

Article

Maps, meanings and loanwords: The interaction of geography and semantics in lexical borrowing

Karlien Franco¹ , Dirk Geeraerts¹, Dirk Speelman¹ and Roeland van Hout²

¹Quantitative Lexicology and Variational Linguistics, University of Leuven, Leuven, Belgium and ²Radboud University Nijmegen, Nijmegen, the Netherlands

Abstract

The use of loanwords is generally attributed to a social feature, like social prestige, and to semantic features, like the need to fill a lexical gap. However, few studies take into account variation in the use of loanwords within a speech community, and directly compare the frequency of loanwords from more than one source language. This paper contributes to research on lexical borrowing by comparing the distribution of loanwords from three different source languages in two large databases of dialect data. We take an onomasiological perspective, which allows us to gauge the frequency of borrowed lexical items vis-à-vis alternative expressions. Using Generalized Additive Mixed Modeling, we show that the usage of loanwords can only be explained by taking into account the interaction between semantics and geographical diffusion. Our analysis confirms that the patterns that occur almost exclusively reflect changes in socio-cultural history.

Keywords: lexical borrowing; dialectology; dialectometry; Dutch; Generalized Additive Modeling

(Received 30 November 2017; accepted 20 November 2018; first published online 2 August 2019)

Introduction

Why and when do language users rely on loanwords? Lexical borrowing is often explained by means of an interaction between the social prestige of the language varieties involved and semantic properties of the concept to be expressed (e.g., Hock & Joseph, 1996; McMahon, 1994; Thomason & Kaufman, 1991; Winford, 2010). More specifically, the presence of loanwords in a particular receptor language is generally analyzed by referring to the necessity to express a particular concept for which no native equivalent exists (e.g., when a novel concept, like *COMPUTER*¹, is introduced), or to the influence of a particular prestigious culture on domains of everyday life (e.g., the use of French words in English after the Norman conquest). One of the results of the Loanword Typology Project (Haspelmath & Tadmor, 2009), for instance, is that the number of loanwords that are borrowed in the world's languages, differs dramatically per semantic field (Tadmor, 2009: 64–65). Table 1 shows the distribution of loanwords per semantic field across the languages included in the project. Fields that are prone to borrowing across varieties include 'religion & belief' (41.2% loanwords), 'clothing and grooming' (38.6% loanwords) and 'the house' (37.2% loanwords), while only very few instances of lexical borrowing occur in the field of 'sense perception' (11% loanwords), 'spatial relations' (14% loanwords) and 'the body' (14.2% loanwords). According to Tadmor, this has to do with the fact that the former fields are more heavily influenced by cultural interactions. The latter semantic fields consist of concepts that are not prone to borrowing as they contain a larger amount

of core vocabulary, for which native elements are already available (Swadesh, 1955).

However, research on lexical borrowing does not often comprehensively take into account lectal variation within a speech community. Although Paul (1981:698) already asserts that loanword usage can be prone to lectal variation, as the use of non-native material can be restricted to particular groups, connected by social ties or characterized by geographical proximity, an onomasiological perspective on the lectal dimension of lexical borrowing has only recently been receiving more attention (Zenner & Kristiansen, 2014). More specifically, although earlier studies have paid attention to social or geographical correlates of variation in lexical borrowing (e.g., Poplack, Sankoff & Miller, 1988; also see Kruijssen, 1995, and Weijnen, 1967, for loanwords in the dialects of Dutch), most research does not examine the distribution of loanwords from an onomasiological perspective, whereby the success of non-native variants vis-à-vis their native alternatives that express the same meaning, is taken into account (an exception is Zenner, Speelman & Geeraerts, 2012). Additionally, hardly any studies provide a comprehensive comparative account of the influence of different prestigious cultures on a particular receptor language by examining borrowings from more than one source language in a single dataset of naturalistic material (for Dutch, notable exceptions include Geeraerts, Grondelaers & Speelman, 1999, and Daems, Heylen & Geeraerts, 2015; Van der Sijs, 2005 brings together previous research on loanwords from different source languages that occur in Dutch). Consequently, although lectal and semantic features have been acknowledged as influential for the use of loanwords, and although recent approaches have argued that multifactorial, large-scale and mixed-data approaches can further inform the process of (lexical) borrowing (Zenner & Kristiansen, 2014:10), research that examines the interaction between semantics and lectal variation for different source languages at once in a systematic and quantitative way, is lacking.

Author for correspondence: Karlien Franco, Email: karlien.franco@kuleuven.be

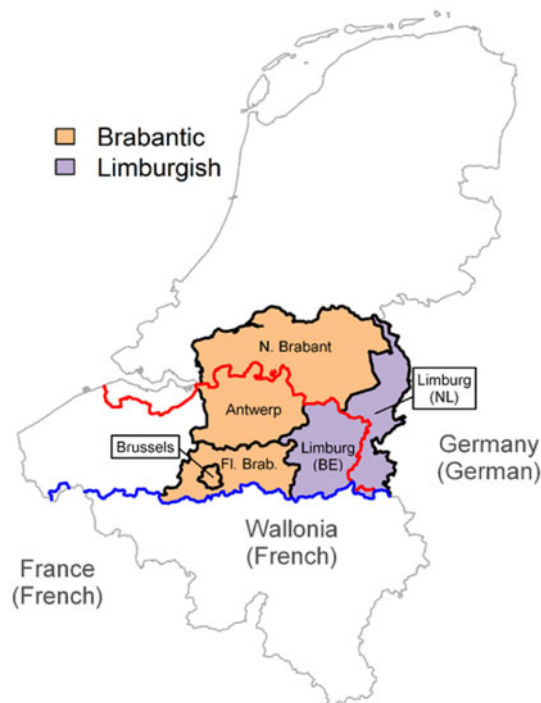
Cite this article: Franco K, Geeraerts D, Speelman D, and van Hout R. (2019) Maps, meanings and loanwords: The interaction of geography and semantics in lexical borrowing. *Journal of Linguistic Geography* 7: 14–32, <https://doi.org/10.1017/jlg.2019.2>

Table 1. Borrowing per semantic field in the Loanword Typology Project (Tadmor, 2009:64)

Semantic field	Loanwords as % of total
Religion and belief	41.2
Clothing and grooming	38.6
The house	37.2
Law	34.3
Social and political relations	31.0
Agriculture and vegetation	30.0
Food and drink	29.3
Warfare and hunting	27.9
Possession	27.1
Animals	25.5
Cognition	24.2
Basic actions and technology	23.8
Time	23.2
Speech and language	22.3
Uuantity	20.5
Emotions and values	19.9
The physical world	19.8
Motion	17.3
Kinship	15.0
The body	14.2
Spatial relations	14.0
Sense perception	11.0

This paper aims to fill this gap by focusing on the use of loanwords from three different source languages in a large dataset of dialect data. More specifically, it contributes to research on lexical borrowing in four ways. First, it ensures an onomasiological perspective. As argued in the Cognitive Sociolinguistics paradigm, an onomasiological perspective is necessary to examine which factors govern the choice of a particular lexical item, like a loanword, instead of an equivalent expression, to convey a particular meaning (Geeraerts, Kristiansen & Peirsman, 2010; Kristiansen & Dirven, 2008; Zenner & Kristiansen, 2014). Second, it explicitly acknowledges that the distribution of borrowed material may be variable within a speech community. We examine the influence of lectal differences by inquiring into the geographical distribution of the borrowed lexemes. Third, it examines the use of loanwords in different semantic fields, which allows us to gauge the extent to which the interaction between the lectal feature, geography, and semantic features influence the use of non-native material. Fourth, it uses an inferential technique, Generalized Additive Mixed Modeling, and a large dataset to gauge variation in the use of loanwords. Consequently, we can establish that the patterns that are found are robust and cannot be attributed to chance.

In practice, we analyze the distribution of loanwords from three different source languages that occur in the Brabantic and Limburgish dialects of Dutch. We rely on large databases of naturalistic and geographically stratified dialect data which form the source material for two onomasiological dialect dictionaries of Dutch, the *Woordenboek van de Brabantse Dialecten* 'Dictionary

**Fig. 1.** The Brabantic and Limburgish dialect areas.

of the Brabantic Dialects' (WBD) and the *Woordenboek van de Limburgse Dialecten* 'Dictionary of the Limburgish Dialects' (WLD). By including four different semantic fields in the analysis, we examine the interaction between semantics and geographical variation. In the following section, we first provide an overview of the relationship between the Brabantic and Limburgish dialect areas and other languages and varieties. Then, we discuss the cultural patterns that have been shown to be relevant for the use of loanwords in Standard Dutch. In section 3, the hypotheses that can be distinguished on the basis of these patterns are outlined. Section 4 describes the data and methodology used in this paper. In section 5, the results of the analyses are presented, followed by a discussion in section 6.

Lexical borrowing in the Dutch language area

Geography and cultural history

Figure 1 is a map of Belgium and the Netherlands that shows the location of the Brabantic and Limburgish dialect areas. It also shows the province borders within these dialect regions. The Brabantic dialect area (in orange) consists of the province of North Brabant (N. Brabant) in the Netherlands and of the provinces of Antwerp, Flemish Brabant (Fl. Brab.), and the Brussels area in Belgium. In the Belgian and Netherlandic provinces of Limburg (in purple), Limburgish dialects are spoken, as well as in a small area in the north of the province of Liège.

Geographically, the two dialect areas are demarcated by the state border with Germany in the east and by the Germanic-Romance border in the south (in blue). In the northern part of Belgium, above the Germanic-Romance border, Dutch is spoken, whereas south of the border, French is used. The state border between Belgium and the Netherlands runs through the dialect areas as well (in red). As a result, two processes can be distinguished that may influence variation in the use of non-native

lexical items in these dialects. First, we may expect to find border effects, as language contact between varieties that are geographically nearby can cause interference between the varieties: the closer to the border, the larger the expected number of loanwords from the nearby language. Second, previous research has shown that state borders can evolve into language borders (Hinskens, Kallen & Taeldeman, 2000). Crucially, the cultural history of the northern, Dutch-speaking part of Belgium and that of the Netherlands is not identical (an overview of the history of the Dutch language situation can be found in Janssens & Marynissen, 2008; Van der Wal & Van Bree, 2008; and Willemyns, 2013). More specifically, throughout western European history, several diglossic constellations have been relevant, in which a particular exoglossic standard exerted its influence on everyday language (Auer, 2005). However, due to the different sociopolitical history of Belgium and the Netherlands, the diglossic constellation of the northern part of Belgium has differed from the Netherlandic one.

The following paragraphs provide an overview of the expected influence of French, Latin and German on the Brabantic and Limburgish dialects. Although loanwords from other languages occur in the WBD and WLD as well, they are too infrequent to be included in the analysis, probably because the databases contain concepts concerning the everyday life of dialect speakers in the traditional agrarian society. In total, only 93 borrowed word types from other source languages occur. The number of loanword types from French is 1644, from Latin 319 and from German 589.

French

According to Van der Sijs (2005), due to a long history of contact with the French people, French loanwords are, overall, widely accepted in Standard Dutch and they occur in a variety of semantic fields, including military (e.g., *artillerie* 'artillery' and *luitenant* 'lieutenant'), the arts (e.g., *melodie* 'melody' and *gravure* 'engraving') and everyday language (e.g., *fauteuil* 'armchair' and *blouse* 'shirt'). Furthermore, French loanwords are frequently used in Dutch for concepts relating to administration and government, which can be explained by the fact that French administration and law were introduced in the Low Countries during the Napoleonic regime (1795–1813). Additionally, French was used for these purposes even longer in the northern part of Belgium, until the Flemish movement gained political ground and Dutch became the official language of politics, education, and administration in the 1930s. Daems et al. (2015) and Geeraerts et al. (1999) inquire into the use of French for a semantic field pertaining to everyday language, viz. clothing concepts (also see Van der Sijs, 2005:184). They clearly find diverging patterns between the northern part of Belgium and the Netherlands. More specifically, French occurs more frequently in Belgian Dutch than in Netherlandic Dutch. However, due to its complex relationship with the French culture, Belgian Dutch seems to react, in a purist fashion, against the abundance of French loanwords in the language, which is apparent from the decreasing number of French loanwords between the 1950s and 2012 in the field of clothing terminology. A similar defensive tendency is absent in Netherlandic Dutch.

Additionally, research on lexical borrowing has indicated that, in the Hesbaye dialects of Dutch located near the Germanic-Romance border, the distance hypothesis holds (Kruijssen, 1990): the further away from the border, the smaller the amount of borrowings from French. However, as the language border only became a political border in the 1960s, the number of French items used by a speaker is also dependent on their age and on the amount

of contact they have with francophones (Van Hout, Kruijssen & Gerritsen, 2014).

Finally, in the south of the Brabantic dialect area, the bilingual city of Brussels is located. This city, which is the capital of Belgium, was frenchified to a much larger extent than other large cities in the northern part of Belgium due to sociopolitical developments (Willemyns, 2013:24–25). Although initially French only served as the language of the nobility, the number of people who used a variety of Dutch decayed over time, in favor of the French language (De Vriendt, 2004:20–29 and 91–94).

Latin

Latin has exerted its influence on Standard Dutch in various domains over time (Van Keymeulen, 2008; Weijnen, 1967). Van der Sijs & Engelsman (2000) mention the influence of Latin on the Germanic languages during the Roman era in semantic fields like military and politics (e.g., *defensie* 'military defense' and *pijl* 'arrow'), trade (e.g., *munt* 'coin' and *kopen* 'to buy') and the names for days of the week and for the months. In medieval times, Latin was mostly important as the language of the Catholic church but it also exerted its influence on Dutch for concepts relating to education (e.g., *school* 'school' and *schrijven* 'to write'), science (e.g., *epidemie* 'epidemic' and *recept* 'recipe'), and for administration and government (e.g., *artikel* 'article' and *decreet* 'decree'). Furthermore, words from Church Latin were borrowed for novel religious concepts when the people of the Low Countries were christened (Van der Sijs, 2005:124). Semantic fields that were influenced by Latin during the Renaissance period include the field of higher education (e.g., *academie* 'academy' and *docent* 'university teacher') and administration and government (e.g., *agenda* 'calendar' and *collega* 'colleague'). Crucially, in comparison to French, the use of Latin is probably less prone to geographical variability, as Latin has predominantly been influential as a written, academic language. Political conflicts between the Germanic tribes and Roman people, who spoke a variety of Latin as their native tongue, only occurred in the Roman era.

German

According to Van der Sijs (2005:257–259; also see Weinreich, 1968:1–2), the fact that German and Dutch are closely related languages, results in a smaller amount of loanwords that are clearly German in Standard Dutch, because they are often borrowed in a "dutchified" form (e.g., *bespreken* 'to discuss' from German *besprechen*, *drukknop* 'press-stud' from German *Druckknopf*, and *warenhuis* 'department store' from German *Warenhaus*). Furthermore, although the German language area shares a border with the region where Dutch is spoken, the influence of French has always been larger, because of the great importance of French culture throughout Europe since the Middle Ages (Van der Sijs, 2005:268). In Standard Dutch, the semantic fields in which the influence of the German language and culture are clear, are trade, religion, science and warfare (Van der Sijs, 2005:274–286). Trade terminology was predominantly borrowed through trade contacts with the Hanse in the Middle Ages. As a result, the Dutch language contains Middle Low German words like *eigenwijs* 'precocious', *daalder* 'thaler' and *kroeg* 'pub'. After the Middle Ages, High German became the dominant variety. Many religious loanwords stem from after the Reformation, when the Luther Bible was translated from (High) German into Dutch, like *afvallig* 'unfaithful', *heftig* 'fierce, intense' and *slachtoffer* 'victim'. In the 19th century, German culture was influential in areas like science (e.g., *bewusteloos* 'unconscious', *psychoanalyse* 'psychoanalysis',

and *volks-etymologie* ‘folk etymology’), socialism and politics (e.g., *jeugdbeweging* ‘youth movement’, *kartel* ‘cartel’, and *autobaan* ‘motorway’) and industry (*erts* ‘ore’, *benzine* ‘petrol’, and *Fahrenheit*). Finally, German words in Dutch having to do with warfare and army are *schermutselen* ‘to skirmish’, *hamsteren* ‘to hoard’, and *concentratiekamp* ‘concentration camp’.

In the east of the province of Limburg, a border with Germany is found, which was installed at the beginning of the 19th century. This border is interesting as German and Dutch are closely related West-Germanic languages. More specifically, the Germanic and Dutch dialects historically form a continuum: some of the dialects spoken in the south of Limburg in the Netherlands can even be considered dialects of German, as they underwent the second Germanic consonant shift (viz. the Riparian dialects, see Van de Wijngaard & Keulen, 2007). Research into the effect of the border with Germany in the Kleverland dialect continuum in the north of Netherlandic Limburg has shown that it has come to serve as a social and linguistic boundary and that the dialects on each side of the border show signs of convergence with their respective standard varieties (De Vriend et al., 2008; Giesbers, 2008). The extent to which the Riparian dialects have been influenced by the language border has been less systematically researched (but see Cornelissen, 2007, for an overview of relatively recent loanwords from German).

Hypotheses

To investigate the interaction between semantics and geography on variation in the use of loanwords in the Brabantic and Limburgish dialects, we focus on four volumes (i.e., semantic fields) of the digitized databases of the WBD and WLD:

- III.1.3: clothing & personal hygiene
- III.1.4: personality & feelings
- III.3.1: society, school & education
- III.3.3: church & religion

As will be discussed in more detail below, these semantic fields were chosen because they are expected to show clear patterns of geographical and cultural variation in the use of loanwords. Furthermore, while most of the fields are prone to borrowing, according to Tadmor (2009), the field of personality & feelings takes up a special place.²

As outlined above, concepts from the semantic field of clothing & personal hygiene are part of the everyday language of a dialect speaker and are prone to lexical borrowing (38.6% borrowed items in Tadmor, 2009). Additionally, detailed research into this field has shown that the use of French loanwords is especially frequent, although clear differences between Belgian and Netherlandic Dutch occur (Daems et al., 2015; Geeraerts et al., 1999). As the dialectal data we use come from an early time period (87.5% of the clothing data in the database were collected in the 1960s) and from a different variety (viz., from the base dialects of Dutch), our data serve as a historical, differently stratified alternative to the oldest data used in Geeraerts et al. (1999).

In the semantic field of society, school & education, lexical borrowings from both French and German are expected. Table 2 shows the subdomains in this field in the WLD.³ On the one hand, it contains concepts relating to the military, politics and education, which have been argued to often be expressed with French items. Additionally, as French culture was dominant for a longer period in the northern part of Belgium, in this field, we expect to find

Table 2. Subdomains of the field of society, school & education in the WLD

Subdomain	Examples
Man and society	e.g., trade, money, property, labour, language, communication
Societal organisation	e.g., societal institutions, taxes, elections, police, law and crime, defence and war
Transportation	by road, by railway, by air, over water
Education	e.g., people in school, the school building

differences between Belgium and the Netherlands. On the other hand, trade and industry concepts, which can be related to German culture, are included as well. In as far as this field is included in Tadmor’s division, it is expected to show a relatively high number of loanwords as well (law: 34.3% loanwords; social and political relations: 31.0% loanwords; warfare and hunting: 27.9% loanwords; possession: 27.1% loanwords; speech and language: 22.3% loanwords).

The field of church & religion is chosen because, according to Tadmor (2009), this field is highly susceptible to borrowing. The use of Latin lexical borrowings is expected to be especially frequent in this field, although some German loanwords may be used as well. However, we expect to find no geographical variation for the distribution of loanwords in this field, as the concepts in the database refer to practices in the Catholic church, which were frequent throughout Limburg and Brabant (Schmeets, 2014). Additionally, many of the church-related Latin words were introduced as names for novel concepts.

Finally, we also include the semantic field of personality & feelings. Table 3 shows the subdomains of this semantic field in the WLD. On the one hand, this table contains some concepts, relating to feelings/emotions and values, that are not prone to borrowing according to Tadmor (2009, see Table 1), because they contain universal/core vocabulary concepts. On the other hand, some subsections of this semantic field, like behavioral traits or affect-sensitive concepts (e.g., concepts related to indecency or stupidity), may also require a certain degree of personal involvement. As a result, if we do find a large number of loanwords in this field, perhaps this has to do with the “need for synonyms” of the speakers, which allows them to retain the expressive force of affect-laden concepts (Weinreich, 1968:58–59). However, if we do not find that loanwords are used for personality & feelings concepts, we can provide further evidence for the universal stability of some semantic domains.

In sum, a complex interaction between semantic and geographical features is expected to influence variation in the use of French, Latin, and German loanwords in the Brabantic and

Table 3. Subdomains of the field of personality & feelings in the WLD

Subdomain	Examples
Intellectual capacity and memory	e.g., thinking, knowing, smart, dumb, to judge/to consider
Personality	e.g., (un)reliable, (in)sincere, diligent-lazy, brave-frightened, conceited(ness)
Feelings	e.g., fun, laughter, anger, sadness, disappointment
Behaviour	e.g., to behave, dominance, to (dis)obey, success-failure, (in)decency

Limburgish dialects of Dutch. First, expansional border effects are expected to show up near the border with Germany and near the Germanic-Romance language border in the semantic fields that are assumed to be prone to borrowing. Furthermore, French is also expected to be more frequent around the city of Brussels, where it holds a stronger position than in the rest of the language area. Second, cultural contact will probably show up as well, through the influence of an exoglossic standard on particular semantic fields. Such cultural effects are predominantly expected for French, especially for concepts relating to society, school & education and clothing & personal hygiene, and for Latin, in the field of church & religion, although German items may be used for concepts relating to these fields as well. However, differences in the geographical distribution between the former two exoglossic standards can also show up. In the use of Latin and German for church concepts, no geographical patterns are expected, as most of these loanwords were probably introduced as names for novel, institutionalized concepts, both in Standard Dutch and in the base dialects. For French, in contrast, we do expect geographical differences. The French culture held a stronger position in Belgium than in the Netherlands, which may result in geographical variation between the countries. These differences will probably be especially relevant in the semantic field of society, school & education, which contains institutionalized concepts relating to administration and politics. On the other hand, the importance of French culture on everyday life will most likely show up in the field of clothing & personal hygiene: we also expect to find differences between Belgium and the Netherlands here. The hypotheses are summarized per semantic field and per source language in Table 4.

Data and methodology

Measuring the amount of loanwords per location

The data used in this paper come from the digitized databases that form the source material for two large-scale dialect dictionaries of Dutch, viz. the *Woordenboek van de Brabantse dialecten* ‘Dictionary of the Brabantic Dialects’ (WBD) and the *Woordenboek van de Limburgse dialecten* ‘Dictionary of the Limburgish Dialects’ (WLD). These dictionaries are onomasiological dictionaries, which contain all the lexical dialect variants available for a large number of concepts throughout the Brabantic and Limburgish dialect areas. The databases consist of several volumes, with each volume representing one semantic field. We restrict attention to the dialectal data in the databases that were collected by means of questionnaires distributed throughout the Brabantic and Limburgish dialect area to ensure that the data were collected similarly and systematically, so that the relative frequency of loanwords vis-à-vis other lexical variants can be taken into account.⁴

Table 5 shows an example of the relevant columns of the databases, from the semantic field of clothing & personal hygiene in the WBD for the concepts ARMBAND ‘bracelet’ and BORSTROK ‘undervest’. In total, the Brabantic dataset contains 153 and 1046 observations for these concepts, respectively. The first column shows the concept for which information is available. The second column contains the lexical variant, a dutchified form of the dialect response of a respondent. In the third column, the number of the questionnaire that was used to elicit this response is provided and in the final column, the location where the lexical variant was elicited, is presented. For some locations, more than one response is available. Crucially, the databases contain loanword tags, like ‘fr.’, ‘du.’, or ‘lat.’, which were added manually by the lexicographers, to indicate whether a particular (dutchified) lexeme for a concept has a non-native origin.⁵ For instance, the word *bracelet* for the concept ARMBAND is marked as French. This allows us to calculate automatically how frequently a variant from French, Latin or German is used. Table 6 contains example loanwords from every semantic field for every source language.

We used the loanword tags in the dictionaries to automatically collect the number of native and non-native French, German and Latin tokens per location and per semantic field. For instance, when focusing on the French terms, the Latin and German lexical items are considered as native (i.e., not French). The same procedure is used for the other source languages. In a few cases, lexical items that were marked as non-native in the Limburgish data were not given a loanword tag in the Brabantic data and vice-versa.

Table 5. Example of the relevant columns from the semantic field of clothing & personal hygiene in the WBD

Concept	Lexical item (dutchified)	Source	Location
Armband ‘bracelet’	Bracelet (fr.)	N 86 (1981)	Aarschot
Armband ‘bracelet’	Bracelet (fr.)	N 86 (1981)	Aarschot
Armband ‘bracelet’	Armband	N 86 (1981)	Aarschot
Armband ‘bracelet’	Armband	N 86 (1981)	Halsteren
Armband ‘bracelet’	Bracelet (fr.)	N 86 (1981)	Landen
...
Borstrok ‘undervest’	Borstrok	N 02 (1960)	Attenhoven
Borstrok ‘undervest’	Slaaplijf	N 02 (1960)	Landen
Borstrok ‘undervest’	Borstrok	N 02 (1960)	Steenbergen
Borstrok ‘undervest’	Hemdruk	N 25 (1964)	Steenbergen
Borstrok ‘undervest’	Borstrok	N 25 (1964)	Steenbergen
Borstrok ‘undervest’	Hemdruk	N 25 (1964)	Steenbergen

Table 4. Overview of hypotheses in relation to the semantic fields

	Society, school & education	Clothing & personal hygiene	Church & religion	Personality & feelings
Expected source language(s)	- French - German	French	- Latin - German	No borrowing
Expected geographical patterns	<i>French:</i> - Belgium vs. Netherlands - Expansional pattern near Germanic-Romance border and Brussels area <i>German:</i> Expansional pattern near border with Germany	- Belgium vs. Netherlands - Expansional pattern near Germanic-Romance border and Brussels area	No geographical variation	No geographical variation

Table 6. Examples of loanwords per source language and semantic field

Source language	Society, school & education	Clothing & personal hygiene		
French	<i>coupon</i> <i>portefeuille</i>	'ticket (transportation)' 'wallet'	<i>bijou</i> <i>winterpaletot</i>	'jewel' 'warm coat'
Latin	<i>statie</i> <i>tribunal</i>	'station' 'cantonal court'	<i>stola</i> <i>stool</i>	'stole' 'bonnet of the "poffer"'
German	<i>rad</i> <i>flik</i>	'bike' 'police officer'	<i>absatz</i> <i>smuk</i>	'shoe heel' 'ornament'
Source language	Church & religion	Personality & feelings		
French	<i>medaille</i> <i>voile</i>	'scapular' 'headdress for girls during Holy Communion'	<i>bleu</i> <i>caractère</i>	'shy' 'personality'
Latin	<i>crucifix</i> <i>monstrans</i>	'crucifix' 'monstrance'	<i>permitteren</i> <i>pretentie</i>	'to rant and rave' 'pride'
German	<i>bleien venster</i> <i>dirigent</i>	'leaded window' 'choirmaster'	<i>juxig</i> <i>geschaft</i>	'comical' 'artificial, forced'

For example, the lexical variant *zich ambeteren* for ZICH VERVELEN 'to be bored' is marked as French in the Limburgish data, while it does not have a loanword tag in the WBD. To ensure maximal comparability between the dictionaries, we used an automatic tagging procedure to ensure that every variant that is labeled as French, Latin, or German in one dictionary, has the same tag in the other dictionary.

The onomasiological perspective is safeguarded because our calculation relies on the frequency of non-native variants for a set of concepts per semantic field, for which lexical variants were elicited throughout each dialect area. Additionally, for most of the locations, we only have one or two observations per concept at our disposal (mean = 1.417, sd = 0.561). Only two larger towns have, on average, more than 5 observations per concept, namely Maastricht (mean = 8.987, sd = 5.880) and Tilburg (mean = 5.053, sd = 4.935).

Table 7 provides an overview of the total number of French, Latin and German tokens in the semantic fields that were included per dictionary, and the proportion of these non-native tokens per dictionary. Clearly, the overall proportion of French is much higher than the proportion of Latin and German. Interestingly, the proportion of French is also almost identical in the

Table 7. Absolute and relative number of French, Latin and German tokens per dictionary

	WBD	WLD
French	16443 (0.051)	13015 (0.059)
Not French	305848 (0.949)	208353 (0.941)
Latin	4361 (0.014)	5810 (0.026)
Not Latin	317930 (0.986)	215558 (0.974)
German	318 (0.001)	2317 (0.010)
Not German	321973 (0.999)	219051 (0.990)

Brabantic and Limburgish data, while both Latin and German occur more frequently in Limburg. The full datasets, which show the amount of French, Latin and German per location, are available in the supplementary materials.⁶

Generalized additive modeling

To measure the effect of the interaction between semantic field and geography on variation in the amount of French, Latin, or German per location, we use Generalized Additive Mixed Modeling (GAMM). These models can be considered an extension of Generalized Linear Models that allow for a combination of parametric and non-parametric relationships, which do not have to be specified *a priori*, between the response and the explanatory variables (Wood, 2006; see Zuur et al., 2009, and Crawley, 2007, chapter 18 for an accessible introduction). More specifically, they employ non-parametric smooth functions on specified model terms as part of the model fitting procedure. The models we discuss below use thin-plate regression splines to represent these smooth terms. The amount of smoothing depends on a type of cross-validation, which in practice entails that the model finds the optimal amount of smoothing while avoiding badness of fit.

All the analyses were carried out with R (R Core Team, 2018). We build one model per source language. For each source language, we start from the same model to compare the influence of the interaction between geography and semantic field on variation in the number of loanwords per location. The response variable is the ratio between the number of French, Latin, or German tokens and the number of native tokens per location. The model contains a smooth term for the interaction between longitude⁷ and latitude for each semantic field, and a random intercept for location, as the total number of observations differs per location (although this factor does not reach significance in the model for the Latin variants). In our model fitting procedure, we follow the suggestions of Crawley (2007:chapter 19), Van Rij (2015), Wieling (2017, 2018) and Wood (2006:221–233) and outlined in the mgcv vignette (Wood, 2017). We compare AIC values and use significance tests to check whether all the predictor variables, interaction effects and smooth terms contribute to the explanatory power of the models. Finally, we visualize the predicted and the fitted values, and the residuals to assess the fit of the model to the data.

The R-code that was used to conduct the analyses and the datasets are available in the supplementary materials.

Results

Figure 2a–c are maps of the Limburgish and Brabantic dialect areas that show the distribution of the proportion of borrowed lexical items in the raw data per source language in the form of bubble plots.⁸ The plots are interactive: clicking on a black or red dot reveals the precise proportion of French, Latin, or German tokens in the specific location. The size of the black symbols is proportionate to the variable under scrutiny (viz. the proportion of French, Latin, or German tokens per location). Black dots indicate that one or more loanwords were found. The larger the black dot, the more loanwords occur in that location. If a red dot is shown, this means that, while data for this location are available in the dictionaries, they do not contain any non-native tokens.

Figure 2a–c indicate that, overall, French loanwords are much more frequent than lexical items from the other source languages. Furthermore, the figures show clear geographical patterns. French is more frequent near the Germanic-Romance language border in Belgium, and in the Dutch-speaking part of Belgium in general.



Fig. 2. Proportion of French (a), Latin (b) and German (c) tokens per location.

German occurs the most near the border with Germany, and specifically in the south of Limburg in the Netherlands. Loanwords from this language are especially frequent in three locations near the Dutch-German border (viz. in Simpelveld, Vaals, and Kerkrade, which all belong to the Ripuarian dialect area). Unexpectedly, the distribution of the Latin tokens does show a geographical pattern: they occur the most in the Limburgish provinces, especially in the Netherlands. The following sections aim to explain the patterns of variation in these bubble plots on the basis of the interaction between geography and semantics.

French loanwords

Using the formula described above, with an interaction between longitude and latitude by semantic field and a random effect for location, we constructed a Generalized Additive Mixed Model.

All the variables that were included in the model reach significance (the numerical output of the GAMM is included in Appendix 1A). Overall, the model performs well. It explains 92% of the null deviance. Adjusted R^2 , a value that ranges from 0 to 1 and another estimate for the amount of variation explained, is high as well: 0.908. We also used diagnostic plots to verify that the assumptions of the model were met.⁹

Figure 3a–d show the visual output of the GAMM, with the predicted surface for each semantic field presented in a separate panel. In each panel, the Brabantic and Limburgish dialect areas are depicted, with province and country borders indicated in black. A continuous color scale is plotted over this geographical area, with yellow hues indicating that the ratio of French to non-French tokens is high and red hues indicating that the number of French tokens is lower. In areas where the predicted number of French tokens is smaller than 0.03 (the lower bound of the

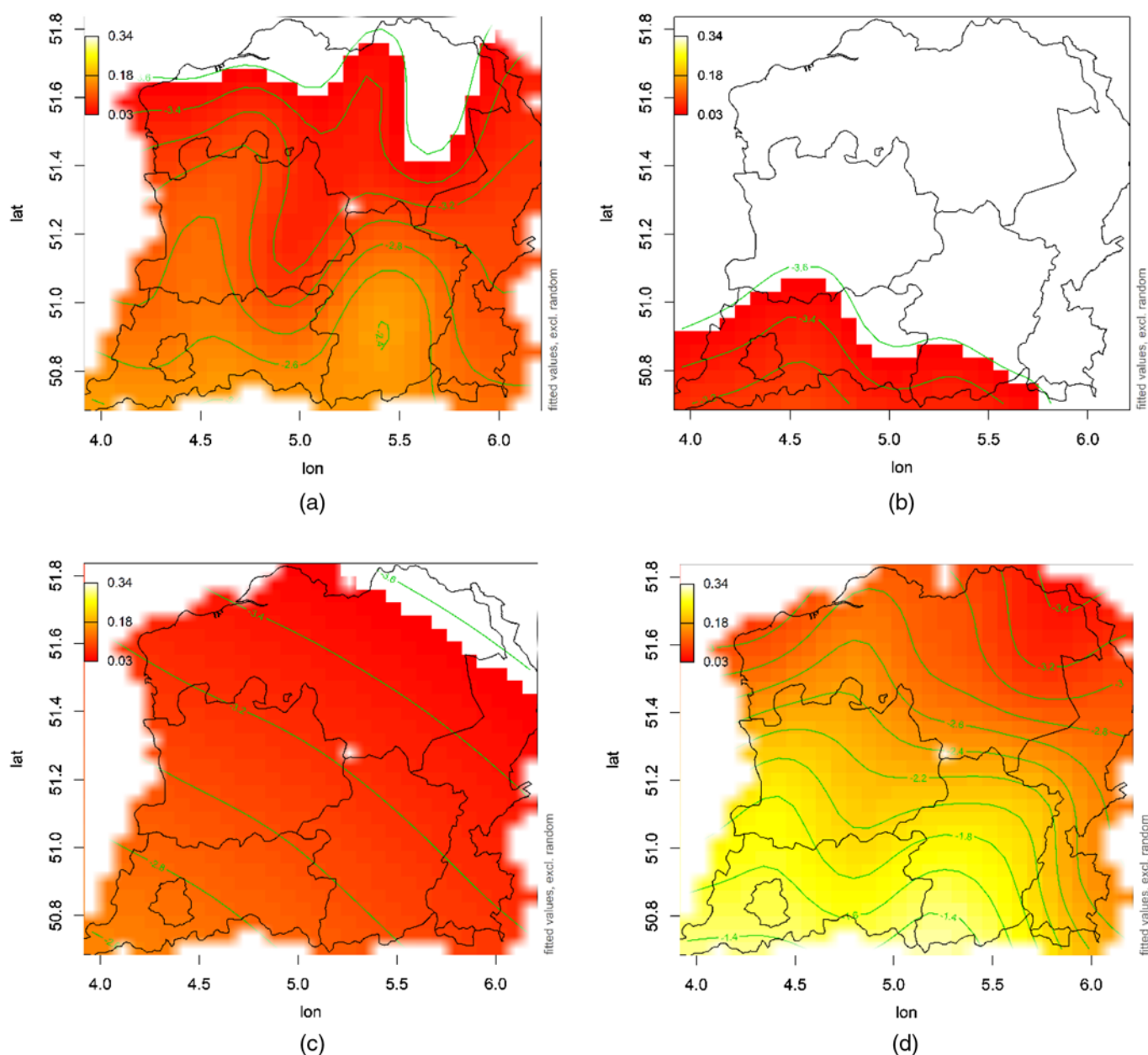


Fig. 3. French tokens per location per semantic field. (a) society, school & education. (b) personality & feelings. (c) church & religion. (d) clothing & personal hygiene.

continuous color scale), the plots show no color. The colors used in this figure, as well as in the other figures in this paper, require some further attention. First, it is important to note that stronger, darker color hues (the reddish ones) indicate a smaller number of non-native tokens. Additionally, white regions can indicate that no borrowed lexemes are available; in this case, the white areas are demarcated from the rest of the plot with non-smooth, jagged boundaries (like in Fig. 3a–c). However, very pale hues of yellow may resemble white as well, although these light hues indicate that the number of non-native tokens is very large. Crucially, in the latter case, smooth transitions rather than discontinuous borders are shown on the plots (like at the bottom of Fig. 3d). The numerical interpretation of the color scheme is provided in the legend at the top left of each panel. The legends and color schemes are kept stable for each map per source language to ensure comparability across semantic fields. The minimum and maximum values for the legend are based on the predicted values for the semantic field where French occurs the most, *viz.* clothing & personal hygiene. Additionally, the plots also show a number of green lines that run throughout the dialect areas. These lines can be interpreted as isoglosses.

The figures confirm that, as expected, the number of French tokens is very high in the semantic field of clothing & personal hygiene. French occurs even more often in this field than in the field of society, school & education. This is surprising, because military, politics and education-related terms have often been mentioned as prime candidates for the use of French loanwords. Additionally, the geographical patterns on the maps for society, school & education, on the one hand, and clothing & personal hygiene, on the other, are not identical, although both maps show clear patterns of geographical diffusion. For society-related concepts, the amount of French is high in Belgium, and in the province of Limburg in the Netherlands. The map for clothing terminology clearly demarcates Belgium from all the Netherlandic provinces. French tokens are less frequent in the field of church & religion. The isoglosses, which do not show a large amount of smoothing in this field, seem to indicate that the larger the geographical distance from the city of Brussels, where French has always held a strong position, the smaller the predicted amount of French. The map for the semantic field ‘personality & feelings’ also shows the effect of the bilingual city of Brussels. Additionally, a language border effect seems to show up near the border with Wallonia, the Germanic-Romance border. However, overall, the amount of French is very small in this field.¹⁰

In sum, two types of diffusion patterns are apparent. On the one hand, the country where a dialect speaker lives, impacts their usage of French loanwords to a large extent: French is used much more frequently in Belgium than in the Netherlands. The effect is so strong that the isoglosses on the map for the clothing concepts even seem to follow the state border between Belgium and the Netherlands. Additionally, the further away from the Germanic-Romance border and from the bilingual city of Brussels, the smaller the amount of French that is used. The maps, thus, seem to confirm a wave pattern of loanword diffusion, which is further reinforced by repeated exposure to French within the Belgian state border.

Latin loanwords

To determine the influence of semantic field and geography on variation in the use of loanwords from Latin, the same model formula was used, with an interaction between longitude and latitude by semantic field and a random effect for location. However, the

random effect for location did not reach significance and was therefore removed from the model (see Appendix 1B). Additionally, the smooth term for clothing & personality did not differ significantly from 0 ($p < 0.1$). The model performs very well: 94.4% of the null deviance is explained and adjusted R^2 is very high as well (0.957). We also verified the assumptions of the model. Although the model seems to struggle to a certain extent with the large differences in the amount of smoothing needed per semantic field, the results presented here are robust for models in which different numbers of basis functions are allowed for the calculation of the smooth term

Figure 4a–d present the visual output of the model for the amount of Latin per location. In these graphs, the color scale ranges from the odds of encountering a Latin token equal to 0.01 (red hues) to 0.1 (yellow hues). The highest predicted value for the odds of encountering a Latin loanword is 0.09, which indicates that these borrowings are clearly less frequent overall than lexical borrowings from French. The maps for the semantic fields personality & feelings and clothing & personal hygiene do not contain any color or isoglosses. This indicates that the predicted odds of encountering a Latin token in a location in these fields is even smaller than 0.01: the maximum predicted value for clothing concepts is 0.004 (in Deurne, province of Antwerp) and 0.002 for personality-related terms (in Vaals, province of Limburg in the Netherlands). For concepts from the field of society, school & education, Latin tokens occur in some locations, albeit very infrequently. Overall, Latin seems to be used the most for concepts of this field in the Netherlands, although the pattern seems to indicate that these tokens are geographically almost randomly distributed. Only a few locations in the north of the Belgian provinces show predicted odds between 0.01 and 0.06.

The field of church & religion shows, as expected, the largest number of Latin tokens. As outlined above, many of the Latin names were introduced into Dutch as names for novel concepts. Whether or not a lexical item is borrowed out of necessity (*i.e.*, to avoid a lexical gap) or not is frequently mentioned as a semantic factor that increases the borrowability of a lexeme. As a result, this factor may serve as an explanation for the success of the Latin source language in this semantic field. In contrast with the maps for the French loanwords, we do not find that the geographical distribution of the Latin loanwords forms an expansional pattern, like the wave pattern that was discussed above. However, although we did not expect to find geographical patterns in the spread of these variants for religion-related concepts, the results from the GAM indicate that Latin is used more in the two provinces of Limburg than in the Brabant dialect area. On the one hand, the picture, thus, may reflect the boundaries between the two dictionaries that are combined in the data. It may be the case that the Latin loanwords were more consistently tagged in the WLD than in the WBD, resulting in a seemingly larger number of Latin *vis-à-vis* other items in Limburg. However, as outlined above, we controlled for differences between the sources by labeling word forms that were marked in one dictionary as French, Latin, or German, as non-native in the other dictionary as well (and vice versa). As a result, another interpretation seems more likely.

More specifically, this spatial distribution can also reflect cultural patterns: perhaps people in Limburg are more oriented towards the Roman Catholic church than the Brabant dialect users and, as a result, are more familiar with the traditional Latin names. Perhaps the Limburgish dialect speakers find it more important to retain these traditional Latin names for the church concepts than the people from Brabant, who also rely on other

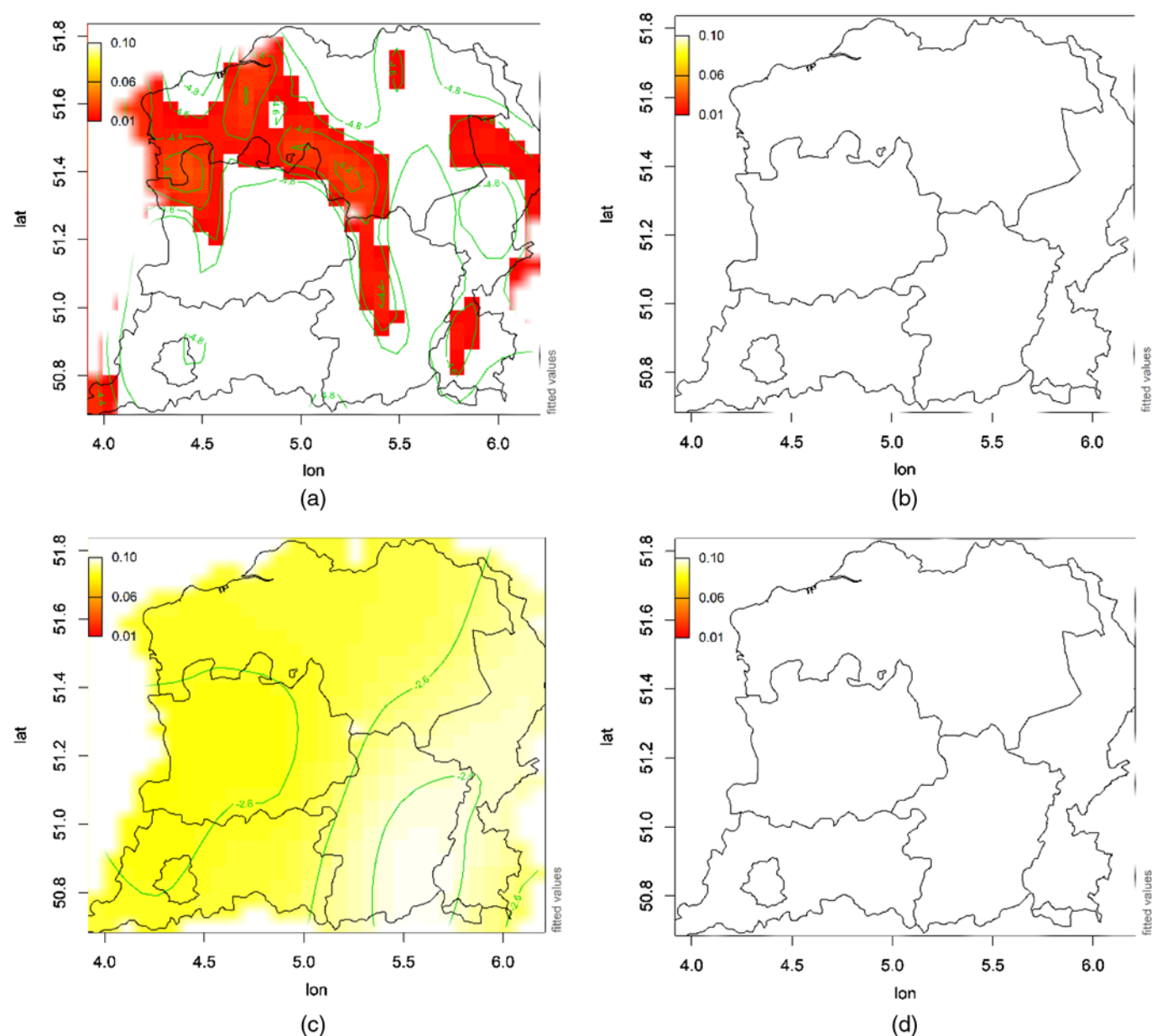


Fig. 4. Latin tokens per location per semantic field. (a) society, school & education. (b) personality & feelings. (c) church & religion. (d) clothing & personal hygiene.

variants. There is some evidence which suggests that this interpretation holds. First, Limburg in the Netherlands has a higher density of pilgrimage locations that were installed in the last 200 years (Margry & Caspers, 2000:11–12). Second, in Belgium, the percentage of catholic baptisms, weddings, and funerals and of people who attend church on Sundays in 2009 is smaller in the central cities, which are mostly located in the Brabantic dialect region, than on the countryside, which includes Limburg (Havermans & Hooge, 2011). Third, the cultural differences between the Brabantic and Limburgish dialect region are also corroborated by self-reported census data collected in the Netherlands between 1849 and 2013 (Schmeets, 2014). Figure 5 shows the percentage of Catholics in the provinces of Limburg (in the Limburgish dialect area) and North Brabant (in the Brabantic dialect area) in the Netherlands (Schmeets, 2014:6). The Figure shows that the percentage of Catholics has been decreasing more quickly in North-Brabant than in Limburg. Since the 1980s, the gap between North Brabant and

Limburg has become particularly large. Crucially, most of the dialect data for the field of church & religion in both dictionaries were only collected in the late 1980s.

German loanwords

For the German loanwords, we also used the same formula to determine the influence of semantic field and geography on the variation in the data, with an interaction between longitude and latitude by semantic field and a random effect for location (the numerical output of the GAMM is included in Appendix 1C). We verified the significance of these predictors and the assumptions of the regression model. Although the model diagnostics show that the model struggles somewhat with the general infrequency of German tokens in the dialect data, overall, it performs well. It explains 89.2% of the variation in the German versus non-German tokens (adjusted $R^2 = 0.928$). In Fig. 6a–d, the predictions

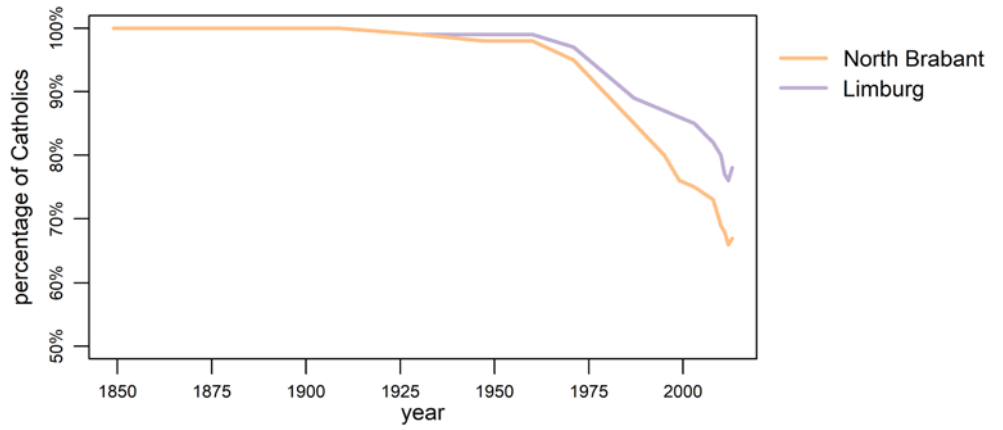


Fig. 5. Percentage of (self-reported) Catholics in the provinces of North Brabant and Limburg in the Netherlands from 1849 until 2013 (Schmeets, 2014:6).

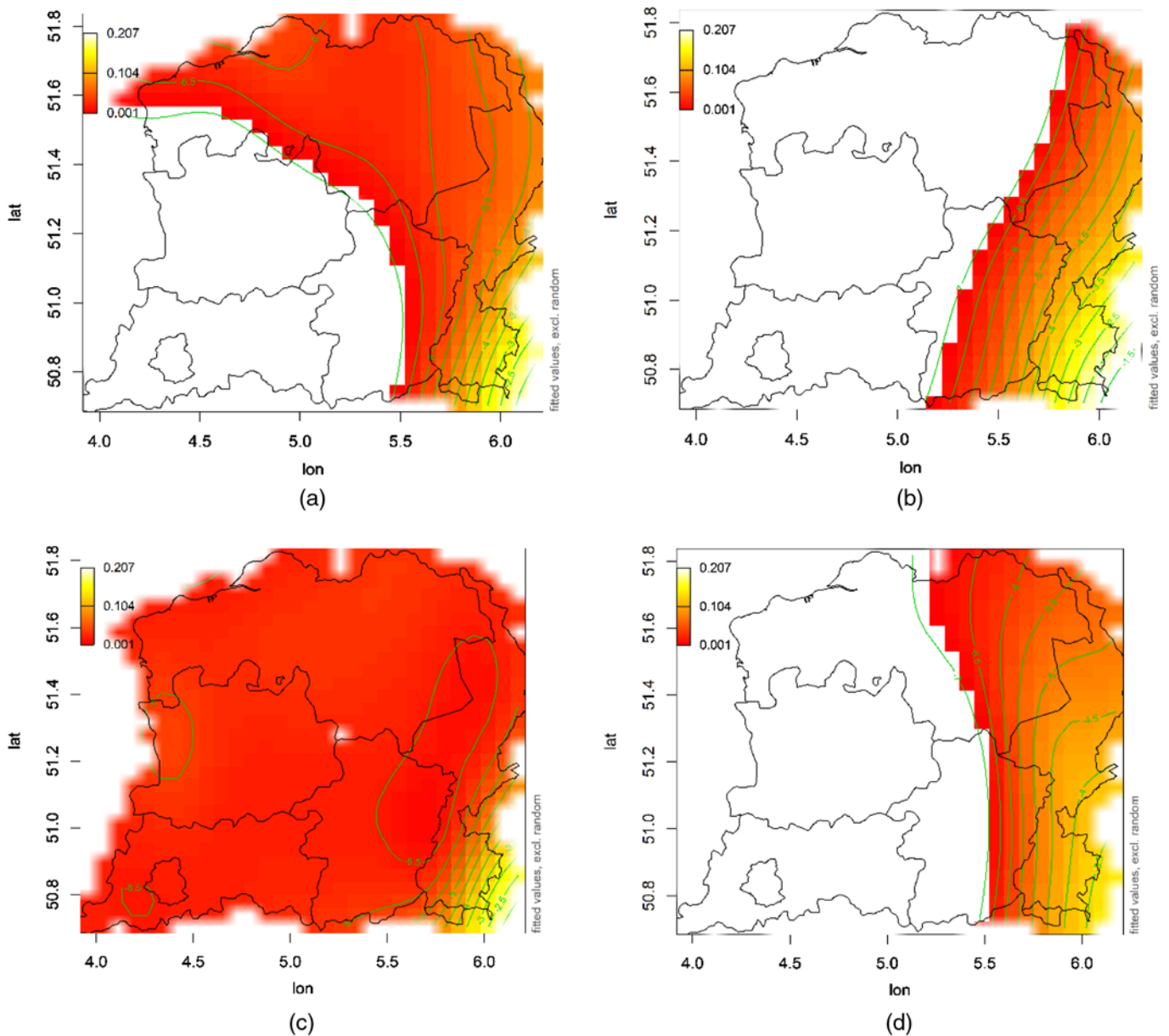


Fig. 6. German tokens per location per semantic field. (a) society, school & education. (b) personality & feelings. (c) church & religion. (d) clothing & personal hygiene.

of the GAMM are presented visually. As for a large number of locations the predicted odds of encountering a German loanword are equal to 0, the lower bound of the maps is set to a very small number (0.001). The upper bound is equal to the maximum of the predicted odds for German (0.207), to ensure comparability between the different semantic fields.

Figure 6 shows that for German, border effects are again present in the data. More specifically, in the semantic field of society, school & education, the GAMM clearly shows that there is a difference between Belgium and the Netherlands: while no German is predicted in Belgium, it occurs much more in the Netherlands, especially near the border with Germany. Interestingly, as section 4.1 showed, French tokens occur more frequently in this field in Belgium. As a result, the use of German and French may be distributed complementarily: while dialect speakers from the Netherlands use German tokens, Belgian dialect users rely on French. This is not surprising as French culture held a stronger position for a longer period in Belgium than in the Netherlands. However, it should be noted that the odds of encountering a German token in the Netherlands are much lower than the odds of finding a French token in this field in Belgium, so the Dutch dialect speakers rely on other naming strategies for society-related concepts as well. In the other semantic fields, expansional patterns show up as well, although the difference between Belgium and the Netherlands is no longer visible. More specifically, the use of German tokens is geographically the most widespread in the semantic field of church & religion. While, unsurprisingly, German tokens occur in the entire Brabant and Limburgish dialect area, they are especially frequent closer to the border with Germany. In the semantic fields of personality & feelings and clothing & personal hygiene, the amount of German that is used also becomes smaller when dialect speakers live further away from the border with Germany.

Interestingly, the border effect in three of these semantic fields (viz. society, school & education, personality & feelings and church & religion) is not only related to geographical closeness to Germany, but it also stems from a small area in the south-east of Limburg in the Netherlands. In fact, the green lines on each of these maps, which can be interpreted like isoglosses, seem to demarcate a small area, where the use of German is exceptionally high, from the rest of Limburg in the Netherlands (also see the bubble plots in Fig. 2). Table 8 shows the five locations with the highest proportion of German tokens per semantic field.¹¹ Notably, locations belonging to three municipalities take up a high position in every semantic field: Kerkrade, Simpelveld, and Vaals. Importantly, in these locations, east of the Benratherlinie (the *machen/maken* line), Ripuarian dialects are spoken (Van de Wijngaard, 2007; Van de Wijngaard & Keulen, 2007). These dialects differ from the other Limburgish dialects due to the fact that they did undergo the second Germanic (High German) consonant shift.

Traditionally and until the beginning of the 20th century, the locations belonging to the Ripuarian dialect region in the Netherlands were oriented towards Aachen. As a result, it is not surprising that the use of German in these locations is high for the semantic fields society, school & education and church & religion, two fields that are prone to borrowing. However, Table 8 indicates that dialect speakers of this region also rely on loanwords related to the field of personality & feelings, which contains concepts that are generally thought to be universal and, thus, not prone to borrowing. In fact, in these locations, the proportion of German tokens for this semantic field is higher than the proportion of German for other semantic fields (it reaches an observed value

Table 8. Locations with the highest proportion of German in three semantic fields

Semantic field	Location (municipality)	Number of German tokens	Number of non-German tokens	Proportion of German tokens
Society, school & education	Amstenrade (Schinnen)	4	78	0.049
	Nieuwenhagen (Landgraaf)	34	595	0.054
	Mechelen (Gulpen-Wittem)	16	253	0.059
	Kerkrade (Kerkrade)	56	380	0.128
	Vaals (Vaals)	16	105	0.132
Personality & feelings	Eys (Gulpen-Wittem)	93	827	0.101
	Nieuwenhagen (Landgraaf)	71	617	0.103
	Kerkrade (Kerkrade)	68	308	0.181
	Vaals (Vaals)	68	244	0.218
	Simpelveld (Simpelveld)	21	74	0.221
Church & religion	Waubach (Landgraaf)	28	1005	0.027
	Chèvreumont (Kerkrade)	4	91	0.042
	Bocholtz (Simpelveld)	22	384	0.054
	Montzen (Montzen)	40	580	0.065
	Kerkrade (Kerkrade)	31	415	0.070

of 0.18 or more). Two interpretations for this finding can be envisaged. On the one hand, the differences between the semantic fields may reflect an older dialect situation. More specifically, if we assume that universal concepts, like those of the field of personality & feelings are not prone to borrowing (in this case from Standard Dutch) and, thus, to language change at large, it may be the case that Ripuarian dialect speakers still use the old Ripuarian words, which happen to be part of Standard German as well, as they are not marked in the dictionary as typically Ripuarian lexemes. Recall that for the French loanwords, we found a similar pattern (more French for personality-related concepts near the border with Wallonia). This may serve as evidence for the fact that the use of foreign material for personality-concepts in contact situations actually reflects bilingualism of the speakers. On the other hand, the concepts from the field of personality & feelings show a large amount of variation in general. Perhaps these concepts are prime candidates for geographical variability (in this case: the use of German loanwords in a relatively limited area) because they are highly expressive. Weinreich (1968:58–59), for instance, argues that affect-laden concepts quickly lose their expressive meaning, which makes them prone to borrowing as language users need to be able to convey this expressive meaning. As a result, it may be the case that the dialect users rely on the German loanwords for this reason.

Tentative evidence for the second explanation comes from the fact that not every concept belonging to the field of personality & feelings for which data is available in more than one of the Ripuarian dialect locations (*viz.* Kerkrade, Vaals and Simpelveld), is expressed with the same German word. More specifically, for 21 out of the 94 concepts for which a German type is used in the Ripuarian dialect area, more than one German type occurs (see Appendix 2), which contains the data for the “German”¹² concepts that occur in at least two locations in the Ripuarian dialect area in the semantic field of personality & feelings). Furthermore, the mean proportion of German tokens per concept from this field in these locations is only 0.541, which means that only half of the “German” concepts are expressed with a German token in more than one Ripuarian dialect location in the Netherlands. A BRAT (SNOTNEUS), for instance, is named a *vorwitzig* (German) in Vaals, while it is called a *muilenjan*, *snotnaas*, *kute-naas* or *kute-nelis* in Kerkrade. However, for five (not very expressive) concepts out of the 94 (*viz.* DECENT (*gründlich*), TO FORCE (*zwingen*), SIMPLE (*einfach*), SOBER (*einfach*) and CHASTE (*anständig*)), one German word type does seem conventionalized to some extent, as it is used in more than one Ripuarian town.

In conclusion, most of the German words for personality & feelings concepts, are not highly entrenched and conventionalized, which makes it less likely that they stem from an older period. For only five concepts, one German word type occurs in more than one Ripuarian location. As a result, for most of the German tokens in this semantic field, the second explanation outlined above seems like the most likely one: personality & feelings concepts can be highly expressive, which results in dialect speakers relying on loanwords to convey extra (social) meaning. However, for the five concepts that do show a high amount of conventionalization, the German type may reflect an older language situation. Additional research is necessary to corroborate these explanations further.

Discussion and conclusion

The main aim of this paper was to analyze the distribution of loanwords from different source languages in the Brabant and Limburgish dialects of Dutch. We used Generalized Additive Modeling to ensure that predictors with a non-linear relationship to the response variable could be included in the analysis as well. This methodology allowed us to show that the number of loanwords is highly dependent on the interaction between semantics and geography. On the one hand, we find clear differences between the source languages. Lexical borrowings from French occur the most, while German is infrequent overall. This reflects the large influence of French culture throughout history. On the other hand, clear and systematic patterns of variation in the use of loanwords from a single source language are significant as well. More specifically, these patterns reveal different types of spatial diffusion of lexical variants through language contact. As they often differ between semantic fields, they confirm that historical and sociocultural differences characterize the use of loanwords in the Brabant and Limburgish dialect area.

As predicted in Table 4, French was found to be especially frequent in the field of clothing terminology and, to a lesser extent, also in the field of society, school & education. Interestingly, the geographical pattern is not the same in the two semantic fields (Fig. 3): the map for clothing & personal hygiene clearly shows that French is much more frequent for these concepts in Belgium than in the Netherlands. In the field of society, school & education, French additionally also occurs relatively frequently in the province of Limburg in the Netherlands.

Latin loanwords are, as expected, especially dominant in the field of church & religion. This is not surprising as many of these words were introduced as necessary loanwords: as novel concepts enter the language, the original (non-native) name for the concept is borrowed as well. However, in contrast with what we predicted in Table 4, we did find geographical differences in the spread of the Latin variants. These differences were explained as the result of a cultural difference between the Limburgish and Brabant dialect area. More specifically, it may be the case that the dialect speakers from the Limburgish dialect area are more oriented towards the Roman Catholic tradition than people from the Brabant region, which is reflected in their more systematic use of the Latin variant.

German loanwords occur throughout the dialect areas in the semantic field of church & religion. Additionally, two interesting geographical patterns show up in this and other semantic fields as well. First, for concepts from the field of society, school & education, German is only used in the Netherlands. Interestingly, we found the opposite pattern for the loanwords of French, which are more frequent in Belgium. This may indicate that French and German are complementarily distributed in this semantic field, which can be related to the fact that French culture held a stronger position in Belgium than in the Netherlands. Second, in three out of the four GAMM maps for German, the Ripuarian dialect area is clearly demarcated. In this region, German tokens are always more frequent, even in the semantic field of personality & feelings. Thus, the expansional patterns that we predicted in Table 4 show up in more semantic fields than expected. A small-scale analysis of the systematicity in the use of German tokens in this region for personality-related concepts, revealed that the use of German is not highly systematic. Only for five concepts is a single German word type used in every Ripuarian location. As most of the German words in the Ripuarian region are, therefore, not highly conventionalized and as many of these personality-related concepts are relatively expressive, it is possible that people living close to the German border use these words to convey extra social meaning.

Geographically, the GAM(M)s, thus, revealed different types of spatial diffusion. First, for French and German loanwords, expansional patterns show up. The closer to the Brussels area and to the Germanic-Romance border a dialect speaker lives, the more French loanwords (s)he uses. The closer to the Ripuarian dialect area and to the state border with Germany, the more German occurs. In other dialectometric research, expansional patterns have been observed as well, that are related to geographical closeness with a prestigious geographical region. Wieling, Nerbonne & Baayen (2011), for instance, show that the geographically farther away from the economic and politically dominant region a dialect is located, the larger the pronunciation difference with Standard Dutch. A second spatial pattern that was apparent concerns the effect of the state border, again for French and German. In the northern part of Belgium, much more French is used for clothing terms and for society-related terminology and in the former semantic field, the isoglosses even seem to follow the border between Belgium and the Netherlands (Fig. 3d). German also occurs much more frequently in the Netherlands than in Belgium for society-related concepts. These within-state patterns have probably also come about through more intensive contact with speakers of the particular source language. Crucially, the spatial patterns for loanwords from Latin are of a completely different nature. We do not find any type of expansional or within-state spatial distribution. If geographical variation is present, like in the semantic field of church & religion, it can only be related to cultural differences between the dialect areas, not to a higher degree of contact with the source language.

On the basis of these findings, we can distinguish general implications for the borrowability of lexical material. First, in every semantic field and for every source language, systematic patterns show up that correlate to a large extent with historical evolution and the sociocultural environment of a dialect user. As has been noted frequently in previous research (e.g., Backus, 2014), loanword usage reflects cultural contact. This is apparent from the different types of geographical patterns that show up. Most of these patterns can only be explained by taking into account the sociocultural history of the dialect speakers. The influence of an exoglossic variety shows up in differences concerning the use of French between the two countries where Dutch is spoken, and between the Limburgish and Brabantic language area in the use of Latin. Additionally, lexical items are also borrowed directly into the base dialects, often due to a small geographical distance to the speakers of the source language. More specifically, the border effects that were distinguished for German in the Ripuarian dialect area and for French at the Germanic-Romance language border and in the Brussels region, indicate that the distribution of loanwords in a dialect area is not only dependent on culture, but also reflects geographical closeness.

In sum, the full system of loanword usage only becomes clear when the complex interaction between culture and geography is taken into account. This is in line with recent work on lexical borrowing which advocates a multifactorial approach to lexical borrowing (Zenner & Kristiansen, 2014:8–10). Another implication for the borrowability of linguistic data is apparent from the comparison of the distribution of lexical items in the field of personality & feelings with the other semantic fields. We provide further evidence for the fact that more universal concepts are less prone to borrowing. These concepts are hardly ever expressed with non-native lexical items, except in regions that have a higher degree of bilingualism, like the Ripuarian dialect area. Finally, some scholars have argued that loanwords are copied easier from a less closely related variety (e.g., Weinreich, 1968:1–2), while lexical borrowings from a genetically close language are phonologically adapted to the language system more easily (Van der Sijs, 2005:257). We can only answer this question tentatively, because both French and Latin culture were more important for people from Belgium and the Netherlands than German culture. The data indicate that the proportion of loanwords from Latin and, especially, from French is much higher than the loanwords from German. As a result, the data corroborate the observation that the number of lexical items that are borrowed from a closely related language in their original form is smaller.

However, a shortcoming of this study is that we did not take into account characteristics of the concepts themselves. Previous research has shown that properties like the sensitivity to affect of a particular meaning may influence the amount of variation that is found (Franco et al., 2019; also see Weinreich, 1968:58–59). Additionally, it may be necessary to take into account the period when a particular concept or artifact was introduced into the Dutch culture and language. For Latin, for instance, Van der Sijs (2005) provides a list of loanwords that were already borrowed in the Romance era, like *defensie* ‘military defense’, and *munten* ‘coin’. These older lexical items, which have been present in the dialects for a longer time, are probably more conventionalized and, thus, probably, more widespread. As a result, it is possible that the concepts for which these types of lexemes are used, show less lexical variation in general. Furthermore, microlevel geographical patterns in the distribution of separate variants probably differ as a result of cultural or political changes as well. Consequently, taking into

account a factor like the age of the lexeme or concept, or comparing these relatively recent dialect data to material from an older time period, would elucidate the importance of diachronic evolution further. Further, we only focused on the use of borrowed material, but follow-up research can be envisaged that investigates whether other naming strategies also differ as a function of historical or sociocultural factors. Taking into account this type of variation can offer more insight into the question of how a particular lexical item becomes entrenched: if several options are available, why are some concepts expressed with loanwords in one location, while language users from a different place rely on names that are, for instance, based on a property of the referent itself, or on hyperonymic variants? Such research could, for instance, provide additional explanations for the use of German tokens in the Ripuarian dialect area. Finally, in this study, we did not pay attention to the context in which a particular lexical item occurs. More specifically, in the analyses, we relied on data collected by means of large-scale dialect questionnaires, but it may be the case that discursive features, like register or speech partner, also influence the variants that are used. Consequently, the geographical patterns that are found should be examined in differently stratified data as well.

However, overall, in this paper we were able to show that the use of loanwords varies as a function of geography and semantic field. By taking an onomasiological perspective to variation in loanword usage and by using sophisticated quantitative techniques and a semantically diverse dataset, we demonstrated that the patterns that occur, almost exclusively reflect changes in socio-cultural history.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/jlg.2019.2>.

Financial support. This research was supported by the Research Foundation Flanders (FWO, grant no. G042114N).

Notes

- 1 In this paper, we use SMALLCAPS to refer to concepts. *Italics* are used for the words to express these concepts.
- 2 It should be noted that the semantic fields distinguished by Haspelmath and Tadmor are not identical to the semantic fields in the WLD and WBD, nor do they contain exactly the same concepts. However, as both were collected on a very large scale, we assume that the general patterns are comparable.
- 3 The subdivision into subdomains is almost identical in the WLD and WBD across dictionary chapters.
- 4 Other data that are available in the WBD include lexical items elicited by means of questionnaires distributed on a smaller geographical scale, or data collected from local dictionaries. These data are excluded from the analysis.
- 5 We consider lexical items marked as Picardic, Old French or Walloon as loanwords from French and lexemes marked as Ripuarian as loanwords from German. Overall, only 14 types (100 tokens) from Ripuarian, 13 from Old French (261 tokens), 2 from Walloon (6 tokens) and 1 from Picardic (1 token) occur. Two lexical items, *proces* for PROCES-VERBAAL ‘report of an offence’ (43 tokens) and *tribunaal* for KANTONGERECHT ‘cantonal court’ (20 tokens), are marked as French and Latin. We considered these word types as Latin loanwords, as they are both borrowed from Latin via French (Philippa et al., 2003–2009).
- 6 These datasets are based on an offline copy of the original dictionary data. The dictionary data are also available online at e-wld.nl and e-wbd.nl (available from 14 December 2017 onwards). We only include the data that were collected by systematically distributed questionnaires by the Centre for Dialectology and Onomastics in Nijmegen (Franco, 2017:chapter 2 for a detailed description of the data).
- 7 We collected longitude and latitude information semi-automatically, using the Google Maps API (see <https://www.r-bloggers.com/using-google-maps-api-and-r/>, Accessed on 3 July 2017).

8 The maps are OpenStreetMaps (OpenStreetMap contributors, Accessed on 20 December 2018) that were constructed with the leaflet package in R. The R-code that was used to create the maps is included in the supplementary materials.

9 More specifically, we verified that there is no harmful structure in the residuals (i.e., homoscedasticity), that the residuals are normally distributed, that there is a linear relationship between the predicted and observed values of the response variable, and that a sufficient number of basis dimensions was used to construct the model.

10 French loanwords also occur frequently in Standard Dutch. Additionally, at the time of the collection of the dialect questionnaires, they were more frequent in the colloquial variety of Dutch in Belgium than in the Netherlands (Geeraerts et al., 1999). However, if French variants that were also accepted in Standard Dutch at the time of the questionnaire collection, are excluded from the analysis, the geographical patterns in the data remain highly comparable in every semantic field. Although the overall proportion of French is smaller, we still find more French in the fields of clothing & personal hygiene and society, school & education, and the difference between Belgium and the Netherlands remains clearly visible in these fields as well.

11 While we included both lexical items that were marked as German and lexemes that were marked as Ripuarian in the calculation of the ratio of German tokens per location, only one token in this area is marked as Ripuarian in the data (*Bohei* 'fuss', which was recorded in Kerkrade). All the other tokens that are presented in Table 8 are marked as High-German in the dictionaries. Furthermore, 95 out of these 100 German word types occur in the online version of the German *Duden* dictionary (<http://www.duden.de>, Accessed on 17 July 2017).

12 We use the term "German" concept informally to indicate that the concept has at least one German token.

References

- Auer, Peter. 2005. Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In Nicole Delbecque, Johan Van der Auwera & Dirk Geeraerts (eds.), *Perspectives on Variation. Sociolinguistic, Historical, Comparative*, 7–42. Boston: De Gruyter Mouton.
- Backus, Ad. 2014. A usage-based approach to borrowability. In Eline Zenner & Gitte Kristiansen (eds.), *New Perspectives on Lexical Borrowing: Onomasiological, Methodological and Phraseological Innovations* (Language Contact and Bilingualism 7), 19–40. Boston: De Gruyter Mouton.
- Cornelissen, Georg. 2007. Uit de woordenschat van het Kerkraads: Duitse woorden [The vocabulary of the Kerkraads dialect: German words]. In Ronny Keulen & Herman Crompvoets (eds.), *Riek van Klank: Inleiding in de Limburgse Dialecten [Wealth of Sound: Introduction to the Limburgish Dialects]* (Veldeke taalstudies 2), 60–65. Sittard: Veldeke Limburg.
- Crawley, Michael J. 2007. *The R book*. Chichester: Wiley.
- Daems, Jocelyne, Kris Heylen & Dirk Geeraerts. 2015. Wat dragen we vandaag: Een hemd met blazer of een shirt met jasje? [What to wear today: een hemd met blazer ('a dress shirt with suit jacket') or een shirt met jasje ('a dress shirt with suit jacket')? Convergence and divergence in Dutch clothing terminology]. *Taal en Tongval*, 67(2), 307–342.
- De Vriend, Folkert, Charlotte Giesbers, Roeland Van Hout & Louis Ten Bosch. 2008. The Dutch-German border: Relating linguistic, geographic and social distances. *International Journal of Humanities and Arts Computing*, 2(1–2), 119–134.
- De Vriendt, Sera. 2004. *Brussels* (Taal in stad en land 2). Tiel: Lannoo.
- Franco, Karlien. 2017. *Concept features and lexical diversity: A dialectological case study on the relationship between meaning and variation*. PhD dissertation. Leuven & Nijmegen: University of Leuven & Radboud University.
- Franco, Karlien, Dirk Geeraerts, Dirk Speelman & Roeland van Hout. 2019. Concept characteristics and variation in lexical diversity in two Dutch dialect areas. *Cognitive Linguistics*, 30(1), 205–242.
- Geeraerts, Dirk, Stefan Grondelaers & Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbalterminen [Convergence and divergence in the vocabulary of Dutch: A study into clothing and football terminology]*. Amsterdam: Meertens Instituut.
- Geeraerts, Dirk, Gitte Kristiansen & Yves Peirsman. 2010. *Advances in cognitive sociolinguistics* (Cognitive linguistics research 45). Berlin: De Gruyter Mouton.
- Giesbers, Charlotte. 2008. *Dialecten op de grens van twee talen: een dialectologisch en sociolinguïstisch onderzoek in het Kleverlands dialectgebied [Dialects at the border of two languages: A dialectological and sociolinguistic study in the Kleverland dialect area]*. PhD dissertation. Nijmegen: Radboud University.
- Haspelmath, Martin & Uri Tadmor. 2009. *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wold.cld.org>. (19 July, 2017).
- Havermans, Nele & Marc Hooge. 2011. *Kerkpraktijk in België: Resultaten van de zondagstelling in oktober 2009. Rapport ten behoeve van de Belgische Bischoppenconferentie [Church practices in Belgium: Results of the Sunday count in October 2009. Report for conference of Belgian bishops]*. Leuven: KU Leuven, Centrum voor Politicologie.
- Hinskens, Frans, Jeffrey L. Kallen & Johan Taeldeman. 2000. Merging and drifting apart: Convergence and divergence of dialects across political borders. *International Journal of the Sociology of Language*, 145, 1–28.
- Hock, Hans H. & Brian D. Joseph. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics* (Trends in linguistics. Studies and monographs 93). Berlin: De Gruyter.
- Janssens, Guy & Ann Marynissen. 2008. *Het Nederlands vroeger en nu [Dutch past and present]*. Leuven: Acco.
- Kristiansen, Gitte & René Dirven. 2008. *Cognitive sociolinguistics* (Cognitive Linguistics Research 39). Berlin & New York: De Gruyter Mouton.
- Kruijssen, Joep. 1990. Woordgeografie van ontlenen in een taalcontactsituatie [Word geography of loanwords in a language contact situation]. *Taal en Tongval*, 52, 4–45.
- Kruijssen, Joep. 1995. *Geografische patronen in taalcontact. Romaans leengoed in de Limburgse dialecten van Haspengouw. [Geographical patterns in language contact. Roman borrowings in the Limburgish dialects in the Hesbaye.]* PhD dissertation. Nijmegen: Radboud University.
- Margry, Peter J. & Charles Caspers. 2000. *Bedevoertplaatsen in Nederland: Limburg [Places of pilgrimage in The Netherlands: Limburg]*. Amsterdam: Meertens Instituut.
- McMahon, April. 1994. *Understanding language change*. Cambridge: Cambridge University Press.
- OpenStreetMap contributors. *s.d. OpenStreetMap*. <https://www.openstreetmap.org>. (20 December, 2018).
- Paul, Hermann. 1981. *Principles of the history of language. Translated from the Second Edition of the Original by H.A. Strong, M.A., L.L.D.* London: Longmans, Green and co.
- Philippa, Marlies, Debrabandere Frans, Quak Arend, Schoonheim Tanneke & Noline van der Sijs. 2003–2009. *Etymologisch Woordenboek van het Nederlands [Etymological Dictionary of Dutch]*. Amsterdam: Amsterdam University Press.
- Poplack, Shana, David Sankoff & Christopher Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26, 47–104.
- R Core Team 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schmeets, Hans. 2014. *De religieuze kaart van Nederland, 2010–2013 [The religious map of the Netherlands, 2010–2013]*. S.D.: Centraal Bureau voor de Statistiek.
- Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, 21(2), 121–137.
- Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor, *Loanwords in the World's Languages*, 55–75. Berlin & Boston: De Gruyter Mouton.
- Thomason, Sarah G. & Terrence Kaufman. 1991. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Van de Wijngaard, Ton. 2007. De Ripuarische dialecten [The Ripuarian dialects]. In Ronny Keulen & Herman Crompvoets (eds.), *Riek van klank: Inleiding in de Limburgse dialecten [Wealth of sound: introduction to the Limburgish dialects]* (Veldeke taalstudies 2), 45–59. Sittard: Veldeke Limburg.
- Van de Wijngaard, Ton & Ronny Keulen. 2007. De indeling van de Limburgse dialecten. In Ronny Keulen & Herman Crompvoets (eds.), *Riek van klank: Inleiding in de Limburgse dialecten [Wealth of sound: Introduction to the Limburgish dialects]* (Veldeke taalstudies 2), 15–23. Sittard: Veldeke Limburg.
- Van der Sijs, Noline. 2005. *Groot leenwoordenboek*. Utrecht: Van Dale Lexicografie.
- Van der Sijs, Noline & Jaap Engelsman. 2000. *Nota bene: De invloed van het Latijn en Grieks op het Nederlands [Nota bene: The influence of Latin and Greek on Dutch]*. 's-Gravenhage: SDU.

- Van der Wal, Marijke & Cor Van Bree. 2008. *Geschiedenis van het Nederlands [History of Dutch]*. Utrecht: Spectrum.
- Van Hout, Roeland, Joep Kruijssen & Marinel Gerritsen. 2014. Exosmosis along the Romance-Germanic language border in Belgium: The diffusion of French borrowings in the Dutch dialects of Haspengouw. In Raquel Casesnoves-Ferrer, Montserrat Forcadell Guinjoan & Núria Gavaldà Ferré (eds.), *Ens queda la paraula: Estudis de lingüística aplicada en honor a M. Teresa Turell, 197–220*. Barcelona: Institut Universitari de Lingüística Aplicada.
- Van Keymeulen, Jacques. 2008. Latijnse leenwoorden in het Nederlands en de Nederlandse dialecten [Latin loanwords in Dutch and in the Dutch dialects]. *Van Mensen en Dingen. Tijdschrift voor Volkscultuur in Vlaanderen*, 6. 69–85.
- Van Rij, Jacqueline. 2015. *Overview GAMM analysis of time-series data*. <http://www.sfs.uni-tuebingen.de/~jvanrij/Tutorial/GAMM.html>. (4 July, 2017).
- WBD = *Woordenboek van de Brabantse Dialecten [Dictionary of the Brabantic dialects]*. 1967–2005. Assen: Van Gorcum & Amsterdam: Gopher.
- Weijnen, Antonius A. 1967. Leenwoorden uit de Latinitas stratigrafisch beschouwd. *Verslagen en mededelingen van de Koninklijke Academie voor Nederlandse taal- en letterkunde*, 5. 365–480.
- Weinreich, Uriel. 1968. *Languages in contact: Findings and problems*. The Hague: De Gruyter Mouton.
- Wieling, Martijn. 2017. *Statistics course on generalized additive Modeling*. <http://www.let.rug.nl/~wieling/statscourse/>. (4 July, 2017).
- Wieling, Martijn. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70. 86–116.
- Wieling, Martijn, John Nerbonne and R. Harald Baayen. 2011. Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLOS ONE*, 6(9). e23613.
- Willemyns, Roland. 2013. *Dutch: Biography of a language*. New York: Oxford University Press.
- Winford, Donald. 2010. Contact and borrowing. In Raymond Hickey (ed.), *The handbook of language contact*, 170–187. Chichester: Wiley-Blackwell.
- WLD = *Woordenboek van de Limburgse Dialecten [Dictionary of the Limburgish dialects]*. 1983–2008. Assen: Van Gorcum & Amsterdam: Gopher.
- Wood, Simon. 2006. *Generalized additive models: An introduction with R* (Texts in statistical science). Boca Raton: Chapman and Hall/CRC.
- Wood, Simon. 2017. *Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*. <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>. (15 March, 2017).
- Zenner, Eline & Gitte Kristiansen. 2014. Introduction: Onomasiological, methodological and phraseological perspectives on lexical borrowing. In Eline Zenner & Gitte Kristiansen (eds.), *New perspectives on lexical borrowing: Onomasiological, methodological and phraseological innovations* (Language contact and bilingualism 7), 1–17. Boston: De Gruyter Mouton.
- Zenner, Eline, Dirk Speelman & Dirk Geeraerts. 2012. Cognitive sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. *Cognitive Linguistics*, 23(4). 749–792.
- Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R* (Statistics for biology and health). New York: Springer.

Appendix 1. Output of the GA(M)Ms

A. Numerical output of the GAMM for French loanwords

Parametric coefficients			
	Estimate	SE	p-value
Intercept	−3.030	0.017	<0.001
Semantic field (<i>personality & feelings</i>)	−1.187	0.028	<0.001
Semantic field (<i>church & religion</i>)	−0.164	0.021	<0.001
Semantic field (<i>clothing & hygiene</i>)	0.740	0.018	<0.001
Approximate significance of smooth terms			
	edf	p-value	
s(lon, lat): sem. field (<i>society, school & education</i>)	24.85	<0.001	
s(lon, lat): sem. field (<i>personality & feelings</i>)	18.03	<0.001	
s(lon, lat): sem. field (<i>church & religion</i>)	3.83	<0.001	
s(lon, lat): sem. field (<i>clothing & hygiene</i>)	21.96	<0.001	
s(location)	329.29	<0.001	
Explanatory power			
Null deviance explained			92%
Adjusted R ²			0.908

B. Numerical output of the GAM for Latin loanwords

Parametric coefficients			
	Estimate	SE	p-value
Intercept	−4.850	0.035	<0.001
Semantic field (<i>personality & feelings</i>)	−3.801	0.269	<0.001
Semantic field (<i>church & religion</i>)	2.226	0.037	<0.001
Semantic field (<i>clothing & hygiene</i>)	−1.651	0.083	<0.001
Approximate significance of smooth terms			
	edf	p-value	
s(lon, lat): sem. field (<i>society, school & education</i>)	64.07	<0.001	
s(lon, lat): sem. field (<i>personality & feelings</i>)	15.94	<0.001	
s(lon, lat): sem. field (<i>church & religion</i>)	16.10	<0.001	
s(lon, lat): sem. field (<i>clothing & hygiene</i>)	17.04	<0.1	
Explanatory power			
Null deviance explained			94.4%
Adjusted R ²			0.957

C. Numerical output of the GAMM for German loanwords

Parametric coefficients			
	Estimate	SE	p-value
Intercept	-6.879	0.171	<0.001
Semantic field (<i>personality & feelings</i>)	-0.116	0.244	0.634
Semantic field (<i>church & religion</i>)	0.714	0.184	<0.001
Semantic field (<i>clothing & hygiene</i>)	-0.027	0.222	0.902
Approximate significance of smooth terms			
	edf		p-value
s(lon, lat): sem. field (<i>society, school & education</i>)	13.91		<0.001
s(lon, lat): sem. field (<i>personality & feelings</i>)	11.58		<0.001
s(lon, lat): sem. field (<i>church & religion</i>)	14.61		<0.001
s(lon, lat): sem. field (<i>clothing & hygiene</i>)	12.50		<0.001
s(location)	59.72		<0.001
Explanatory power			
Null deviance explained			89.2%
Adjusted R^2			0.928

Appendix 2. Distribution of “German” concepts that occur in at least two locations in the Riparian dialect area in the semantic field of Personality and feelings

Concept	Translation	Number of types	Number of observations	Number of German types	Number of German observations	Proportion of German types	Proportion of German tokens
onnozel persoon	nitwit	9	10	1	1	0.111	0.100
snotneus	brat	5	10	1	1	0.200	0.100
prutsen	to mess about	6	6	1	1	0.167	0.167
begrip, besef	understanding	3	5	1	1	0.333	0.200
mokken	to sulk	5	5	1	1	0.200	0.200
bezorgd	concerned	3	4	1	1	0.333	0.250
kouwe drukte	much ado about nothing	4	4	1	1	0.250	0.250
mopperen	to grumble	4	4	1	1	0.250	0.250
potsachtig	comical	4	4	1	1	0.250	0.250
stiekem	sneaky	4	8	2	2	0.500	0.250
teleurgesteld (worden)	(to be) disappointed	3	4	1	1	0.333	0.250
tevreden, tevredenheid	satisfied, satisfaction	3	4	1	1	0.333	0.250
beestachtig persoon, beestachtig	savage (person)	3	3	1	1	0.333	0.333
beteuterd	dismayed	3	3	1	1	0.333	0.333
de baas spelen	to boss someone around	3	3	1	1	0.333	0.333
genoegen (doen)	(to be) satisfactory	2	3	1	1	0.500	0.333
gluiperd	shifty character	3	3	1	1	0.333	0.333
lasteren	to insult	3	3	1	1	0.333	0.333
onwennig (voelen)	(to feel) ill at ease	3	3	1	1	0.333	0.333
treuren	to be sorrowful	3	3	1	1	0.333	0.333
van katoen geven	to give at all one has got	3	3	1	1	0.333	0.333
zich vergissen	to be mistaken	3	3	1	1	0.333	0.333
zich zeer slecht gedragen	to behave very badly	2	3	1	1	0.500	0.333

Concept	Translation	Number of types	Number of observations	Number of German types	Number of German observations	Proportion of German types	Proportion of German tokens
bangerik	coward	5	9	1	4	0.200	0.444
aandringen	to insist	2	2	1	1	0.500	0.500
aanstoot geven	to give offence	2	2	1	1	0.500	0.500
aarzelen	to hesitate	2	2	1	1	0.500	0.500
angst	fear	2	2	1	1	0.500	0.500
baldadig (persoon)	rowdy (person)	2	2	1	1	0.500	0.500
bedrieger	fraud	2	2	1	1	0.500	0.500
bestemmen	to reserve	2	2	1	1	0.500	0.500
bezadigd	steady	2	2	1	1	0.500	0.500
boertig	coarse	2	2	1	1	0.500	0.500
geheimzinnig	mysterious	2	2	1	1	0.500	0.500
gemakkelijk	easy	2	2	1	1	0.500	0.500
grapjas	joker	4	4	2	2	0.500	0.500
gril	whim	2	2	1	1	0.500	0.500
hansworst	buffoon	2	2	1	1	0.500	0.500
hopen	to hope	2	2	1	1	0.500	0.500
hulp, bijstand	help	2	2	1	1	0.500	0.500
iemand kwaad maken	to anger someone	2	2	1	1	0.500	0.500
ingetogen	modest	2	2	1	1	0.500	0.500
jaloers	jealous	2	2	1	1	0.500	0.500
kalm, bedaard	calm	2	2	1	1	0.500	0.500
keus	choice	2	2	1	1	0.500	0.500
kiezen	to choose	2	2	1	1	0.500	0.500
kniezen	to mope	2	2	1	1	0.500	0.500
knoeier	sloppy person	2	2	1	1	0.500	0.500
konkelen	to scheme	2	2	1	1	0.500	0.500
lichtgeraakt, kregel	touchy	2	2	1	1	0.500	0.500
lui	lazy	2	2	1	1	0.500	0.500
nauwgezet, nauwgezet persoon	conscientious (person)	2	2	1	1	0.500	0.500
niet helder van geest	not lucid	2	2	1	1	0.500	0.500
schuchter	shy	2	2	1	1	0.500	0.500
slecht mens, slechte kerel	bad person, bad fellow	2	2	1	1	0.500	0.500
slordig	sloppy	2	2	1	1	0.500	0.500
troosten, troost	(to) comfort	2	2	1	1	0.500	0.500
verstandig	sensible	2	2	1	1	0.500	0.500
verzuimen	to neglect	2	2	1	1	0.500	0.500
zich gedragen	to behave	2	2	1	1	0.500	0.500
zich kwaad maken	to get angry	2	2	1	1	0.500	0.500
zonder opzet	unintentionally	2	2	1	1	0.500	0.500
zwoegen	to labour	4	4	2	2	0.500	0.500
slim	smart	4	7	1	4	0.250	0.571
treiteren	to torment	3	5	1	3	0.333	0.600
geestig	witty	3	3	2	2	0.667	0.667
huichelaar	hypocrite	3	3	2	2	0.667	0.667
iemand prijzen	to commend someone	3	3	2	2	0.667	0.667

Concept	Translation	Number of types	Number of observations	Number of German types	Number of German observations	Proportion of German types	Proportion of German tokens
informereren (onoverg.)	to inform	2	3	1	2	0.500	0.667
leep, doortrapt	cunning	3	3	2	2	0.667	0.667
prettig	pleasant	3	3	2	2	0.667	0.667
schelm	crook	3	3	2	2	0.667	0.667
uitbrander	dressing-down	2	3	1	2	0.500	0.667
vanzelfsprekend	obvious	2	3	1	2	0.500	0.667
vermaak	entertainment	3	3	2	2	0.667	0.667
wijs	wise	2	3	1	2	0.500	0.667
zich bedenken	to change one's mind	3	3	2	2	0.667	0.667
begrijpen	to understand	2	4	1	3	0.500	0.750
ellende (lijden)	misery	3	4	2	3	0.667	0.750
plezier maken	to have fun	2	4	1	3	0.500	0.750
pret, schik	fun	3	4	2	3	0.667	0.750
pretmaker	merrymaker	3	4	2	3	0.667	0.750
vrolijk	cheerful	4	6	3	5	0.750	0.833
degelijk	decent	1	2	1	2	1.000	1.000
dwingen	to force	1	2	1	2	1.000	1.000
eenvoudig	simple	1	2	1	2	1.000	1.000
gunst	favor	2	2	2	2	1.000	1.000
ophouden met het werk	to end the working day	2	3	2	3	1.000	1.000
schipperen	to compromise	2	2	2	2	1.000	1.000
slimmerik	smart number	3	4	3	4	1.000	1.000
sober	sober	1	2	1	2	1.000	1.000
vreugde	joy	2	2	2	2	1.000	1.000
zedig	chaste	1	2	1	2	1.000	1.000
zijn tevredenheid betuigen	to express one's satisfaction	2	3	2	3	1.000	1.000
mean		2.606	3.106	1.245	1.553	0.528	0.541