chronology.  $H$  represents the stratificationally higher situation, where we may expect language use that is representative of or at least closer to standard language use (to the extent that standardization exists at all in the linguistic situation at hand), and  $L$  a lower-ranking situation. If  $t_1$  and  $t_2$  represent an earlier and a later point in time, then the three types of destandardization (in the order in which they were introduced above) are defined as follows.

(7) Destandardization<sub>1</sub> occurs

$$\text{if } U(H_{t_1}, L_{t_1}) > U(H_{t_2}, L_{t_2}) \\ \text{or if } U'(H_{t_1}, L_{t_1}) > U'(H_{t_2}, L_{t_2})$$

(8) Destandardization<sub>2</sub> occurs

$$\text{if } U(L_{t_1}, H_{t_2}) > U(H_{t_1}, L_{t_2}) \\ \text{or if } U'(L_{t_1}, H_{t_2}) > U'(H_{t_1}, L_{t_2})$$

(9) Destandardization<sub>3</sub> occurs

$$\text{if } I(H_{t_1}) > I(H_{t_2}) \\ \text{or if } I'(H_{t_1}) > I'(H_{t_2})$$

Corresponding to the three processes introduced above, the formulae will be self-evident, except perhaps in the second case. Formula (8) measures the direction of change by comparing the similarity between, on the one hand, the lower level at time  $t_1$  and the higher level at  $t_2$ , and on the other, that between the higher level at time  $t_1$  and the lower level at  $t_2$ . If the former is bigger than the latter, the attraction exerted by the originally lower level is stronger than the attraction of the higher level in the initial stage, or in other words, the change is from bottom to top rather than from top to bottom.

To illustrate the formulae and the phenomena they capture, we have a brief look at a longitudinal study on the lexical development of Dutch in the lexical field of clothing terms. Although, as we shall see, the results can be plausibly interpreted in the light of the recent evolution of Dutch, it will be clear that a single lexical field is not enough to yield general conclusions about the evolution of Dutch. We would need to know more about other parts of the vocabulary and other levels of linguistic structure for a comprehensive picture. In this sense,

the results are primarily meant to illustrate the method rather than to support far-reaching descriptive statements.

The study drawn on here is a replication of Geeraerts/Grondelaers/Speelman (1999), in which clothing terms and football terms were followed from 1950 over 1970 to 1990 in Netherlandic Dutch and Belgian Dutch sources. These sources primarily comprised supraregional written data from national newspapers and magazines, with the addition of shop window materials for the 1990 clothing terms. These ‘shop window materials’ took the form of price tags in local shops, with the exclusion of national or international chain stores. In this way, a second situational layer is added to the dataset: if naming practices differ in less formalized contexts, this is one communicative situation in which less formal usage may be found. The shop window data were collected in two Dutch and two Flemish towns with similar characteristics: the centrally located traditional university towns Leiden and Leuven, and the peripheral towns Maastricht and Kortrijk, each with a smaller university. The replication study of 2012 (see Daems/Heylen/Geeraerts 2015 for an extended description) repeated the 1990 clothing terms study, so that we now have real time data for two stratigraphic levels at two points in time – a crucial condition for applying the definitions in (7)–(9). In quantitative terms, the dataset contains 8,797 observations for Belgian Dutch in 1990, and 3,761 for 2012. For Netherlandic Dutch, the figures are 6,205 and 5,255 respectively.

The fourteen concepts included in the analysis are the following: shirt<sub>M</sub>, shirt<sub>F</sub>, t-shirt<sub>MF</sub>, sweater<sub>MF</sub>, cardigan<sub>MF</sub>, trousers<sub>MF</sub>, jeans<sub>MF</sub>, leggings<sub>F</sub>, skirt<sub>F</sub>, dress<sub>F</sub>, suit jacket<sub>M</sub>, suit jacket<sub>F</sub>, jacket<sub>MF</sub>, suit<sub>MF</sub>. The subscripts indicate whether the item of clothing is meant for women or men. This could either mean that the clothing type is gender-specific (like ‘skirt’) or that the same type receives different names when worn by men or women (as in a jacket as part of a suit, which is often called *colbert* in the case of men, but hardly ever so in the case of women). If the gender distinction does not correlate with differences of naming pattern, the concept is considered gender-neutral. The lexical alternatives involve synonyms like *jeans*, *jeansbroek*, *spijkerbroek*. Only in the case of ‘skirt’ no alternatives emerge: skirts are always called *rok* (but because we want to have an aggregate level view of the lectometric relations, concepts with little or no lexical variation have to be retained as part of the calculations). Overall, statistical significance is checked by applying a log-likelihood ratio test with a threshold of 5% to the naming patterns under comparison.

If we then collect the results for the Belgian Dutch dataset, the three types of possible destandardization captured by (7)–(9) appear as follows. Restricting the overview to non-weighted averages, the *B* figures refer to the higher level stratum of national magazine data, while the *LeuKor* figures are based on the shop window materials in Leuven and Kortrijk.

- (10) Destandardization<sub>1</sub>  
 $U(B_{90}, LeuKor_{90}) = 50.47$   
 $U(B_{12}, LeuKor_{12}) = 73.72$   
 $U(B_{90}, LeuKor_{90}) < U(B_{12}, LeuKor_{12})$
- (11) Destandardization<sub>2</sub>  
 $U(LeuKor_{90}, B_{12}) = 60.25$   
 $U(B_{90}, LeuKor_{12}) = 53.12$   
 $U(LeuKor_{90}, B_{12}) > U(B_{90}, LeuKor_{12})$
- (12) Destandardization<sub>3</sub>  
 $I(B_{90}) = 69.21$   
 $I(B_{12}) = 74.96$   
 $I(B_{90}) < I(B_{12})$

For the Netherlandic Dutch dataset, the *N* figures refer to the higher level stratum of national magazine data, while the *LeiMaa* figures are based on the shop window materials in Leiden and Maastricht.

- (13) Destandardization<sub>1</sub>  
 $U(N_{90}, LeiMaa_{90}) = 69.07$   
 $U(N_{12}, LeiMaa_{12}) = 73.62$   
 $U(N_{90}, LeiMaa_{90}) < U(N_{12}, LeiMaa_{12})$
- (14) Destandardization<sub>2</sub>  
 $U(LeiMaa_{90}, N_{12}) = 61.57$   
 $U(N_{90}, LeiMaa_{12}) = 84.93$   
 $U(LeiMaa_{90}, N_{12}) < U(N_{90}, LeiMaa_{12})$
- (15) Destandardization<sub>3</sub>  
 $I(N_{90}) = 68.48$   
 $I(N_{12}) = 71.06$   
 $I(N_{90}) < I(N_{12})$

The evolutions contained in these figures turn out to point to standardization, rather than destandardization. In both national varieties of Dutch, the distance between the stratigraphic layers diminishes and the internal uniformity of the upper layer increases. In the Netherlandic case, the directionality of the compression corresponds to a traditional conception of standardization: the lower level moves in the direction of the upper level. In the Belgian Dutch data, on the other hand, the opposite is the case, and this is the only example of ‘destandardization’ as defined above that may be found in the dataset. This destandardizing aspect

of the Belgian Dutch development needs to be understood in a broader historical context. (For more background, see Geeraerts/Van de Velde 2013 for a comprehensive view of recent developments in Netherlandic and Belgian Dutch.)

In Flanders, the standardization process that started off (as in most European countries) in the Early Modern Period was slowed down as a result of Flanders' political separation from The Netherlands during the Eighty Years' War. Standard Dutch started to develop in The Netherlands in the course of the 17th century, but as Flanders was politically separated from The Netherlands, remaining under foreign rule, it did not link up with this process of standardization. Rather, French was used more and more as the language of government and high culture, a practice that received an important impulse after the birth of the Belgian state in 1830. Dutch then survived predominantly in the form of a range of Flemish dialects. However, as a result of a social and political struggle for the emancipation of Flanders and the Dutch-speaking part of the Belgian population, Dutch again gained ground as a standard language (the language of learning, government, and high culture) in Flanders. This process started somewhat hesitantly in the late 19th century as a typically romantic movement, gained momentum during the first half of the 20th century, and finally made a major leap after World War II and during the booming 1960s. Importantly, the official linguistic policy of Belgian Dutch during this process of standardization was based on a normative dependency on Netherlandic Dutch: when the use of Dutch as a language of higher education and culture spread, the existing Netherlandic Dutch norm was officially promoted, in educational practices and elsewhere, as the model to be taken over. This linguistic policy was successful: if we look at our dataset for the evolution of  $U(B, N)$  over sixty years, we see a steady increase from 1950 over 1970 to 1990:  $U'$  figures rise from 69.21 over 77.50 to 86.50. From 1990 to 2012, however, the uniformity drops from 86.50 to 81.50. If this drop signals a growing independence of Belgian Dutch with regard to Netherlandic Dutch, then the 'destandardizing' directionality revealed in (11) makes sense. At the same time as looking away from (or at least looking less attentively at) Netherlandic Dutch as a norm to be adopted, Belgian Dutch makes more room for its own forms of linguistic usage. When all aspects of the evolution are taken into account, the 'destandardizing' change of Belgian Dutch does not signal an abandonment of the traditional model of standardization, but rather reveals that the Belgian Dutch standardization process has acquired a dynamics of its own, with more autonomy with regard to Netherlandic Dutch than used to be the case.

Even with all the caveats mentioned earlier, this case study illustrates the fruitfulness of an analytic, multidimensional perspective on so-called 'destandardization'. Distinguishing between different ways in which the evolution of stratigraphic continua may depart from the traditional model of standardization – increasing distances between layers, top-down rather than bottom-up direc-

tionality, growing variation in the top level – yields a more accurate and detailed picture of ongoing changes than a simple opposition between ‘destandardization’ and ‘demotization’ would do.

## References

- Auer, Peter (2005): Europe’s sociolinguistic unity, or: A typology of European dialect/standard constellations. In: Delbecque, Nicole/van der Auwera, Johan/Geeraerts, Dirk (eds.): *Perspectives on variation. Sociolinguistic, historical, comparative.* (= Trends in Linguistics. Studies and Monographs (TiLSM) 163). Berlin, pp. 7–42.
- Auer, Peter (2011): Dialect vs. standard: a typology of scenarios in Europe. In: Kortmann, Bernd/van der Auwera, Johan (eds.): *The languages and linguistics of Europe. A comprehensive guide.* Berlin, pp. 485–500.
- Coupland, Nikolas/Kristiansen, Tore (2011): SLICE: Critical perspectives on language (de)standardisation. In: Kristiansen, Tore/Coupland, Nikolas (eds.): *Standard languages and language standards in a changing Europe.* Oslo, pp. 11–35.
- Daems, Jocelyne/Heylen, Kris/Geeraerts, Dirk (2015): Wat dragen we vandaag: Een hemd met blazer of een shirt met jasje? Convergentie en divergentie binnen Nederlandse kledingtermen. In: *Taal en Tongval. Tijdschrift voor Taalvariatie* 67, 2, pp. 307–342.
- Geeraerts, Dirk/Grondelaers, Stefan/Bakema, Peter (1994): The structure of lexical variation. Meaning, naming, and context. (= *Cognitive Linguistics Research (CLR)* 5). Berlin.
- Geeraerts, Dirk/Grondelaers, Stefan/Speelman, Dirk (1999): *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen.* Amsterdam.
- Geeraerts, Dirk/van de Velde, Hans (2013): Supra-regional characteristics of colloquial Dutch. In: Hinskens, Frans/Taeldeman, Johan (eds.): *Language and space. Vol. 3: Dutch.* (= *Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK)* 30.3). Berlin, pp. 532–556.
- Goebel, Hans (2006): Recent advances in Salzburg dialectometry. In: *Literary and Linguistic Computing* 21, 4, pp. 411–435.
- Gooskens, Charlotte/Heeringa, Wilbert (2004): Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. In: *Language Variation and Change* 16, pp. 189–207.
- Grieve, Jack (2007): Quantitative authorship attribution: An evaluation of techniques. In: *Literary and Linguistic Computing* 22, 3, pp. 251–270.
- Heeringa, Wilbert (2004): *Measuring dialect pronunciation differences using Levenshtein distance.* PhD thesis. Groningen.
- Kristiansen, Tore (2016): Contemporary standard language change. Weakening or strengthening? In: *Taal en Tongval* 68, pp. 93–118.
- Luyckx, Kim/Daelemans, Walter (2011): The effect of author set size and data size in authorship attribution. In: *Literary and Linguistic Computing* 26, 1, pp. 35–55.
- Mattheier, Klaus J. (1997): Über Destandardisierung, Umstandardisierung und Standardisierung in modernen europäischen Standardsprachen. In: Mattheier, Klaus J./Radtke, Edgar (eds.): *Standardisierung und Destandardisierung europäischer Nationalsprachen.* (= *Variolinguaria. Nonstandard – Standard – Substandard* 1). Frankfurt a.M., pp. 1–9.

- Nerbonne, John/Kleiweg, Peter (2003): Lexical distance in LAMSAS (Linguistic Atlas of the Middle and South Atlantic States). In: *Computers and the Humanities* 37, 3, pp. 339–357.
- Ruette, Tom et al. (2014): Semantic weighting mechanisms in scalable lexical sociolectometry. In: Szmrecsanyi, Benedikt/Wälchli, Bernhard (eds.): *Aggregating dialectology, typology, and register analysis. Linguistic variation in text and speech.* (= *Linguae & Litterae* 28). Berlin, pp. 205–230.
- Ruette, Tom/Ehret, Katharina/Szmrecsanyi, Benedikt (2016): A lectometric analysis of aggregated lexical variation in written Standard English with Semantic Vector Space models. In: *International Journal of Corpus Linguistics* 21, pp. 48–79.
- Séguy, Jean (1971): La relation entre la distance spatiale et la distance lexicale. In: *Revue de Linguistique Romane (RLiR)* 35, pp. 335–357.
- Speelman, Dirk/Grondelaers, Stefan/Geeraert, Dirk (2003): Profile-based linguistic uniformity as a generic method for comparing language varieties. In: *Computers and the Humanities* 37, 3, pp. 317–337.
- Speelman, Dirk/Geeraerts, Dirk (2009): The role of concept characteristics in lexical dialectometry. In: *International Journal of Humanities and Arts Computing* 2, 1–2, pp. 221–242.
- Speelman, Dirk/Impe, Leen/Geeraerts, Dirk (2014): Phonetic distance and intelligibility in Dutch. In: da Silva, Augusto Soares (ed.): *Pluricentricity. Language variation and sociocognitive dimensions.* (= *Applications of Cognitive Linguistics (ACL)* 24). Berlin, pp. 227–241.
- Szmrecsanyi, Benedikt (2013): *Grammatical variation in British English dialects. A study in corpus-based dialectometry.* (= *Studies in English Language*). Cambridge.
- Van Bezooijen, Renée/Heeringa, Wilbert (2006): Intuitions on linguistic distance: Geographically or linguistically based? In: Koole, Tom/Nortier, Jacomine/Tahitu, Bert (eds.): *Artikelen van de vijfde sociolinguïstische conferentie.* Delft, pp. 77–87.
- Zenner, Eline/Speelman, Dirk/Geeraerts, Dirk (2012): Cognitive sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. In: *Cognitive Linguistics* 23, 4, pp. 749–792.