Sound Change Estimation in Netherlandic Regional Languages

Buurke, Raoul Sergio Samuel Jan; Wieling, Martijn

Published in:
Taal en Tongval

DOI:
10.5117/tet2023.1.002.buur

Amsterdam University Press

TAAL & TONGVAL 75.1 (2023) 7-28
HTTPS://DOI.ORG/10.5117/TET2023.1.002.BUUR

Sound Change Estimation in Netherlandic Regional Languages: Reducing Inter-Transcriber Variability in Dialect Corpora

Raoul Sergio Samuel Jan Buurke
Rijksuniversiteit Groningen
raoul.buurke@rug.nl

Martijn Wieling
Rijksuniversiteit Groningen

Abstract
Large phonetic corpora are frequently used to investigate language variation and change in dialects, but these corpora are often constructed by many researchers in a collaborative effort. This typically results in inter-transcriber issues that may impact the reliability of analyses using these data. This problem is exacerbated when multiple phonetic corpora are compared when investigating real time dialect change. In this study, we therefore propose a method to automatically and iteratively merge phonetic symbols used in the transcriptions to obtain a more coarse-grained, but better comparable, phonetic transcription. Our approach is evaluated using two large phonetic Netherlandic dialect corpora in an attempt to estimate sound change in the area in the 20th century. The results are discussed in the context of the available literature about dialect change in the Netherlandic area.

Keywords: aggregated sound change, phonetic transcriptions, reducing inter-transcriber issues, Levenshtein distance, Netherlandic language area

1 Background[1]

To investigate pronunciation variation and change, researchers often rely on large data collections of transcribed speech. Depending on their interests, researchers typically investigate a few linguistic variables in detail (the dialectological tradition; see Boberg, Nerbonne, and Watt 2018) or many at the same time to obtain an aggregate view of dialect variation (the dialectometric tradition; see Nerbonne and Kretzschmar 2013). The increase in computing power in recent decades has enabled large-scale quantitative analyses, including novel dialectometric approaches, for investigating pronunciation variation and change (see Wieling and Nerbonne 2015 for an overview). A dialectometric approach has the advantage of avoiding researcher bias due to not having to select a small set of specific linguistic variables, but the reliance on large data collections may still present some problems. Specifically, inter-transcriber variability may be an issue when many transcribers are involved in the data collection process. Furthermore, comparing different corpora in order to investigate language change may be hampered by different transcription practices in the corresponding projects.

In this study, we focus on two corpora investigating dialect variation in the Netherlands and Flanders: the Reeks Nederlandse Dialectatlassen (RND; Blancquaert and Pée 1930) and the Goeman-Taeldeman-van Reenen Project (GTRP; Taeldeman and Goeman 1996). Goeman and Taeldeman indicated that their project, for which data were collected between 1980 and 1995, may be seen as a refinement of the RND, for which data were collected between 1922 and 1975.

Both datasets are described in detail below, but it is worth noting that inter-transcriber issues have been specifically noted for the GTRP before. Hinskens and Van Oostendorp (2006) analyzed nasal-plosive clusters (e.g. [nd]/[nt]) in the GTRP and noticed that they were differently transcribed by transcribers from different areas (contrasting, for example, transcriptions made by transcribers for Frisian varieties with transcriptions made for Dutch varieties). In addition, the set of International Phonetic Alphabet (IPA) symbols used by the GTRP transcribers in the Netherlands is much larger than the set used by the transcribers in Flanders (Van Oostendorp 2007; Wieling, Heeringa and Nerbonne 2007).

Inter-transcriber differences are not a novel phenomenon. They are a well known problem for phonetic transcriptions, and such differences are only expected to increase when transcribers have different linguistic backgrounds. Even among experienced transcribers, the inter-rater reliability is rarely higher than 80% (Amorosa et al. 1985), and it is even lower when narrow phonetic transcriptions are made as opposed to broader ones (Shriberg and Lof 1991). Naturally, these effects do not devalue phonetic transcriptions, but their detrimental effects should be minimized, if possible.

The degree to which inter-transcriber issues are a problem depends on the scale and the type of analysis. A synchronic study of a comparatively small geographical area is less likely to suffer extensively from such systematic inconsistencies than, for example, a diachronic (real-time) study of a large geographical area. In the latter scenario, more transcribers are usually involved, and the increased territory also introduces more linguistic variation to consider as a transcriber. These factors substantially increase the difficulty of the task. In this study, we focus on exactly this type of scenario. Consequently, we will try to alleviate inter-transcriber issues in a relatively extreme case. If the approach proves effective, it can be fine-tuned for less extreme cases.

Specifically, in this study, we attempt to estimate the magnitude of sound change across the Netherlandic area, by comparing the GTRP and the RND datasets. In order to compare two datasets which differ in how narrow the phonetic transcriptions are, we successively merge similar phonetic symbols until a shared set of phonetic symbols remains. While information about variation is potentially lost prior to the final analysis through this procedure, not all variation is equally meaningful. By merging symbols in the phonetic transcriptions that are most similar to each other, we effectively generate a less narrow phonetic transcription with likely fewer inter-transcriber issues. This approach is described in detail in the methods section, where we also explain how we estimate sound change from the resulting transcriptions.

In order to evaluate the results that we obtain after simplifying the phonetic complexity, it is necessary to compare them to what is known about the regional languages within the geographical area. In the Netherlands and Flanders, our focus area, variants of three main language families are spoken: Frisian, Low Saxon, and Low Franconian. Dialects of these families are influenced by the prestigious standardized variety of Dutch (Standard Dutch; Heeringa and Nerbonne 2000; Heeringa and Hinskens 2015). We expect this vertical convergence to be the main driver of sound change (cf. Heeringa and Hinskens 2015).

We expect that the level of sound change is going to vary between the language families of interest, as well as within these language families. For Frisian varieties, we expect sound change to be relatively low, because the language is taught in school, shows a relatively high intergenerational transmission (Driessen 2012) and is protected politically (Hoekstra 2003). Both Low Saxon and Limburgish (a dialect group within the Low Franconian language family) are recognized as official regional languages in the Netherlands under the European Charter for Regional or Minority Languages (ECRML).[2] We expect a pattern similar to Frisian for Limburgish, due to its stable intergenerational transmission (Driessen 2006), which is unique within the Low Franconian family. For Low Saxon, however, we expect a much larger change (towards Standard Dutch). While the regional language is recognized under the ECRML and positive attitudes towards the language exist (Ter Denge 2012), the intergenerational transmission has declined rapidly (Driessen 2012; Bloemhoff et al. 2013).

As the Low Franconian area is so large, we also provide some predictions about other dialect groups within the family. A small to moderate amount of change is expected for Brabantish and Zeelandic. While these language varieties are not recognized under the ECRML, and intergenerational transmission is not very high (for Brabantish even very low; Driessen 2012), they are nowadays already quite similar to Standard Dutch, which limits the amount of possible sound change. We expect the level of sound change for varieties in Flanders[3] to be in between Low Saxon (comparatively much change) and Limburgish or Frisian (comparatively little change), because variable rates of dialect leveling have been noted for these varieties in recent decades (e.g. Vandekerckhove 2013; Taeldeman 2013; Swanenberg and Van Hout 2013). The West Flemish varieties in the area likely exhibit the least change and more eastern varieties more change, due to differences in intergenerational transmission (Vandekerckhove 2013; Taeldeman 2013).

2 Methods

2.1 Data

As mentioned before, we analyze the phonetic transcriptions of two large dialect corpora across the Netherlands and Flanders: the Reeks Nederlandse Dialectatlassen (RND) and the Goeman-Taeldeman-van Reenen Project (GTRP). The RND consists of a series of 16 dialect atlases, which were published between 1923 and 1982. Researchers visited informants in 1956 locations and asked them to translate 141 (Flemish) Dutch sentences into their local dialects. The target informants were non-mobile older rural males (NORMs), who were seen as desirably conservative language users (Chambers and Trudgill 1980). The phonetic transcriptions for these sentences (split into single words) and locations are partially available in digitized form (see Gabmap; Leinonen, Çöltekin and Nerbonne 2016). The available subset we use in this study consists of the phonetic transcriptions of 166 target words across 347 locations.

The GTRP, as successor to the RND, employed a similar methodology. The data were largely collected between 1980 and 1995 (see Taeldeman and Goeman 1996 for details), and the target informants were also NORMs. For this project, however, the informants were largely tasked with translating single (or a sequence of a few) words as opposed to sentences. In total, 1876 items were translated by informants in 613 locations. The translations were recorded, and later transcribed phonetically. In this study, we use the subset extracted by Wieling, Heeringa and Nerbonne (2007), which consists of phonetic transcriptions of 562 (single) words across the 613 locations.

In order to estimate the change of local variants between the RND and GTRP time periods, we select locations and target words that overlap between these corpora. The 192 overlapping locations are presented in Figure 1, and the 61 Dutch target words that overlap are presented in Table 1.

Figure 1. Overlapping locations (192) between the RND and GTRP.

2.2 Levenshtein distance

To quantify the differences between the phonetic transcriptions, as well as for reducing the phonetic inventories of the corpora, we use a variant of the Levenshtein distance (Levenshtein 1966) that has been optimized for linguistic purposes. The inputs of this algorithm are two phonetic strings, and the result is a count of how many binary operations are *minimally* necessary to turn one transcription into the other. There are three possible operations to achieve this: insertions, deletions, or substitutions of two sounds simultaneously. An example is given in Table 2, for which the transcriptions are taken from Heeringa and Hinskens (2015).

Table 1. Overlapping words (61) between the RND and the GTRP.

bakken	dun	kaas	op	ver
bier	duwen	komen	potten	vier
binden	eieren	koud	rijp	vijf
blauw	flauw	krijgen	saus	voor
brengen	gaan	krom	sneeuw	vuur
buigen	geld	laten	spannen	weg
doen	geweest	licht	springen	wijn
dopen	goed	maart	stenen	zee
dorsen	gras	melk	tegen	zes
dorst	groen	moe	twee	zijn
drie	hebben	nog	vader	zuur
drinken	hooi	ook	veel	zwemmen
droog				

Table 2. Example Levenshtein alignment between dialectal variations of Dutch 'straat': [stʀodə] and [stʀɔət].

	1	2	3	4	5	6	7
String 1	s	t	ʀ	o		d	ə
String 2	s	t	ʀ	ɔ	ə	t	
Operation	-	-	-	sub.	ins.	sub.	del.
Cost	0	0	0	1	1	1	1

The following operations are sufficient to transform [stʀodə] into [stʀɔət]. The first 3 sounds are already equal and therefore require no transformation. After this, the [o] and [ɔ] are substituted, the [ə] inserted, and the [t] and [d] are substituted. Finally, the remaining [ə] is deleted. This requires 4 operations in total, which is the Levenshtein distance between these transcriptions. In order to correct for phonetic sequences of different lengths (longer sequences are less likely to be identical), we divide this distance by the alignment length, so the *normalized* Levenshtein distance is 4/7 (≈57%).

Note that it is possible to achieve a transformation of the first transcription into the second in 3 operations (all substitutions), when vowels and consonants are allowed to substitute each other. This is not linguistically sensible, however, as these are different categories of sounds. We avoid such alignments by setting the associated cost of such an operation to be very high, so that vowel-consonant substitutions do not occur.

The Levenshtein distance can be optimized further for phonetic purposes. The binary weights used in the prior example treat all substitutions as equal, but it is sensible to penalize a substitution involving two very different sounds (such as an [i]-[u]-substitution) more than a substitution involving similar sounds (such as an [i]-[ɪ]-substitution). In that case, the binary weights can be adjusted to a value between 0 and 1 (instead of either 0 or 1) that reflects the distance between sounds in phonetic space. Deriving reliable weights is not a trivial problem, however, as is demonstrated in detail by Heeringa (2004, pp. 79-120).

Here we use the approach proposed by Wieling, Margaretha and Nerbonne (2012), which is based on the co-occurrence patterns of sound segments in alignments (such as the one above) of phonetic transcriptions on the basis of corpora, such as the RND and GTRP. The underlying idea is that more similar sounds will be more often substituted by each other (Wieling, Margaretha and Nerbonne 2012). This data-driven approach (applicable when the dataset from which the alignments are generated is of sufficient size) has been shown to result in meaningful sound distances (Wieling, Margaretha and Nerbonne 2012), and the Levenshtein distance algorithm incorporating these sensitive sound distances has been found to correlate well with the perception of pronunciation differences by listeners (Wieling et al. 2014). In the approach of Wieling, Margaretha and Nerbonne (2012), the sound distances between phonetic symbols X and Y are determined via pointwise mutual information (PMI; Church and Hanks 1990):

(1) $$PMI(X,Y) = log_2\left(\frac{p(X,Y)}{p(X)p(Y)}\right)$$

The numerator reflects how often the relevant sounds are aligned *together* in Levenshtein alignments on the basis of corpus data (where the pronunciations of each individual word are compared between every pair of locations). The denominator is the multiplication of the probability of each sound occurring *individually* in these Levenshtein alignments (i.e. the probability of the two sounds to align simply due to chance). When sounds align more often than would be expected on the basis of chance, the formula results in a positive value, and otherwise a negative value (the value 0 indicates that they co-occur exactly as often as expected). More details about this algorithm can be found in Wieling, Margaretha and Nerbonne (2012).

To convert the PMI scores to sound distances which are scaled between 0 and 1, we first invert the PMI values as follows:

(2) $$SndDist_{(X,Y)} = 0 - PMI_{(X,Y)}$$

And we subsequently normalize these values by scaling them between 0 and 1 as follows:

$$(3) \quad SndDistNorm_{(X,Y)} = \frac{SndDist_{(X,Y)} - min(SndDist)}{max(SndDist) - min(SndDist)}$$

As a next step, a matrix can be constructed between all the sounds in the corpus, such as the hypothetical one in Table 3. The values in this matrix represent the normalized sound distances and can, be used to weight the operations of the Levenshtein distance algorithm (cf. Wieling, Margaretha and Nerbonne 2012). This matrix of sound distances can then be used to combine phonetic inventories by merging the most similar sounds in a successive approach.

Table 3. PMI-based segment distance matrix (example).

	e	ɪ	ɛ	i	ø
e	0				
ɪ	0.018	0			
ɛ	0.019	0.022	0		
i	0.020	0.021	0.022	0	
ø	0.023	0.028	0.030	0.030	0

2.3 Combining phonetic inventories

The PMI-based segment distance matrix can also be used to (successively) determine which phonetic segments should be merged (i.e. those with the smallest distances) to make two sets of phonetic inventories more comparable. Specifically, for two separate sets of phonetic symbol inventories, which contain both overlapping sounds and distinct sounds, our procedure works as follows. For the first set, we merge each phonetic symbol (on the basis of the PMI-based segment distance matrix) with its closest alternative in the first set. The possible alternative symbols for the first set are the symbols that occur in the second set, of which the closest is chosen. After this procedure, we do the same for the sounds in the second set that do not occur in the first set. Note that after each individual phonetic symbol merger, the phonetic transcriptions are correspondingly updated, and a new PMI-based segment distance matrix is generated using these transcriptions. The end result is a shared set of (partly merged) phonetic symbols.

In this specific study, we follow a three-step approach. First, we integrate the Flemish and Netherlandic sound symbol inventories used in the GTRP (as these differed substantially; Wieling, Heeringa, and Nerbonne 2007). Second, we integrate the (resulting) GTRP sound symbol inventory and the RND sound symbol inventory. Third, as there may still be symbols in the resulting symbol inventory which occur very infrequently (which may be indicative of transcriber inconsistencies), we further merge those symbols with more frequently occurring symbols. Specifically, we require the minimum frequency of each sound symbol to be at least 1% of the total number of segments occurring in all (pairwise) transcriptions. For example, if all transcriptions together contain 1000 sound segments, any sound symbol that occurs fewer than 10 times across the corpus is merged with the most similar phonetic symbol that occurs sufficiently frequently (and phonetic transcriptions are adjusted accordingly after every merger).

2.4 Estimating sound change

After reducing the phonetic inventories according to the procedure illustrated above, we can determine the amount of sound change by calculating the Levenshtein distance (with PMI-based data-driven sound distances) between the RND and the GTRP transcriptions (using the shared phonetic symbol inventory) for each word in every location.

Note that this analysis focuses on the word level, and not on the level of individual segments. Consequently, the values may exceed 1 as the cost of the differences are summed. However, we normalize the sum of the substitution, insertion and deletion costs (determined by the PMI weights) by dividing the total sum by the length of the Levenshtein alignment. As a result, the value that is predicted for each RND-GTRP pair of transcriptions therefore lies between 0 and 1.

In order to evaluate and visualize the aggregate change across the Netherlandic area, we model sound change (aggregated per word) on the basis of geographical coordinates. More specifically, these coordinates are modeled as a two-dimensional smooth in a generalized additive model (GAM; see Wieling, Nerbonne and Baayen 2011). GAMs can be seen as an extension to linear regression, but with the capacity to deal with non-linear relationships and predictors (Wood 2017).

Due to the bounded nature of our normalized dependent variable, we use beta regression, which is suitable for predicting data in the interval (0,1). Sound change values of exactly 0 or exactly 1 are increased or decreased by a small number (10^{-6}) in order to adhere to the assumptions of the beta regression family. Likewise, instead of reporting the adjusted R^2 of the models, we report the explained deviance of the models, because this is better suited for non-Gaussian models (see Wood 2017, p. 128, for the computation). Explained deviance may be seen as a generalization of and interpreted in a similar way: a higher percentage of deviance explained reflects a better fit of the model to the data.

3 Results

3.1 Phonetic inventory reduction

The overlapping RND-GTRP corpus has a total symbol inventory of 70 IPA symbols, which are presented in the Appendix. In Table 4, we present the 32 symbols that are absent in at least one of these subsets: the Dutch part of the GTRP, the Belgian part of the GTRP, or the RND (i.e. not split by country). The check mark indicates the presence of a symbol in a particular subset.

Table 4. List of IPA symbols which are not present in all subsets.

IPA	GTRP-NL	GTRP-BE	RND
c	✓		✓
h	✓		✓
ɕ	✓		
ð	✓		
ħ	✓		
ɒ	✓		✓
ɡ	✓		✓
ɢ		✓	
ɤ	✓		
ɦ	✓	✓	
ɨ	✓		
ɪ	✓		✓
ɫ	✓	✓	
ɱ	✓	✓	
ɵ	✓	✓	
ɶ	✓		
ɷ	✓		
ɸ	✓		
ɹ	✓		
ɼ	✓		
ɾ	✓	✓	
ʁ	✓		

IPA	GTRP-NL	GTRP-BE	RND
ʉ	✓	✓	
ʊ	✓		✓
ʌ	✓		
ʍ	✓		
ʋ	✓		✓
ʎ	✓		
ʏ	✓		✓
ʝ	✓		
β	✓		
χ	✓		

First, the GTRP-NL symbol inventory (70 symbols) is merged with the GTRP-BE symbol inventory (40 symbols), by taking as reference first the GTRP-NL set and subsequently the GTRP-BE set. Using GTRP-NL as reference, a total of 29 symbols (which occurred in GTRP-NL, but not GTRP-BE) were merged to their optimal (i.e. most similar) alternative symbol. Using the GTRP-BE set as reference, a single symbol ([ɢ]) was merged with another, as this symbol did occur in GTRP-BE but not in GTRP-NL. Finally, the GTRP and RND symbol inventories are merged, by first taking the GTRP as reference, followed by the RND as reference. The resulting set of 38 symbols consisted of 14 vowels and 24 consonants.

As mentioned earlier, we continue merging symbols based on a frequency constraint. If any phonetic symbol occurs less than 1% of the time, it is replaced by its closest phonetic alternative within the GTRP-RND reduced subset. After this procedure, the final phonetic inventory consisted of 29 symbols (11 vowels, 18 consonants) and is presented in Table 5.

Table 5. Combined phonetic sound inventory showing the merged (29) symbols.

Used symbol	← Merged symbol(s)
a	
ɑ	ɤ[4]
b	β
d	ð
e	ɪ

Used symbol	← Merged symbol(s)
ə	ɨ
ɛ	æ
f	ɸ
ɣ	ħ ɡ j h ɦ ɕ
i	
k	c ʔ ɢ
l	ɫ ʎ
m	
n	ɲ ɱ
ŋ	
ø	ɶ
œ	ʏ ɵ ʌ
o	ɷ
ɔ	ʊ
p	
r	ɹ ɾ ʀ ʁ
s	ʃ ʝ ɼ
t	θ
u	ʉ
v	
w	ʊ ʍ ʋ
x	ç χ ɥ
y	
z	ʒ

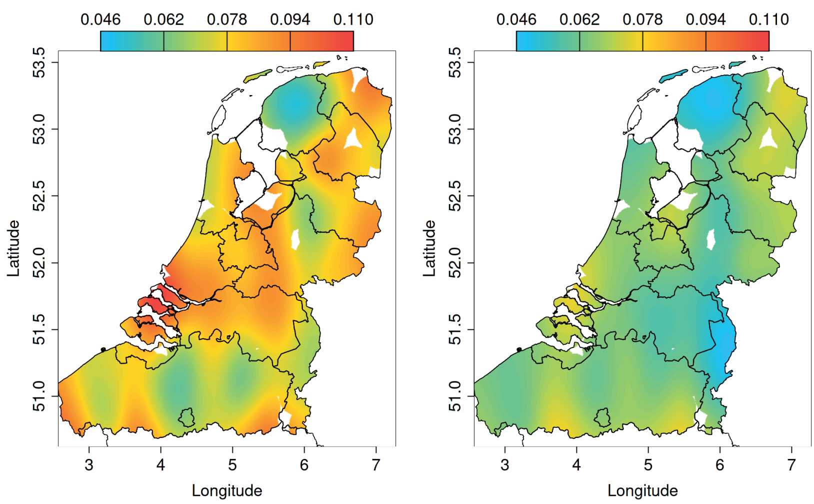
3.2 Aggregated sound change estimation

With the shared (smaller) phonetic inventory, but also with the original transcriptions for the sake of comparison, we estimate change in each location across the target words. The geographical distribution of sound change is modeled using a two-dimensional smooth on the basis of the longitude and latitude coordinates. Additionally, to account for the word-specific variability (location-specific variability is modeled by the two-dimensional smooth), a random intercept for word was included. The accompanying visualizations are presented in Figure 2. The model specification for both models is as follows (i.e. only the transcriptions and consequently the phonetic distances themselves are different for each model):

```
normalized_sound_change ~ s(longitude, latitude,
k = 50, m = 1) + s(word, bs="re")
```
[5]

The geographical smooth is significant in both models ($p < 0.01$), which indicates that the geographical distribution of change is not random. The explained deviance of the model based on the original transcriptions is 22% and 17.6% for the model based on the combined inventory.

Note that the values are proportions of change, so a value of 0.2 indicates an average change of 20% across all words in that area. The sound change estimations are generally higher when the original transcriptions are used than when the combined transcriptions are used, which stems from an inherently higher probability of differences.



(a) Original inventory. (b) Combined inventory.

Figure 2. Proportions of sound change predicted on the basis of geography (using PMI-based weights). Red: more change. Blue: less change.

There are several noticeable differences between the geographical patterns in the visualizations. For example, for the model based on the original inventory (Figure 2a), there is relatively much change around the border between the provinces Gelderland and North Brabant. For the corresponding model based on the combined transcriptions (Figure 2b), this area shows relatively little change. Moreover, some areas have become slightly more homogeneous after combining the symbol inventories, such as the Low Saxon area around the Groningen–Drenthe border and the Overijssel–Gelderland border.

The patterns in Flanders are less affected by the combination of the symbol inventories, which was expected, because symbols were combined in the GTRP data concerning the Netherlands. The main difference before and after combining the symbol inventories is the area in the southwest of Flanders. This was an area of relatively much change before the procedure, but is an area of little to moderate change after the procedure. The areas of stability seem consistent before and after the procedure, because the areas of relatively much change around the border with Wallonia remain clearly distinct.

There are also areas that overlap between the models. The relative stability of local varieties in Fryslân and Limburg can still be observed after reducing the phonetic inventory. Similarly, the relatively high level of change in the Low Saxon area can be identified in both models. The area around South Holland and Zeeland consistently shows a relatively high level of change.

4 Discussion

It is difficult to imagine a study into the sound change of dialects (at an aggregate level) without the reliance on large linguistic corpora, such as the RND and GTRP. Gathering the relevant data is a costly endeavor, in terms of both time and monetary expenses, and impossible to accomplish for a single researcher. Consequently, such large corpora typically rely on many researchers contributing data, and potential transcriber inconsistencies may present a problem (such as when comparing the Netherlandic and Flemish part of the GTRP; Wieling, Heeringa and Nerbonne 2007). For this reason, this study has presented a method that may mitigate the influence of inter-transcriber inconsistencies. While the method, which relies on co-occurrence patterns of sound segments in the corpora (and builds upon the work of Wieling, Margaretha and Nerbonne 2012), was tested on data from the Netherlandic area, it may be applied to different areas as well. Specifically for the Netherlandic area, however, we posed several predictions, which we use to evaluate the proposed method.

After combining the phonetic inventories to a common one, we created two geographical models to interpret sound change (at an aggregate level) across the area. These models showed that the patterns were not geographically random. We discuss the patterns in the Netherlands and Flanders separately, as the former are much more affected by combining the phonetic inventories than the latter (which was due to a smaller phonetic inventory being used in Flanders). When we observe the patterns in the Netherlands, most findings in the models were in line with what was expected. The Frisian and Limburgish varieties turned out to be relatively resistant to change, while the variants spoken in the Low Saxon area were most prone to change. Most of the other areas in the Netherlands fell in between these two extremes.

A noticeable exception is the area on the border of the provinces Zeeland and South Holland, which showed a relatively high level of change, for which we do not have a clear explanation. Upon closer inspection, it seems that this area concerns a particularly large range in years between the RND and GTRP recordings, which may be a contributing factor. At the same time, the area in Fryslân that consistently showed little change also concerns an area where there were many years between the datasets, so it is likely that more factors contribute to these patterns.

As the changes in the GTRP transcriptions in the Netherlands were much more substantial, the areas in Flanders were much less affected by reducing the symbol inventories. One consistent finding between the models was that the area in West Flanders showed relatively little change. The West Flemish varieties are subject to dialect loss, but it has been observed that they remain relatively prolific (Vandekerckhove 2013), which seems to be consistent with our findings. There is also a southern area in Flanders that showed relatively much change (i.e. near the level of the Low Saxon area), which we did not express predictions about. It is striking that this change is clearly on the border with Wallonia, which may potentially be explained by the linguistically tumultuous history of that area in the 20th century, during which Flanders and Wallonia became monolingually Dutch and French in the 1930s (see Willemyns 2002 for details about directly identifiable border effects).

Comparing the results on the basis of the original inventory to those on the basis of the combined inventory reveals that the patterns are more clear (globally) and less noisy (locally) when using the latter approach than the former. The combined inventory approach seems therefore capable of revealing the most important (and robust) areas of change. In some further analyses (not reported here), we tested whether the patterns changed considerably after reducing the inventories even further, but the patterns remained stable until fewer than 10 symbols were left in the inventory, which demonstrates that the observed patterns of change are quite robust. Overall, the findings indicate that the method can be used to extract the most prominent patterns from transcription data that are influenced less by inter-transcriber issues, as long as sufficient care is taken to determine a suitable stopping criterion for the merging of sound symbols in a language area.

However, we also stress that the using the proposed method of combining symbol inventories does not inherently reflect an ideal analysis. It is useful for reducing complexity in a meaningful way, but determining the appropriate and optimal level of complexity is not trivial. The frequency constraint, for example, was rather arbitrary and could have been chosen differently. For the low frequency symbols, the combination of symbols does not influence the results much, because we know from previous work that aggregate analyses that are conducted at the word-level are not strongly altered by such small differences (see e.g. Wieling, Prokić and Nerbonne 2009). It is even possible that the situation of language variation is inherently noisy, which means that forcefully reducing the complexity produces results that are further away from reality instead. There is no obvious reason to avoid the method altogether, but it is clearly important to be informed of the language area to which the method is applied, so that any known salient differences are not removed by accident. To further assess this, simulation studies might be insightful.

There are a few shortcomings that we still want to explicate here. One potentially problematic point that we have not discussed so far is that the RND data collection spans many years compared to the GTRP. Consequently, in some locations, change may be tracked during a much longer time period (i.e. where the RND data were collected at the beginning of the project) than in others (i.e. where the RND data were collected towards the end of the project). Of course, this may have impacted the observed patterns of sound change. However, for the locations where the difference in years is smallest (around the middle of the Netherlands and the Veluwe), we do not see the smallest change. Similarly, there are relatively many years between the RND and GTRP recordings for Fryslân, where we see relative stability. Consequently, at the very least, the observed sound change patterns are not a direct consequence of the difference in recording years.

Some of the more general shortcomings of more dialectometric studies also apply to this study. Most importantly, only data from a single informant in each location were analyzed. It requires little imagination to see how this may influence the results, particularly in regions where major language families meet. Appelscha, which lies in southeastern Fryslân, is historically a Low Saxon area, but over the years many Frisians have migrated to this village as well. Upon closer inspection of the data, it turned out that the GTRP speaker followed a very clear Frisian sound pattern, while the RND speaker followed a more typical Low Saxon pattern. Leaving out this data point did not significantly influence the results, but such issues can be avoided better if data from several speakers are available from any particular location.

One final point for improvement stems from the fact that we only looked at the role of geography on sound change. Dialectologists typically lament the (admittedly frequent) shortcoming of dialectometric studies to investigate only a few determinants of language change (Wieling and Nerbonne 2015), despite a wealth of sociolinguistic research suggesting potential (social) determinants of change which could be incorporated in dialectometric studies. It is desirable to steer towards models in future endeavors that incorporate extralinguistic variables (such as gender and social class; Labov 1972, 1994) as well as broader categorical linguistic information (such as frequency and word classes; Bybee and Hopper 2001; Bybee 2002). However, since the main objective of this study was to evaluate a novel approach to compare phonetic transcriptions between different datasets, we have not attempted this here. Nevertheless, we acknowledge that our models only show a limited view on the propagation of sound change. This is also reflected by the low amount of explained deviance of the models. For the model based on the original transcriptions, only 2.3% of the deviance is explained by the geographical predictor, while the random effect for word explains 19.7% of the deviance. For the model based on the transcriptions with the combined inventory, only 1.9% of the deviance is explained by the geographical predictor, while the random effect for word explains 15.7% of the deviance.

In sum, we hope to have provided a potentially attractive solution to generate more coarse-grained phonetic transcriptions in a data-driven way, when this is desirable due to noisy data. As we have shown, our approach can be used to make different phonetically transcribed datasets better comparable, and to potentially reduce transcriber-related differences.

References

Amorosa, H., U. Von Benda, E. Wagner and A. Keck. 1985. 'Transcribing phonetic detail in the speech of unintelligible children: A comparison of procedures' *International Journal of Language & Communication Disorders* 20 (3): 281–287.

Blancquaert, Edgard, and Willem Pée. 1930. *Reeks Nederlandse Dialectatlassen*. De Sikkel.

Bloemhoff, Henk, Hermann Niebaum, Jan Nijen Twilhaar and Harrie Scholtmeijer. 2013. 'Recent evolutions in the position and structure of Low Saxon. With an excursus on the new polders in Flevoland'. In *Dutch*, edited by Frans Hinskens and Johan Taeldeman, 495–511. Language and Space. Walter de Gruyter.

Boberg, Charles, John Nerbonne and Dominic Watt. 2018. *The Handbook of Dialectology*. John Wiley & Sons.

Bybee, Joan. 2002. 'Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change'. *Language variation and change* 14 (3): 261–290.

Bybee, Joan, and Paul Hopper. 2001. *Frequency and the emergence of linguistic structure.* Vol. 45. John Benjamins Publishing.

Chambers, Jack, and Peter Trudgill. 1980. *Dialectology*. Cambridge University Press.

Church, Kenneth Ward, and Patrick Hanks. 1990. 'Word association norms, mutual information, and lexicography'. *Computational linguistics* 16 (1): 22–29.

Driessen, Geert. 2006. 'Ontwikkelingen in het gebruik van streektalen en dialecten in de periode 1995–2003'. *Toegepaste taalwetenschap in artikelen* 75 (1): 103–113.

Driessen, Geert. 2012. 'Ontwikkelingen in het gebruik van Fries, streektalen en dialecten in de periode 1995-2011'. Nijmegen: ITS.

Heeringa, Wilbert. 2004. 'Measuring dialect pronunciation differences using Levenshtein distance'. PhD diss., University of Groningen.

Heeringa, Wilbert, and Frans Hinskens. 2015. 'Dialect change and its consequences for the Dutch dialect landscape'. *Journal of Linguistic Geography* 3 (1): 20–33.

Heeringa, Wilbert, and John Nerbonne. 2000. 'Change, convergence and divergence among Dutch and Frisian'. In *Philologia Frisica Anno*, edited by P. Boersma, Ph. H. Breuker, L. G. Jansma and J. Van Der Vaart, 88–109.

Hinskens, Frans, and Marc van Oostendorp. 2006. 'De palatalisering en velarisering van coronale nasaal-plosief clusters in GTR. Talige, dialectgeografische en onderzoekerseffecten'. *Taal en Tongval* 58: 103–122.

Hoekstra, Eric. 2003. 'Frisian: Standardization in progress of a language in decay'. In *Germanic Standardizations: Past to Present*, edited by Ana Deumert and Wim Vandenbussche, 193–209. John Benjamins Publishing Company.

Labov, William. 1972. *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, William. 1994. *Principles of Linguistic Change. Volume 1: Internal factors*. Wiley-Blackwell.

Leinonen, Therese, Çağrı Çöltekin and John Nerbonne. 2016. 'Using Gabmap'. *Lingua* 178: 71–83.

Levenshtein, Vladimir. 1966. 'Binary codes capable of correcting deletions, insertions, and reversals'. *Soviet Physics Doklady* 10 (8): 707–710.

Nerbonne, John, and William Kretzschmar. 2013. 'Dialectometry++'. Literary and *Linguistic Computing* 28 (1): 2–12.

Shriberg, Lawrence, and Gregory Lof. 1991. 'Reliability studies in broad and narrow phonetic transcription'. *Clinical Linguistics & Phonetics* 5 (3): 225–279.

Swanenberg, Jos, and Roeland van Hout. 2013. 'Recent developments in the mid southern dialects'. In *Dutch*, edited by Frans Hinskens and Johan Taeldeman, 319–335. Language and Space. Walter de Gruyter.

Taeldeman, Johan. 2013. 'Recent developments in the East Flemish dialect area'. In *Dutch*, edited by Frans Hinskens and Johan Taeldeman, 255–276. Language and Space. Walter de Gruyter.

Taeldeman, Johan, and Antonie Goeman. 1996. 'Fonologie en morfologie van de Nederlandse dialecten: Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten'. *Taal en Tongval* 48: 38–59.

Ter Denge, G. J. M. 2012. 'Linguae Intergermanica: The Use of Low Saxon, English, Dutch, German, and Receptive Multilingualism in Northern Dutch–Northern German Communication'. Master's Thesis, Utrecht University.

Van Oostendorp, Marc. 2007. 'Kenmerkeconomie in de GTRP-database'. *Taal en Tongval* 18 (1): 148–159.

Vandekerckhove, Reinhild. 2009. 'Dialect loss and dialect vitality in Flanders'. *International Journal of the Sociology of Language* 2009 (196–197): 73–97.

Vandekerckhove, Reinhild. 2013. 'Southwestern dialects today'. In *Dutch*, edited by Frans Hinskens and Johan Taeldeman, 194–210. Language and Space. Walter de Gruyter.

Wieling, Martijn, Jelke Bloem, Kaitlin Mignella, Mona Timmermeister and John Nerbonne. 2014. 'Measuring foreign accent strength in English: Validating Levenshtein distance as a measure'. *Language Dynamics and Change* 4 (2): 253–269.

Wieling, Martijn, Wilbert Heeringa and John Nerbonne. 2007. 'An aggregate analysis of pronunciation in the Goeman-Taeldeman-van Reenen-Project data'. *Taal en Tongval* 59 (1): 84–116.

Wieling, Martijn, Eliza Margaretha and John Nerbonne. 2012. 'Inducing a measure of phonetic similarity from pronunciation variation'. *Journal of Phonetics* 40 (2): 307–314.

Wieling, Martijn, and John Nerbonne. 2015. 'Advances in dialectometry'. *Annual Review of Linguistics* 1 (1): 243–264.

Wieling, Martijn, John Nerbonne and Harald Baayen. 2011. 'Quantitative social dialectology: Explaining linguistic variation geographically and socially'. *PloS ONE* 6 (9): e23613.

Wieling, Martijn, Jelena Prokić and John Nerbonne. 2009. 'Evaluating the pairwise string alignment of pronunciations'. In *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education* (*LaTeCH–SHELT&R 2009*), edited by Lars Borin and Piroska Lendvai, 26–34.

Willemyns, Roland. 2002. 'The Dutch-French language border in Belgium'. *Journal of Multilingual and Multicultural Development* 23 (1-2): 36–49.

Wood, Simon N. 2017. *Generalized additive models: An introduction with R*. 2nd ed. Chapman / Hall/CRC.

Notes

1. We would like to thank the anonymous reviewers for their comments on previous versions of this article. Only the authors are responsible for any remaining inaccuracies.
2. Signatories of the ECRML choose to which degree languages are protected and promoted. Frisian is protected to a larger degree than Limburgish and Low Saxon in the Netherlands, for example.
3. We should note here that we do not expect convergence to Netherlandic Standard Dutch for these varieties, but instead to Belgian Standard Dutch (see Vandekerckhove 2009 for a discussion on dialect-standard dynamics in Flanders).
4. The symbol presented here is the most recent version representing the sound (close-mid back unrounded vowel). Newer fonts do not have the old symbol available, but prior to 1989 (and in the phonetic transcriptions used here) the symbol looked like this: https://upload.wikimedia.org/wikipedia/commons/thumb/b/b2/Latin_letter_small_capital_Gamma.svg/320px-Latin_letter_small_capital_Gamma.svg.png.
5. The parameters *k* and *m* specify the number of basis dimensions for the smooth and a first-derivative penalty to avoid excessive extrapolation of the data. See https://cran.r-project.org/web/packages/mgcv/mgcv.pdf for further details.

Appendix

Table 6. Occurrence of all phonetic symbols in each subset.

IPA	GTRP-NL	GTRP-BE	RND
a	✓	✓	✓
b	✓	✓	✓
c	✓		✓
d	✓	✓	✓
e	✓	✓	✓
f	✓	✓	✓
h	✓		✓
i	✓	✓	✓
j	✓	✓	✓
k	✓	✓	✓
l	✓	✓	✓
m	✓	✓	✓
n	✓	✓	✓
o	✓	✓	✓

IPA	GTRP-NL	GTRP-BE	RND
p	✓	✓	✓
r	✓	✓	✓
s	✓	✓	✓
t	✓	✓	✓
u	✓	✓	✓
v	✓	✓	✓
w	✓	✓	✓
x	✓	✓	✓
y	✓	✓	✓
z	✓	✓	✓
æ	✓	✓	✓
ɕ	✓		
ð	✓		
ø	✓	✓	✓
ħ	✓		
ŋ	✓	✓	✓
œ	✓	✓	✓
ɑ	✓	✓	✓
ɒ	✓		
ɔ	✓	✓	✓
ç	✓		
ə	✓	✓	✓
ɛ	✓	✓	✓
ɡ	✓		✓
ɢ		✓	
ɣ	✓	✓	✓
ɤ	✓		
ɥ	✓		
ɦ	✓	✓	
ᵻ̇	✓		
ɪ	✓		✓
ᵻ	✓	✓	
ɱ	✓	✓	
ɲ	✓	✓	✓
ɵ	✓	✓	
ɶ	✓		
ɷ	✓		
ɸ	✓		
ɹ	✓		

IPA	GTRP-NL	GTRP-BE	RND
ɤ	✓		
ɾ	✓	✓	
ʀ	✓	✓	✓
ʁ	✓		
ʃ	✓	✓	✓
ʉ	✓	✓	
ʊ	✓	✓	✓
ʋ	✓		✓
ʌ	✓		
ʍ	✓		
ʎ	✓		
ʏ	✓		✓
ʒ	✓	✓	✓
ʔ	✓	✓	✓
ʝ	✓		
β	✓		
χ	✓		

About the authors

Raoul Buurke is a PhD candidate at the University of Groningen.

Martijn Wieling is Associate Professor and Professor by special appointment of Low Saxon / Groningen Language and Culture at the University of Groningen.