VIVES
DISCUSSION
PAPER

N° 58
April 2017

On dialects, networks, and labour mobility

Damiaan Persyn
JRC-IPTS, European Commission, Seville
University of Leuven, VIVES

On dialects, networks, and labour mobility*

Damiaan Persyn

20th April 2017

Abstract

I consider dialect differences between Flemish municipalities and the relation with commuting flows. Less commuting takes place between municipalities with larger dialect differences, even after controlling for travel time by car and public transport, and for the number of jobs and workers in the surrounding municipalities. Dialect differences may be indicators of local cultural differences, marking boundaries of local socio-economic networks which act as hurdles to economic exchange. The results point to an upper bound as to how much additional labour mobility can be expected from public policies such as infrastructure investment.

1 Introduction

This paper analyses the relation between local dialects and commuting behaviour, using data for the Belgian region of Flanders. Flanders is characterised by relatively large variations in dialects over small geographic distances. These small language differences may hinder smooth communication on the shop floor, which could cause employers and employees to prefer local, more easily understood, partners. A statistical analysis confirms that there is less commuting between municipalities with more distinct dialects, even when carefully controlling for travel time and the economic surroundings of both origin and destination. However, it is unlikely that the relatively

*I am grateful to Martijn Wieling for allowing the use of his database with dialect distances, and to both Jo Reynaerts and Martijn Wieling for useful remarks and suggestions which have greatly improved the paper.

small frictions in oral communication due to dialects are the cause of this observed 'missing commuting'.

Dialects form and evolve over decades and centuries, under the influence of, for example, topological characteristics such as rivers and mountain ranges, historical migration and trade patterns, or geopolitical factors such as wars and international or regional borders. These factors, however, also are key in determining socio-economic links, cultural exchange and therefore the emergence of distinct cultures. The language(variety) a person speaks therefore also marks the membership of a certain community, network, and (sub)culture. These networks and the accompanying language differences are persistent over time, even when the underlying causes which gave rise to them have long disappeared –say because of changes in political borders, transport infrastructure or trade patterns.

Culture and cultural differences are known to affect a wide variety of economic phenomena. Countries with a common language or shared history trade significantly more with each other (see for example Guiso *et al.*, 2009; Melitz, 2008; Ginsburgh and Weber, 2011). More close to this paper, Falck *et al.* (2012) show how contemporaneous migration patterns coincide with historic dialects in Germany, and point to cultural factors as the underlying explanation. There are several possible reasons why people would interact less over cultural boundaries, even when superficially some cultural differences may seem minor. Apart from factors such as xenophobia, more economic and rational factors are that there is less information exchange across network-borders. Possibly, some matches on the labour market do not materialise when participants lack the necessary information because their networks overlap insufficiently. Alternatively, reputation might be more relevant within a network, reducing trust and economic exchange between partners from different networks (Guiso *et al.*, 2009). There exists both theoretical (Calvó-Armengol and Zenou, 2005) and empirical (Ioannides and Soetevent, 2006) evidence that limited information flow due to a small network size is detrimental to labour market outcomes, leading to higher unemployment, lower wages and economic growth (Glaeser *et al.*, 2000).

Later in this paper I perform a regression analysis to quantify the relation between dialect distances and commuting. Before jumping to conclusions regarding any of the results of this analysis, it is important consider what they might imply.

Correlation between dialect distances and labour mobility could -at least partially- be *caused* by the effect of dialects on mobility, due to the frictions and communication costs they cause. We can imagine that dialect differences make communication somewhat more difficult, and both an employer and employee might therefore have an incentive to match with a partner speaking the same or closely related dialect.

But it is unlikely that friction in oral communication alone is the main cause of any observed correlation. A first reason to doubt this direct causal explanation is the fact that the GTRP pronunciation data which we use was collected mainly in the 1980's and 1990's, and focussed mainly on adults and elderly, in order to record the *original* –and not necessarily the contemporaneous– local dialect as close as possible. The median speaker in the dataset was 63 years old at the time of recording, and the median year of recording is 1987, implying the median speaker in the dataset was growing up during the 1930's. It is clear that these 'historic' dialects do not always correspond to the dialect of younger, contemporaneous inhabitants, or inhabitants with mixed backgrounds or an interregional migration history.

The fact that the original dialects are several decades older than the commuting data reduces the possibility that the observed correlation is due to reverse causality. It is possible and even likely that *historic* limited labour mobility between regions is a cause of the emergence of dialect differences between regions. However, it is unlikely that the historic dialects are subject to influences from factors related to contemporaneous patterns of labour mobility, after controlling for commuting costs and the location of employees and jobs in the neighbouring municipalities. Most people whose pronunciation was used in the database were born before highways were build, or before the demise of the shipbuilding, steel, or coal-mining industries, and their associated geographical location and commuting-patterns. The underlying cause of dialects-structures in Flanders is known to go back many centuries further in time.

An important alternative explanation for the correlation which is observed between these historic dialects and current commuting flows, is the fact that the historic dialects are markers for socio-cultural networks which persisted over time, on a geographically small scale. Studies like Guiso *et al.* (2009) have found a strong correlation between a common language between countries, the level of trust which exists between them, and the intensity of trade and foreign direct investment between them. Similarly, Falck *et al.* (2012) find an effect of dialects recorded in the 19th century on contemporaneous migration flows, while controlling for a host of possible other explanatory factors such as soil type and religion, suggesting underlying cultural differences as the remaining main explanation.

2 Dialect distances

2.1 Measuring dialect distances

The data on dialect differences that are used in this paper have been elaborated by Wieling *et al.* (2007), and have kindly made available by Martijn Wieling for this study. The work of Wieling *et al.* (2007) in turn is based on the data from the Goeman-Taeldeman-Van Reenen-Project (see Goeman and Taeldeman, 1996). The GTRP contains the phonetically transcribed pronunciation of 1876 words in 613 municipalities in Flanders (Belgium) and The Netherlands. In this study I only use the 189 Flemish locations and pronunciations, and a selection of 562 words (see Wieling *et al.*, 2007).

To determine the difference in pronunciation, or dialect distance, between two localities, Wieling *et al.* (2007) use a Levenshtein method. This is a metric for measuring the difference between two sequences of characters, where the minimal number of edits (additions/deletions/substitutions) is calculated which is needed to transform one sequence into the other. The Levenshtein distance and related metrics have applications in for example search engines or spell-checkers. Let us consider the word 'earth' from the GTRP database as an example. For the municipality of Veurne in West-Vlaanderen we have the pronunciation (ærdə). In Wetteren in Oost-Vlaanderen this becomes (eərdə). The Levenshtein distance is calculated as follows

æ		r	d	ə
e	ə	r	d	ə
1	1	0	0	0

With 1 one substitution and 1 addition, we obtain a Levenshtein distance of 2. If we calculate the distance between the pronunciations of Veurne (ærdə) en the pronunciation (jat) of the municipality of Lauw in the province of Limburg, the Levenshtein distance becomes 4.

In an adjustment of the standard Levenshtein-algorithm Wieling *et al.* (2007) additionally penalise interchanging vowels and consonants, to take in to account that the pronunciation of an 'e' is often closer, say, 'a' as compared to 'p'. To avoid giving more weight to longer words, the Levenshtein-distances are expressed relative to word-length. The mean word-distance in the database is 0.53. However, this also reflects the language distance between remote small municipalities between which no commuting takes place. Weighing the dialect-distances with the amount of commuting, the mean becomes 0.23, with a standard deviation of 0.2.

Flanders contains 308 municipalities and therefore $308^2 = 94864$ municipality pairs between which to consider commuting flows. The GTRP only contains information for $189^2 = 35721$ of those pairs. For the missing pairs, the Levenshtein distance was taken for the nearest pair with available data. This approximation is unlikely to significantly affect the results, as for 75 percent of the cases with missing information the nearest municipality with information is at a distance of less than 5 kilometre. For none of the cases this distance is larger than 10 kilometre.

2.2 Dialect borders and clusters

The matrix with bilateral dialect distances between the municipalities summarises all pronunciation differences in a single number. This may masks subtle pronunciation differences which make dialects easily recognisable by humans. In this section I aim to show that, using appropriate methods, well known geographic patterns in dialect differences can nevertheless be revealed using the matrix of dialect distances.

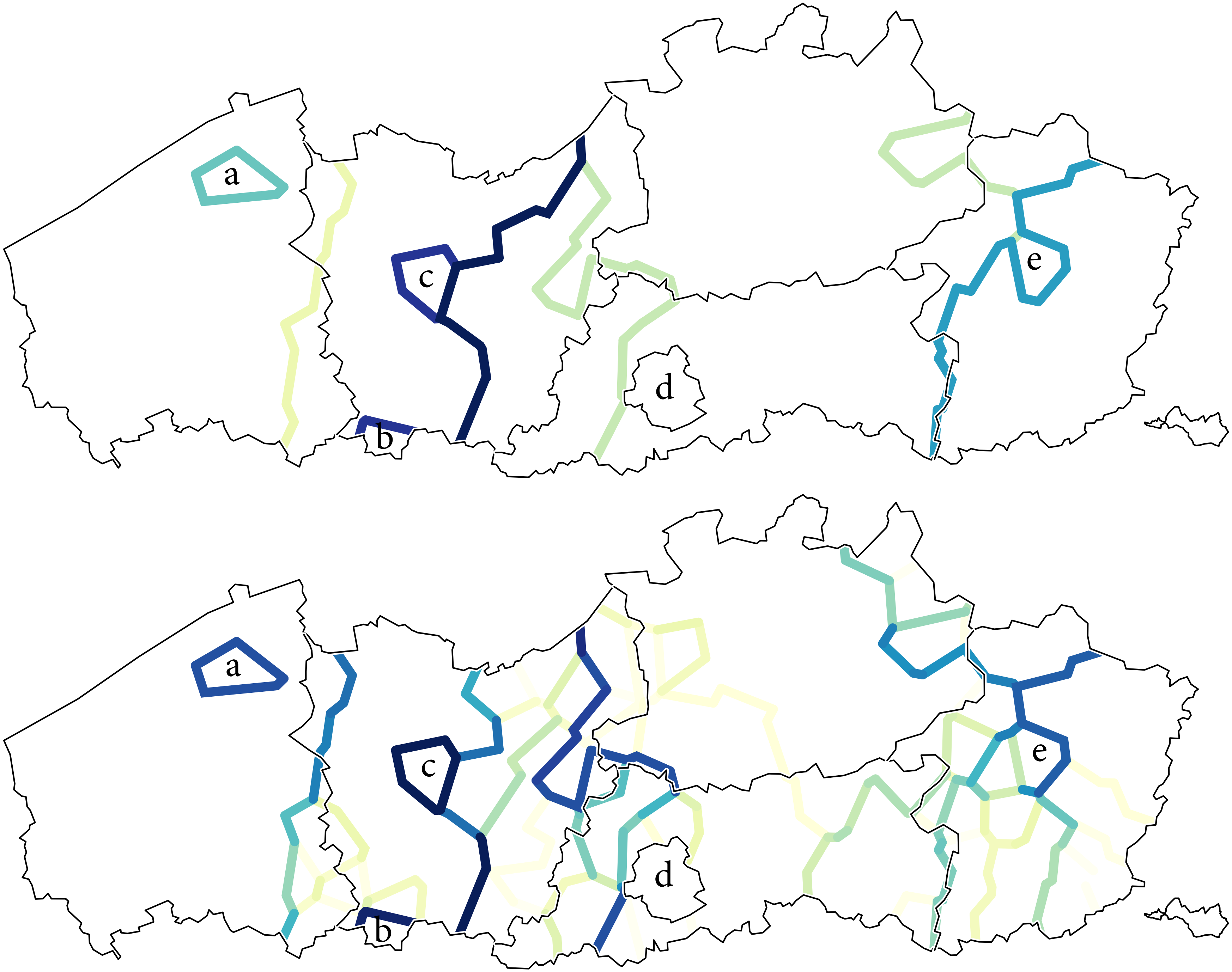

Figure 1: Dialect borders in Flanders as calculated using the GTRP data. Top: division in two clusters (darkest border), than three, etc. Bottom: idem, but repeated 100 times with added noise. a: Brugge; b: Ronse; c: Gent d: Brussel; e: Hasselt

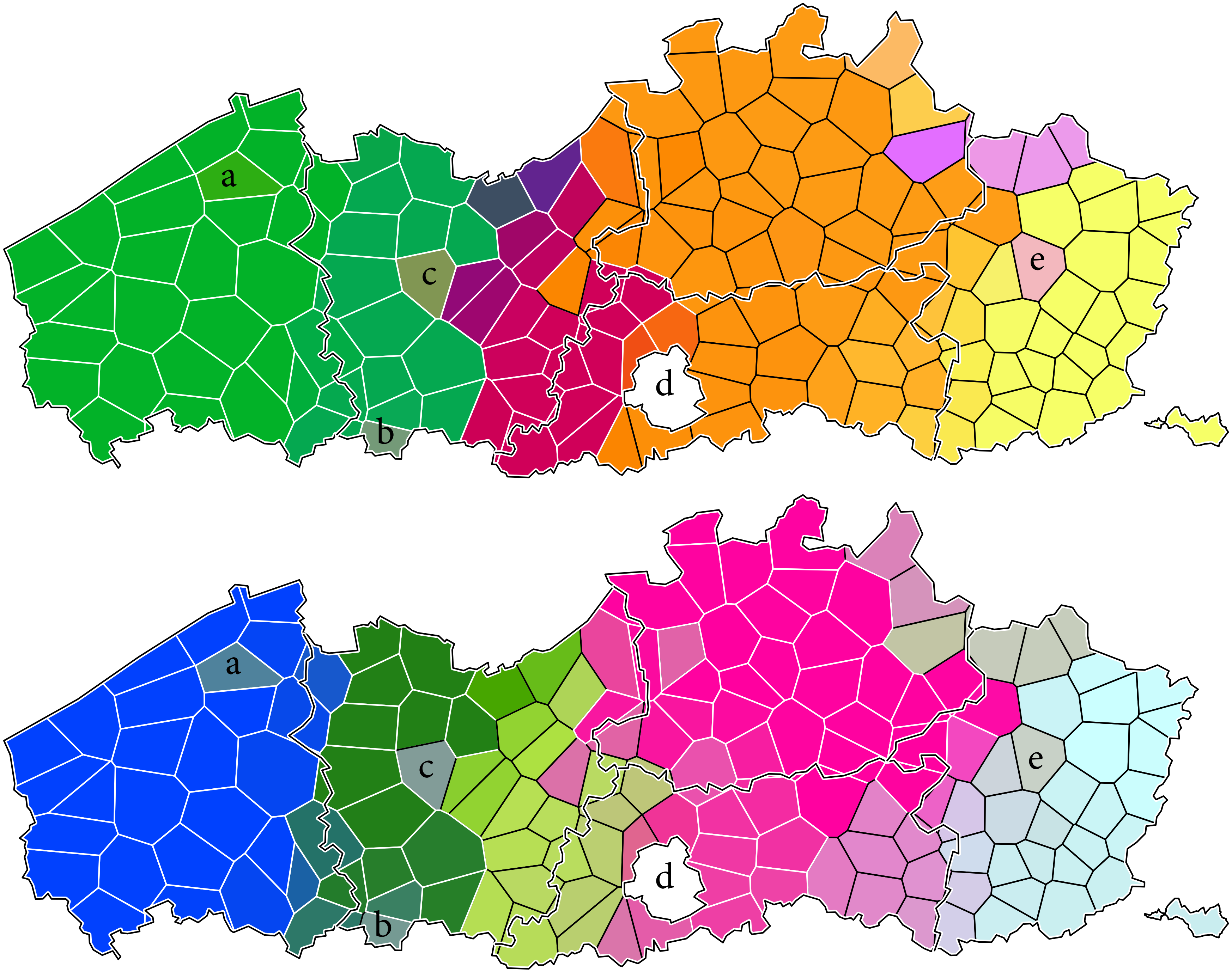

Figure 2: Distinguishing between dialects in Flanders using multidimensional scaling. a: Brugge; b: Ronse; c: Gent; d: Brussel; e: Hasselt

The maps in Figures 1 and 2 were made using the RUGL04 software of Peter Kleiweg.[1] The map in the upper panel of Figure 1 divides the municipalities in different clusters[2], such that the Levenshtein distance is small between the clusters and small within them.

Setting the algorithm to identify the two largest dialect-clusters, a single border is drawn to the east of the city of Ghent. This border divides between East-Flemish and Brabantian, two well known and quite distinctive dialects.[3] Attempting to draw two borders between three dialect groups, two city dialects are taken as a separate cluster. These are the dialects of the city of Ghent and Ronse, which are more similar to Brabantian when compared to the rural municipalities that surround them, because of historic migration and trade patterns.[4]

The algorithm determining clusters in the above maps is sensitive to small changes in the data on dialect distances: A small change can make a dialect border run quite differently, and also the order in which the different clusters are recognised may be affected. This can be seen as a weakness of such methods, but it can also provide additional insights. To exploit this, I repeated the above clustering exercise 100 times, while adding random noise to the data in each step. If a border is unaffected by this addition of noise, it implies that the dialect distance at this point is substantial. If

[1] See http://www.let.rug.nl/~kleiweg/L04/

[2] The GTRP dataset may contains multiple measurements per municipality. Because the commuting data are only available on the municipal level, I chose to take the mean of the dialect distances per municipality, or the measurement in the most populous area. This causes the map to show somewhat less detail when compared to other publications using the GTRP data, but it makes the maps consistent with the statistical analysis of commuting flows.

[3] The discussion of dialects in Flanders can be confusing. The entire region and the various dialects are referred to as Flanders and Flemish, respectively. But this terminology is relatively new. The original county of Flanders, and the original Flemish dialects, are spoken in the province of West-Vlaanderen and the western part of the province of Oost-Vlaanderen. The west of the province Oost-Vlaanderen historically was part of the county of Brabant, and Brabantian dialects are spoken there.

[4] Distinguishing four clusters splits Brabantian and Limburgian dialects in the east of the country. With five groups, West-Flemish is separated from East-Flemish. Dividing in even more clusters takes apart the Brabantian dialects spoken in the East of East-Flanders, and some smaller dialects in the north-east of the province of Antwerp. The latter group forms part of the North-Brabantian dialects (e.g. in Lommel), which are spoken in the Netherlands, but the algorithm takes them together with West-Limburgian dialects which are spoken in some municipalities in Northern-Limburg municipalities (such as Overpelt).

a border disappears or moves, this reveals that there are alternative ways to draw the borders and classify dialects. The bottom panel of Figure 1 shows the result. Although the panel seems to be dotted with many new borders, some interesting conclusions can be drawn. First, it is noticeable that the city dialects of Ghent and Ronse remain quite distinct: they are almost always recognised as separate clusters even if noise is added. The cities of Brugge en Hasselt now also appear as having distinct dialects; Other important borders which remain quite stable are the dialect border between East-Flemish and Brabantian; Brabantian versus Limburgian in the north of the province of Limburg. In contrast, many variants are found for the southern dialect border between Limburgian and Brabantian, but the alternative borders correspond to well know distinct dialects which are linguistically related to both Limburgian and Brabantian[5]

Multidimensional scaling (MDS) is an alternative method to distinguish between dialect clusters, using colours. Using MDS, each municipality with dialect measurements is assigned three coordinates, such that some measure of distance between the coordinates of all municipalities approximates the original matrix with dialect distances as close as possible. The three coordinates are subsequently mapped into intensities of the three basis colours red, green and blue, and represented on a map. The top map in Figure 2 uses parameters which divide the area into few and large dialect groups, clearly showing the three main dialects of Flanders (Flemish, Brabantian and Limburgian). The bottom map allows for more subtle differences between the dialects to be visible. Here East-Flemish appears less close to West-Flemish, and West-Brabantian is closer to East-Flemish. The differences between Limburgian and the deviating North-Brabantian dialects in the north of Limburg are less visible in this map.

To conclude our discussion of the dialect data and how it relates with the known

[5]These dialects are sometimes taken together with the Limburgian cluster or the Brabantian cluster (West-Getelands around Tienen and Oost-Getelands around Sint-Truiden). The dialects of Ravels and Arendonk in the northeast of the province of Antwerp are now often determined as a separate dialectcluster, which is as expected since they are quite distinct from the Brabantian dialects of Flanders and Limburgian. The situation in the northwest of the province of Flemish-Brabant is complex and is not readily matchable with known dialect groups.

structure of dialects in Flanders, it is important to repeat that there may very well exist better methods and data to cluster dialects. Our goal, however, was to illustrate that even in the simple matrix of dyadic dialect distances, which summarises all pronunciation differences between a pair of municipalities in a single number, the most important and well known properties of Flemish dialects can be recognised. In the end, however, in this study the data will not be used to cluster or distinguish dialects, but will be used consider the relation between dialect differences and labour mobility.

2.3 An example: dialect distances for the Zwalm municipality

Should dialect differences be highly correlated with distance or travel time, it would be difficult to distinguish between the effects of these variables on commuting. It are exactly those cases in which dialects abruptly change over a short distance (or do not change over a long distance) which allow us to disentangle the effects of these factors. Abrupt changes occur at the dialect borders which were illustrated above, whereas the more homogeneous clusters show areas where dialects do not change much with distance.

As an example, Figure 3 considers the dialect distances between the municipality of rural Zwalm in East-Flanders and its neighbouring municipalities. The dialect distance to neighbouring Ronse (0.62) or the city of Ghent (0.65) is particularly high when compared to other municipalities at comparable distances in the East-Flemish dialect cluster (dark green). The dialect distances within the dialect cluster do not seem to be directly correlated with distance. The dialect distance with distant rural municipalities above the city of Ghent is small. In contrast, when crossing the nearby dialect border with Brabantian dialects at the east, the dialect distances increase from around 0.3 to 0.45-0.5. This is also the case when considering distances to West-Flemish dialects. Comparing with the Brabantian dialects close to Brussels (Halle, Londerzeel) gives even larger dialect distances.

The example of Zwalm is quite representative: dialect distances do not simply

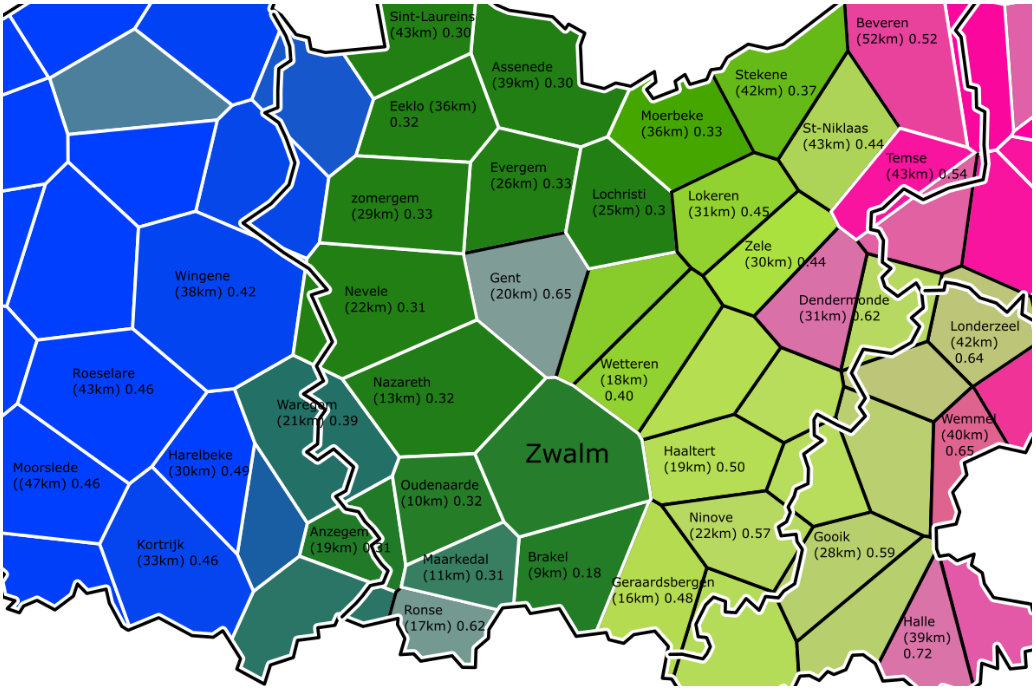

Figure 3: Different Levenshtein dialect distances relative to Zwalm

increment with distance, but are rather constant within a dialect cluster and increase sharply at dialect borders. These deviations between geographical distance or travel time and dialect distance will allow us to consider the effects of these variables separately.

3 Quantitative Analysis

3.1 Methodology

3.1.1 A model for commuting flows

The quantitative analysis uses a gravity model for commuting as discussed, for example, in Persyn and Torfs (2015). The commuting flow F_{od} between an origin community o and destination d is modelled as a function of their economic mass and some measure of the commuting cost between them. For the mass variable in the municipality of origin o, I take the number of workers W_o, and for the destination d mass variable the number of jobs J_d. Using the geographic distance D_{od} as a measure of the commuting cost, an example of a simple model for the commuting flow F_{od} from o to d is the following equation which is based on Newtons gravity law:[6]

$$\mathrm{F}_{od} = \frac{\mathrm{W}_o \times \mathrm{J}_d}{\mathrm{D}_{od}^{\alpha}}. \tag{1}$$

This model has a serious shortcoming, however: the total number of outgoing commuters $\sum_d P_{od}$ from the municipality o can differ substantially from the number of workers W_o residing in the municipality. This is neither logical nor desirable. To address this, the model is extended[7] with two 'balancing factors' A_o and B_d (see equation (2a)). Imposing the two constraints $\sum_d F_{od} = W_o$ and $\sum_o F_{od} = J_d$ allows to derive the required values of A_o and B_d, which ensure that the predicted outgoing and incoming commuter flows match the observed number of workers and jobs.[8] They are shown in equation (2b). A last adaptation we make to the standard gravity model is to replace the geographic distance with a generalised commuting cost factor K_{od}, which determines how costly or attractive commuting is between each pair of municipalities. We assume the generalised commuting cost K depends on travel time using public transport and car, and the dialect distance. The functional form we assume is $K_{od} = \exp[\alpha_0 \log(\text{travel time by car}_{od}) + \alpha_1 \log(\text{travel time by public transport}_{od}) + \alpha_2(\text{dialect distance}_{od})]$, where $\alpha_0, \alpha_1, \alpha_2$ are the parameters which will be estimated by means of the statistical analysis.

[6]Note that o and d are allowed to be the same municipality. The distance and travel time to the own municipality is calculated as a function of the surface area, see Persyn and Torfs (2015). The language distance to the own municipality is assumed to be 0.

[7]This extended model is known as the 'Wilson Doubly Constrained' Model, see Wilson (2010).

[8]In a Poisson regression, these constraints can equally be imposed by including a set of origin and destination dummies, see (Anas, 1983; Fally, 2015). These terms are called 'multilateral resistance' terms in the literature following Anderson and Van Wincoop (2003). Using dummies makes it impossible to perform a counter-factual analysis, as this requires calculating A and B in an alternative scenario (for example with reduced commuting costs).

The resulting model then is given by

$$\mathrm{F}_{od} = \frac{\mathrm{W}_o \times \mathrm{J}_d \times A_o \times B_d}{\mathrm{K}_{od}} \tag{2a}$$

$$A_o \equiv \sum_d \frac{J_d \times B_d}{K_{od}} \quad B_d \equiv \sum_o \frac{W_o \times A_o}{K_{od}}. \tag{2b}$$

This models how a *given* geographic distribution of jobs is filled in by a *given* geographic distribution of workers, and allows to investigate how a change in commuting cost and dialect distances would affect the geographic distribution of commuting flows which match jobs and workers. Through the balancing factors, the constraints also introduce a dependency on the economic surroundings for the flow between any pair o, d: in order to predict how many of a fixed number of jobs J_d in some destination will be filled in by workers from a specific origin o, it matters whether or not there are alternative sources of workers close to the destination. Likewise, it matters what alternative destinations there are for the workers in o.

3.1.2 Estimating the gravity model

No commuting takes place between most municipality pairs in the dataset, as the majority of pairs regard small municipalities which are located at a fair distance. Most commuting happens between municipalities at a short distance, or for which at least one of the members is large. This poses a problem for estimation methods which log-linearise equation (1) before estimating it. In this study, we consider the commuting data to be count data, and estimate it using a Poisson regression model in which 0 is an admissible outcome. More details are available in Silva and Tenreyro (2006) and Persyn and Torfs (2015).

3.2 Other data sources

Dialect distances tend to be larger at natural barriers such as large rivers which also affect commuting costs. We therefore aim to precisely measure and control quite precisely for the effect of travel time to isolate the effect of dialects from these factors. For all 94864 municipality pairs the travel time per car and public transport was calculated. Using Perl and Javascript these travel times were respectively obtained through the Google Maps API and by scraping the Belgian public railway provider NMBS website. We include travel per tramway, bus, metro or bike as 'public transport'.[9] This is mostly the town hall or another central point in the municipality. The commuting data on the municipal level are made available by the social security service (Rijksdienst voor Sociale Zekerheid) and pertain to the year 2008. The commuting data comprises only paid employment and thus excludes the self-employed.

3.3 Results

Table 1 shows the result of the Poisson regression analysis. The results for travel time can be interpret as elasticities when considering small changes. An increase in the travel time with 1 percent is associated with a decrease in the number of commuters by 2.2 percent. This effect is much smaller for public transport.

The average dialect distance between all municipality pairs is 0.53. This average is not very relevant, however, because there is very little or no commuting taking place between the most remote municipalities which also have the largest dialect distances. If we weight the dialect distances by the amount of commuting between the municipalities, the weighted average dialect distance is 0.25 with a standard deviation of 0.2. This weighted average dialect distance is the dialect distance faced by the typical commuter, rather then the dialect distance between a typical municipality pair. The result of -0.439 in Table 1 for the dialect distance variable implies that if the dialect distance decreases with 0.1 (half a standard deviation), the predicted commuting flow increases with about 4.4 percent.

To better understand the order of magnitude of the association between dialect distance and commuting, we consider the predicted change in commuting flows

[9] The coordinates of the departure and arrival point in each municipality were taken from the AdminVector database provided by Statistics Belgium.

log(travel time by car)	−2.206
	(0.137)
log(travel time by public transport)	−0.487
	(0.103)
dialect distance	−0.439
	(0.204)
# observations	94864

Bootstrapped standard errors between brackets

Table 1: The estimated gravity model for commuting in Flanders, with an effect for travel time and dialect distance. The table omits the correction factors A_o and B_b, and the mass variable W_o and J_d, the coefficients of which have been constrained to one.

between the 5 Flemish provinces in hypothetical cases where the dialect distances are reduced. We compare these results with the predicted commuting flows for the originally observed dialect distances. These baseline predicted inter-provincial commuting flows are reported in Table 2. The model predicts that 414018 of a total

	Antwerpen	Limburg	Oost-Vl.	Vl.-Brabant	West-Vl.	Total (W_o)
Antwerpen	441508	19649	30054	36933	5441	533585
Limburg	27546	193041	5248	21828	1986	249649
Oost-Vlaanderen	56549	3791	295278	33043	38582	427242
Vlaams-Brabant	35543	16052	18265	172502	3759	246120
West-Vlaanderen	10582	1581	40119	7470	291860	351612
Total (J_b)	571727	234113	388964	271776	341627	1808207

Table 2: Predicted aggregate commuting flows between Flemish provinces, in the year 2008.

of 1808207 employees in the dataset work in a province which is different from their province of residence, or about 22.9 percent.

We now consider the hypothetical case where the dialect distances are limited to 0.2. This is approximately the lower bound of the typical dialect distance between nearby municipalities in the same dialect cluster (see figure 3). The resulting predicted commuting flows in this hypothetical case are reported in Table 3. It is important

	Antwerpen	Limburg	Oost-Vl.	Vl.-Brabant	West-Vl.	Total (W_o)
Antwerpen	434245	21177	33509	37755	6898	533584
Limburg	28846	189296	6084	22934	2487	249647
Oost-Vlaanderen	59882	4408	284747	35397	42809	427242
Vlaams-Brabant	36581	17325	20382	167083	4749	246120
West-Vlaanderen	12173	1907	44243	8607	284684	351614
Total (J_b)	571727	234113	388964	271776	341627	1808207

Table 3: Predicted aggregate commuting flows between Flemish provinces, for the year 2008, in a scenario where dialect distances between provinces are reduced to 0.2 – a level typical within a dialect cluster.

to note that, as imposed by the doubly-constrained model, the predicted in and outflows in each municipality is unchanged. This then obviously must also hold on the province level (abstracting from rounding errors). Only the geographic distribution of commuting flows is allowed to change between scenario's. The predicted number of cross-province commuters increases from 414018 to 448153, an increase of 8 percent, increasing the share of cross-province commuters from 22.9 to 24.8 percent. Similarly, if we repeat this exercise with 6 dialect clusters rather than the 5 provinces, and consider the complete elimination of dialect differences, the predicted number of cross dialect-cluster-commuters increases from 394622 to 432062, or 9.5 percent.

4 Conclusion

This paper considered the relation between dialects and commuting flows. The results suggest that a reduction of larger dialect differences is associated with an increase intra-regional commuting flows by 8 percent, in a doubly-constrained gravity model. This results takes into account travel time and the economic surroundings of both origin and destination.

These results should not be taken as proof of a direct effect of dialects on labour mobility. Although such a direct effect might partially explain the results, is seems more likely that it are underlying cultural and socio-economic factors which cause dialect differences and commuting to be correlated, and act as a barrier to labour mobility.

The magnitude of the effect we find is surprisingly large, given that we are considering mobility in a region in complete absence of formal hurdles to commuting, and are considering only small variations in a single language. The results suggest that there exist important underlying hurdles to mobility which are unlikely to be removed by public policy in the form of, say, infrastructure investment. To investigate more precisely what the underlying causes which explain these observations is an interesting venue for future research.

References

Anas, A. (1983). Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological*, **17** (1), 13–23.

Anderson, J. and Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review*, **93** (1), 170–192.

Calvó-Armengol, A. and Zenou, Y. (2005). Job matching, social network and word-of-mouth communication. *Journal of urban economics*, **57** (3), 500–522.

Falck, O., Heblich, S., Lameli, A. and Südekum, J. (2012). Dialects, cultural identity, and economic exchange. *Journal of Urban Economics*, **72** (2), 225–239.

Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics*, **97** (1), 76–85.

Ginsburgh, V. and Weber, S. (2011). *How many languages do we need?: The economics of linguistic diversity*. Princeton University Press.

Glaeser, E. L., Henderson, V. and Inman, R. P. (2000). The future of urban research: nonmarket interactions [with comments]. *Brookings-Wharton papers on urban affairs*, pp. 101–149.

Goeman, A. and Taeldeman, J. (1996). Fonologie en morfologie van de Nederlandse dialecten. een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval*, **48**, 38–59.

Guiso, L., Sapienza, P. and Zingales, L. (2009). Cultural biases in economic exchange? *The Quarterly journal of economics*, **124** (3), 1095–1131.

Ioannides, Y. M. and Soetevent, A. R. (2006). Wages and employment in a random social network with arbitrary degree distribution.

Melitz, J. (2008). Language and foreign trade. *European Economic Review*, **52** (4), 667–699.

Persyn, D. and Torfs, W. (2015). A gravity equation for commuting with an application to estimating regional border effects in Belgium. *Journal of Economic Geography*.

SILVA, J. and TENREYRO, S. (2006). The log of gravity. *The Review of Economics and Statistics*, **88** (4), 641–658.

WIELING, M., HEERINGA, W. and NERBONNE, J. (2007). An aggregate analysis of pronunciation in the Goeman-Taeldeman-van Reenen-Project data. *Taal en Tongval*, **59** (1), 84–116.

WILSON, A. (2010). Entropy in urban and regional modelling: Retrospect and prospect. *Geographical Analysis*, **42** (4), 364–394.

Research Center for Regional Economics
Waaistraat 6 - bus 3550
3000 Leuven, Belgium

Damiaan Persyn
damiaan.persyn@gmail.com