MARJOLEINE SLOOS AND MARC VAN OOSTENDORP

THE RELATIONSHIP BETWEEN PHONOLOGICAL AND GEOGRAPHICAL DISTANCE
UMLAUT ON THE DIMINUTIVE IN DUTCH DIALECTS

Abstract[1]

Recent years have witnessed a surge of interest in the application of modern methodology to the study of geographic language variation. One instance of this, applied to the Dutch language area, is the topic of dialectometry. It seems fair to say that the measurements in most dialectometrical projects are fairly superficial from a linguistic point of view. In this article, we propose a view of dialect distance which is radically different both in its methodology and in its goals. We present a view in which the distances between *grammars* are computed rather than the differences between words or constructions, which we view as products of the grammars. In particular, we consider the phenomenon of umlaut in diminutive forms. Umlaut, a phenomenon in which back vowels become front in certain morphological contexts, is virtually absent in the western parts of the Dutch speaking area, but fully productive in many parts of the east. We show how this east-west transition can be elegantly described as a gradual change in grammars.

1 Introduction

Recent years have witnessed a surge of interest in the application of modern methodology to the study of geographic language variation. One

[1] Contactgegevens auteurs: Marjoleine Sloos, Freiburg University, e-mail: marjoleine.sloos@frequenz.uni-freiburg.de.
Marc van Oostendorp, Leiden University and Meertens Instituut, e-mail: marc.van.oostendorp@meertens.knaw.nl.

instance of this, applied to the Dutch language area, is the topic of dialectometry, in the work of John Nerbonne and his coworkers (cf. Heeringa 2004, Nerbonne and Kretzschmar 2006, Spruit 2008, to mention just a few relevant examples). This work uses refined statistical methods in order to measure the relative distance between dialects, which in turn allows us to compare this distance to a variety of other factors which can also be measured with some precision – such as socio-economic difference, geographic distance, or even linguistic distance as measured along some other dimension (e.g. pronunciation vs. syntax).

It seems fair to say that the measurements in these projects are always fairly superficial from a linguistic point of view. In Heeringa (2004) for instance, differences in pronunciation are measured using so-called Levenshtein distances (Levenshtein 1966), as defined on phonetic transcriptions of individual words, which are not directly related to the underlying phonological system. Spruit (2008) analyses syntactic constructions, which are obviously a little more abstract, but still the constructions are analysed as atomary building blocks: a dialect either has these constructions or lacks them. Spruit points out that this methodology could subsequently be used to find (implicational) relations between such constructions, which might benefit the theorist who is interested in setting up a grammatical analysis. Also in this work, however, it is not the direct comparison of grammars which defines the distance but rather the comparison of products of the grammar.

The latter type of model might allow for a more direct interaction with linguistic theory. In this article, we propose a view of dialect distance which is radically different from the one just outlined both in its methodology and in its goals. As a matter of fact, the differences are so radical that the two views seem to complement each other rather than compete: they serve different ends, and could therefore be used side by side. We present a view in which the distances between *grammars* are computed rather than the differences between words or constructions, which we view as products of the grammars. In our view, grammars are formal objects, and this makes it possible to give a precise definition, and even a quantification, of differences between grammars. In particular, we show that it is easy to develop such quantification in an Optimality Theoretic

framework (Prince and Smolensky 1993/2004) and we apply this to three sets of dialects, showing that geographic distance corresponds to grammatical distance in each of these cases. The choice for Optimality Theory is motivated by the fact that it is a grammatical formalism for describing language variation that enjoys a rather high level of formalisation, which makes it relatively easy to calculate. No doubt similar exercises could be performed on other formal theories of grammar; their relative success in handling language-geographical phenomena could even be tested on the basis of their success in describing systems of multiple dialects.

Almost by necessity given our current state of knowledge, sophisticated statistics plays a much less prominent role in our approach. Comparing grammars for a number of dialects implies that one has to set up such grammars for a number of dialects, and this involves a lot of analytical work – as anybody will know who has tried to describe even a fragment of a grammar for some single language variety in any formal framework. For this reason, we have to restrict ourselves, both by selecting only a very limited set of phenomena and secondly by carrying out the analysis for only a handful of dialects. This in turn makes statistical analysis basically irrelevant.

In this article, we present a case study of the type of approach we would like to propose. In particular, we consider the phenomenon of umlaut in diminutive forms. Umlaut, a phenomenon in which back vowels become front in certain morphological contexts, is virtually absent in the western parts of the Dutch speaking area, but fully productive in many parts of the east. It enjoys high productivity particularly in diminutives. Furthermore, it is not the case that one isogloss separates umlauting from non-umlauting areas, but rather, there is a slow transition in which an increasing number of diminutive forms lose the umlauting property. We show how this slow transition can be elegantly described as a gradual change in grammars, travelling along the east-west dimension. Furthermore, the success of this grammatical approach shows a shortcoming of the other, linguistically more superficial approaches: the latter can only account for the fact that fewer and fewer nouns get umlauted, but it cannot explain why groups of words lose their umlauting capacity as a group.

2 Phonological distance and geographic distance

Optimality Theory (OT) arguably has been the dominant framework for formal phonological analysis in the past fifteen years. Within this theory, a grammar is defined as a ranking of a set of constraints. These constraints are universal; at least in classical versions of the theory such as the one we adopt here (Prince and Smolensky 1993/2004 and much related work). For this reason, the only systematic differences between languages or language varieties are ranking differences between universal constraints. It is a clear advantage of OT for our purposes that it has this very narrow and precise restriction on the locus of variation.

We propose to compare the grammars of adjacent dialects in terms of 'minimal ranking differences'. These can be defined in the following way:

(1) Two OT grammars *x* and *y* are at a *minimal ranking difference* iff reranking of two adjacent constraints in the ordering of A gives us the ordering of B.

If dialect *x* has the constraint ranking A » B » C and dialect *y* has the constraint ranking A » C » B, this can be considered as a minimal difference. The grammar *z* = B » A » C is also at a minimal ranking distance of *x*. However *y* and *z* are not at a minimal distance, since reranking of two adjacent constraints does not suffice. Now suppose we define the distance between two grammars in the following way:

(2) A path between two grammars x_0, x_n is a sequence $x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow ... \rightarrow x_n$, such that x_0, x_1, ... are grammars, and for any two grammars $x_1 \rightarrow x_2$, x_1 and x_2 are at a minimal ranking difference

(3) The distance between two grammars x_0, x_n is the length of the shortest path that can be constructed between x_0 and x_n

Minimal ranking differences are assigned one point, but in comparing dialect *y* with dialect *z*, two points are assigned, because two minimal reorderings are needed to change *y* into *z*, for instance: A » C » B → A » B » C → A » C » B. In this paper, we use this computation to define

phonological distance between dialects concerning umlaut on the diminutive.

We will need to make one refinement to this computation. We will find that in some dialects there is variation for certain phenomena. Such variation will usually mean that a location shows the grammatical phenomena of both of its neighbours. Within Optimality Theory, it is usually assumed (after Anttila 1997) that variation corresponds to crucial unranking of constraints, notated as A, B. This notation basically means that the rankings A » B and B » A can both be used. Given this, it seems reasonable to count the distance from A » B to A, B as well as that from B » A to A, B as 0.5.

Once phonological distance has been computed, it can be related to geographical distance, or for that matter to any other independent variable. We will restrict ourselves to geographical distance since it is rather straightforward. A word of caution is nevertheless in order. Since it is well-known that other geographical factors, such as, say, the intervention of rivers, and also socio-economic factors play a significant role in dialect geography, the relation between geographical and phonological distance is not expected to be a one-to-one relation between the number of kilometres and the number of constraint rerankings. However, we hypothesize that a topological relationship could be attested; if geographic location *y* is in between *x* and *z*, we expect the grammar of *y* to be also between *x* and *z* according to the computation just defined.

We will test this hypothesis on three strips of dialects from east to west, in Overijssel, Gelderland and Zuid-Limburg/Noord-Brabant; these are depicted in Figure 1.

Figure 1. All dialects investigated

For this study we used the so-called GTRP (Goeman-Taeldeman-Van Reenen Project) database of the Meertens Institute which is accessible at http://www.meertens.knaw.nl/mand/database/. Dialectological data of 613 places in the Dutch speaking area, including in Vlaanderen, are available, of which a selection was made of 29 places, in such a way that they form three strips on equal distances in (from north to south) the provinces of Overijssel, Gelderland and Limburg/Noord-Brabant as can be seen in figure (1). Each strip covers all relevant locations in the GTRP,

from the most eastern village in the Dutch speaking area to the west, until a location in which no umlaut is attested (or where there is no location to the west at all). The GTRP contains 133 diminutives for all dialects, which were compared with their respective noun stems. Only diminutives which can be paired to stems with a back vowel, were selected (n=94).

We have chosen umlaut within diminutives as our empirical point of departure for a variety of reasons. First, the GTRP data clearly show an east-to-west structure at many different spots in the north-south dimension (we see the change from umlaut to non-umlaut both in the north and in the south). Secondly, although umlaut occurs also in other morphological paradigms, it seems particularly productive in the diminutive. Third, the questionnaire on which the GTRP is based paid close attention to asking relevant forms for many different types of nouns.

In the following section we will first provide a brief overview of our phonological analysis of umlaut, couched in OT. After this, we give an in-depth analysis of the grammatical changes needed to describe each of the three strips. The last section is devoted to a conclusion.

3 Umlaut and the diminutive

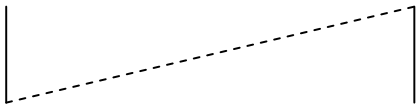
Diminutive forms in dialects of Dutch typically consist of a noun plus a diminutive suffix which has a variety of shapes. Umlaut on the stem is attested in the eastern Dutch speaking area except for the three northernmost provinces (Fig. 1). Umlaut in Germanic languages is usually defined as the alternation of a back stem vowel to a front vowel, due to a morphological suffixation process and the Dutch dialects form no exception to this:

(4) Umlaut on the diminutive in Winterswijk (Gelderland)

high	rutə	→	rytkən	‘window’
mid tense	sxoənə	→	sxøːkən	‘shoe’
mid lax	hɔnt	→	hʏndəkən	‘dog’
low tense	bak	→	bæksən	‘barge’
low lax	ɣɑns	→	ɣɛnskən	‘goose’
diphtong	mau	→	muʏkən	‘sleeve’

What exactly triggers the vowel alternation in (4)? Under autosegmental assumptions, we would expect the diminutive suffix to cause the alternation by some process of spreading. The coronal feature, which under some views (Clements and Hume 1995) represents frontness, spreads to the stem vowel, as illustrated in (5):[2]

(5) [labial] [cor]



b u k + x i n → bykin

Yet the suffix has no front vowel itself in many current dialects of Dutch. As the front vowel of the suffix historically was reduced to schwa in other varieties, the feature [coronal] (henceforth [cor]) became a floating feature under an autosegmental analysis (see for instance Nijen Twilhaar 1990), which becomes linked to the stem. This type of umlaut is still productive in diminutivization in several eastern Dutch dialects, witness forms such as:

[2] The feature [labial] presents roundedness, again following Clements and Hume (1995).

(6)	foto	føtəkə	'photograph'
	traktɔr	trɛktərkə	'tractor'
	blox	bløxskə	'weblog'[3]

Why can the feature [cor] not be attached to the suffix? A logical idea might be that [cor] cannot be attached to a schwa syllable, because schwa is not specified for Place. In many dialects umlaut does not occur in words with a schwa in the second syllable:

(7) Hellendoorn, umlaut on diminutives but not when the stem is disyllabic (data from GTRP)

a.	bɑk	~	bɛkskən	'tray'
b.	foto	~	fotøxin	'photograph'
c.	ɑpəl	~	ɑpəltjə	'apple'

The data in (7) show that [cor] is attached to the rightmost vowel of the stem. But umlaut is attested only when this rightmost vowel is a full vowel, not when it is a schwa. We propose that the attachment of the floating feature is bound to the final syllable of the stem and furthermore that [cor] can only be licensed in a strong position. Because the suffix is a weak position [cor] cannot be licensed there. In Limburg, generally the 'head of a foot' (*viz.* the stressed syllable) is the only possible target for [cor] (cf. Walker 2004):

(8) LICENSE[COR]/HDFT
[cor] is licensed by the head of a foot.

Notice that this constraint works both in dialects in which [cor] is still attached to the suffix (so where the diminutive has a full front vowel) and in dialects in which [cor] has become floating due to vowel reduction. When umlaut occurs, LICENSE[COR]/HDFT dominates a constraint which

[3] Data from Ben Hermans, dialect of Maasbracht (Zuid-Limburg).

prevents umlaut in other dialects. We propose that this is the constraint LOCALITY:

(9) LOCALITY
A feature F should not spread outside its domain.

LOCALITY generally prevents any kind of spreading; but if it occurs it wants the spreading to be in as short a domain as possible. In dialects where umlaut never occurs, the reversed ranking will generally be in place: LOCALITY dominates LICENSE[COR]/HDFT.

It should be noted that the licensing conditions mentioned in (8) differ per dialect region. It is not always necessary for the [cor] feature to reach the stressed syllable. In some dialects, any nucleus may be a licensing position for [cor], as we will see below.

4 Limburg and Brabant

The villages investigated in the provinces Zuid-Limburg and Noord-Brabant are Venlo (B1), Maasbree (B2), Sevenum (B3), Deurne (B4), Bakel (B5), Beek (B6), Sint-Oedenrode (B7), Boxtel (B8), Loon op Zand (B9) and Dongen (B10). Loon op Zand (B9) is the most western village where umlaut is attested.

The following vowel alternations due to umlaut occur in the region:

(10) u → y
o → ø
ɔ → ʏ
a[4] → ɛ (B1,2,5,7,8) , æ (B 5,7,9)
ɑ → ɛ (B1-B5), æ (B5-9)
ɑu → ɛ, ø, œy, ai

[4] In B3/4/6 the low vowel /a/ is not attested in the MAND.

In B4-5, centralizing diphthongs (diphthongs ending in a schwa-like vowel) are frequent, which are not umlauted in B4 but are variably umlauted in B5. Umlaut of centralizing diphthongs affects the first (low) vowel, and sometimes coincides with unrounding. But umlaut on centralizing diphthongs is also sometimes blocked (11):

(11) Umlaut of /ɑi/ in Bakel (B5)

a.	bɑik	bœykskə	'abdomen'
b.	trɑin	trɛntjə	'train'
c.	daim[5]	daimkə	'thumb'
d.	bɑis	kleine bɑisjə	'tube/ little tube'

Apart from this rich variation in alternating vowels, the data seem to be very chaotic at first sight. The overall impression is clear: the westernmost dialects display less umlaut than the more eastern dialects, but it is not so clear how the distribution of umlauting and non-umlauting contexts could be described.

A closer look at the data reveals that there are particular phonological contexts that prevent umlaut from applying. Some stem vowels seem to be blocking factors in some dialects, but final consonants can also block umlaut. Depending on the place of articulation of those final consonants, umlaut surfaces on the vowel or not. In B4-9, coronal consonants variably block umlaut. Besides, the absence of a final consonant also forms a context in which no umlaut can occur as we have seen above. Furthermore, schwa syllables in Limburg/Noord-Brabant can block umlaut. Moreover, these effects do not occur in an obvious pattern. In some dialects they occur, in others they do not and variation between the umlauted and the non-umlauted forms is frequently attested[6]. This variation also occurs

[5] Words ending in another back vowel, followed by –m, actually are umlauted.

[6] We observed a tendency of a gradual decrease of the number of umlauted forms, going from east to west in all regions, but there are not enough data for a statistical analysis.

within one dialect and sometimes even at the level of the individual speaker.

Table (12) shows the different umlaut occurrences in each dialect. The first column shows the phonological contexts that have a blocking effect on umlaut.

(12) Umlaut blocking contexts in Zuid-Limburg/Noord-Brabant

	Venlo	*Maas-bree*	*Sevenum*	*Deurne*	*Bakel*	*Beek*	*Sint-Oedenrode*	*Boxtel*	*Loon op Zand*	*Don-gen*
Code	*B1*	*B2*	*B3*	*B4*	*B5*	*B6*	*B7*	*B8*	*B9*	*B10*
Kloeke Code[6]	*L271*	*L267*	*L266*	*L244*	*L208*	*L204*	*L200*	*K172*	*K164*	*K163*
-V#	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
-Vn	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗
-ar(C)	✓	✓	✓	*var*	*var*	*var*	*var*	*var*	*var*	✗
-t	✓	✓	✓	*var*	*var*	*var*	*var*	*var*	*var*	✗
-s	✓	✓	✓	*var*	*var*	*var*	*var*	*var*	*var*	✗
-Vnt	✓	✓	✓	✓	*var*	*var*	*var*	*var*	*var*	✗
-CəC	-	✓	✓	*var*	*var*	*var*	*var*	*var*	*var*	✗
arm	✓	✓	✓	✓	✓	✗	✗	✗	✓	✗
-χ	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗
-p	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗
-k	✓	✓	✓	✓	✓	✓	✓	✓	*var*	✗
-m	✓	✓	✓	✓	✓	✓	✓	✓	*var*	✗
-Vn	✓	✓	✓	✓	✓	✓	✓	✓	*var*	✗

✓ = umlaut
✗ = no umlaut
var = variably umlaut
- = context does not occur

We will now investigate in detail each of the phonological contexts that prevent umlaut and the phonological processes and representations which cause the blocking effect.

4.1 Final open syllables

Venlo (B1) is the only location in our sample where umlaut occurs across the board or at least in all the contexts which we checked. Moving slightly to the west, the open syllable is the context in which umlaut dis-

appears first; whereas /u/ is regularly umlauted in B2-6, it is not in final position: in /ku-kə/ ‘cow’ and in /mu-kə/ ‘sleeve’. Interestingly, in B7-9 umlaut surfaces as plain /i/ in ku ~ kuikə. We assign the blocking effect of final open syllables to positional faithfulness. The relevant constraint is ANCHOR-R:

(13) ANCHOR-R
The rightmost segment remains unchanged

In B1, where umlaut also occurs on open syllables, the constraint ranking is LICENSE[COR]/HDFT » ANCHOR-R, but in the other dialects the ranking is reversed.

Although (13) may make a somewhat ad hoc impression at first sight, there is in fact evidence that the final segment is somehow more resistant to change than other segments in varieties of Dutch (van Oostendorp 2000). For instance, final vowels are less susceptible to vowel reduction (although unstressed /e/ typically reduces, it does not do so in words like *toffee* ‘toffee’ or *dominee* ‘pastor’) and vowel epenthesis also doesn’t occur at the outer edge of the word (*melk* ‘milk’ is pronounced as [mɛlək] rather than [mɛlkə], even though the latter strictly speaking has a better syllable structure with only one closed syllable rather than two).[7]

4.2 Schwa syllables

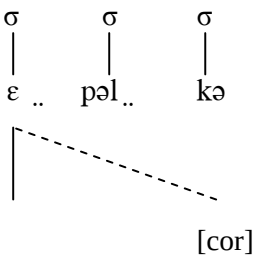
All disyllabic words of which the diminutive is provided in the GTRP database end in a schwa syllable e.g. ɑpəl ‘apple’[8]. The first, full, vowel is umlauted in B1-3, not umlauted at all in B10 and variation is attested in B4-9. When the vowel is umlauted, floating [cor] presumably crosses the

[7] Given that our analysis is cast in Optimality Theory, where constraints are violable, this does not imply that final segments are unchangeable in Dutch; they are not. In particular, final devoicing applies to them (hond /hɔnd/ ‘dog’ is pronounced [hɔnt]) in virtually all dialects of Dutch.

[8] Original disyllabic Germanic words have a schwa syllable as final syllable.

schwa syllable. If no umlaut occurs, we take this as an indication that crossing is not possible:

(14)



Since schwa is not specified for Place, [cor] will not surface. As we have mentioned above, constraint LOCALITY restricts the Licensing of [cor] to the rightmost syllable of the stem.

Thus, in B1-3 the constraint ranking is LICENSE[COR]/HDFT » LOCALITY, while in B10 the reversed ranking holds: LOCALITY » LICENSE[COR]/HDFT. In B4-9 variation is attested, which is accounted for by crucially non-ranked constraints. In Optimality Theory, crucially non-ranked constraints are postulated as a source of variation (e.g. Antilla 1997): both orders of the two constraints are possible, and therefore the results of both rankings will give a valid outcome. This means that, in B4-9, LICENSE[COR]/HDFT and LOCALITY are crucially non-ranked.

4.2.1 Frequency?

The variation in the disyllabic words could point towards language in change; is it the case that umlauted forms are gradually replaced by non umlaut. In that case we would expect high-frequency words to change first.There are six disyllabic words in our selection: ɑpəl ‘apple’, dɔxtər ‘daughter’, *kamər* ‘room’, ovən ‘oven’, torən ‘tower’ and voɣəl ‘bird’. Going from east to west, successively more of these words occur without umlaut. An example is provided in (15):

(15) Umlaut on disyllabic words

	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
ɑpəl		✓			✓	✓	✓		✗	✗
dɔxtə		✓			✓	✓	✗		✗	✗
kamə		✓			✓	✓	✓		✗	✗
ovən		✓			✗	✗	✗		✗	✗
torən		✓			✓	✗	✗		✗	✗
voɣəl		✓			✓	✓	✓		✓	✗

The order in which these words lose their umlaut, differs per region, as shown in (16):

(16) Zuid-Limburg/Noord-Brabant
{ovən}→{torən}→{dɔxtər}→{ɑpəl}→{kamər}→{voɣəl}

Overijssel
{ovən}→{dɔxtər}→{kamər}→{ɑpəl}→{torən}→{voɣəl}

Gelderland
{ɑpəl}→{dɔxtər}→{ovən}→{kamər, torən, voɣəl}

Although the order in which disyllabic words lose their umlaut differs per region, a clear tendency is present, which we cannot explain phonologically. From (15), it can be observed that *ovən* and *dɔxtər* are most likely to appear without umlaut, whereas umlaut on voɣəl is remained in all regions. What can be the reason of this? We compared the frequency of these words in the Corpus of the Institute for Dutch Lexicology (INL) and *Corpus Gesproken Nederlands* (Corpus Spoken Dutch) with the occurrence of umlaut on these words. Since the INL and CGN do not contain regional information, the order of umlaut loss (as in (16)), is averaged for all regions and compared with the frequency of the words. Loss of umlaut for Dutch dialects in general is then {ovən} → {dɔxtər} →

{ɑpəl} → {torən} → {kamər} → {voɣəl}. The frequency of these words is provided in table (17):

(17) Frequency of disyllabic words

order of umlaut loss			*CGN*	*INL*
1.	oven	'oven'	216	457
2.	dochter	'daughter'	717	3944
3.	appel	'apple'	-	304
4.	toren	'tower'	133	873
5.	kamer	'room'	412	13492
6.	vogel	'bird'	342	1489

One thing which can be readily observed from this table is that it is very hard to decide what counts as reliable information on word frequency. It is not clear what explains the discrepancy between these sources: the fact that the data are from different periods, the fact that one is about spoken Dutch, and the other about the written language, etc. makes this hard to establish. However, the sources do agree on a few issues, such as that *dochter* and *kamer* are both relatively frequent whereas *oven* and *appel* are relatively infrequent. As can be seen in (17), word frequency does not correlate very strongly to the gradual decline of umlaut on the diminutive. The order in which these words are unable to be umlauted in the diminutive thus remains unaccounted for.

Taking into account the frequency of the diminutive forms does not help much either. CGN does not provide us with sufficient data on these, but the INL corpus gives a result which is very similar to the forms which are not diminutivized:

(18) Frequency of diminutive forms

order of umlaut loss		*INL*
1.	oventje	10
2.	dochtertje	406
3.	appeltje	59
4.	torentje	60
5.	kamertje	411
6.	vogeltje	239

We can thus only conclude that there is no reason to suspect that word frequency plays a role in the susceptibility of forms to umlauting effects. There is thus no reason to consider this variation as a phenomenon of language change. Since there is no phonological explanation either, and since the patterns seem to be different for every region we studied, this looks like a random effect.

4.3 Coronal consonants

In B4-9, coronal stem final consonants /-s,-t,-n/ form an umlaut blocking context. Our interpretation is that floating [cor] is licensed by the Place node of these coronal consonants, rather than by the head of the foot in dialects B4-9. Thus floating [cor] does not need to surface on a vowel, and umlaut does not occur:

(19) LICENSE[COR]/PLACE
[cor] is licensed by a Place node in the root.

Clearly, (19) is a weaker version of the constraint LICENSE[COR]/HDFT that we have considered so far: whereas the latter wants licensing by a very specific segment with a place node (the stressed vowel), LICENSE[COR]/PLACE is satisfied if the feature can be linked to any seg-

ment in the root. These two constraints are in a so-called Paninian relationship (Prince and Smolensky 1993): a form which violates LICENSE[COR]/ PLACE will always also violate LICENSE[COR]/HDFT by definition, but not the other way around. This means that the relative order of the two constraints will be difficult to determine, except if there is a constraint between them. This means that in the absence of any evidence the two constraints will be unranked, but this doesn't mean variation since without intervening constraints LICENSE[COR]/PLACE » LICENSE[COR]/ HDFT and LICENSE[COR]/HDFT » LICENSE[COR]/PLACE give exactly the same result.

When the final consonant is a coronal nasal, umlaut displays a different pattern. If the vowel is short, umlaut surfaces, except in B9-10. This difference in behaviour is related to the allomorphy of the diminutive suffix. A stem with a long vowel and a coronal coda always selects the allomorph -tjə; however, a stem with a short vowel followed by a sonorant selects the allomorph -əkə.

Still it is not clear why long vowels followed by a coronal nasal cannot be umlauted in B4-9, whereas the other final coronal consonants display variation. Possibly this is caused by the fact that the nasal, being a sonorant, is 'closer' to the vowel in syllable structure than the other consonants. Various ways of formalizing this idea are available.

4.4 Other umlaut blocking contexts

Words ending in *-ar(C)* behave in a special way. In B1-3, they have to be umlauted ɑ → ɛ, but in B4-9 umlaut does not always occur. On the other hand, other vowels followed by *-r* are always umlauted in B4-9:

(20)		stem	diminutive (B6)	diminutive (B7)	gloss
	a.	kɑːr	kɑrskə	kærkə	'chariot'
	b.	kɑːrt	kɑːrtjə	kɑːrtjə	'card'
	c.	arəm	ærəmkə	arəmkə	'arm'
	d.	ɔːr	ʏrkə	ʏrkə	'ear'

The blocking effect of final *–rt* and *–r* is attested in Overijssel and Gelderland as well (see sections 5 and 6 below). We do not have an explanation for these facts. One problem is that there is no consensus in the literature on the representation of /r/, or on its relation to low vowels. So we leave this to future research.

It must be noted that the behaviour of *arm* 'arm' differs from both words ending in *–m* as well as from words with final *–art*. Why doesn't *arm* correspond to one of those categories? We observe that ɑrm is usually pronounced with a schwa: ɑrəm. Does ɑrəm follow the same pattern as disyllabic words? Table (12) shows that in B1-3 *arm* is always umlauted. In the regiolect B4-8 and in B9 it is umlauted in 3 out of 6 cases. Disyllabic words are also always umlauted in B1-3 and variably umlauted in B4-9. This suggests that the underlying representation of ɑrəm has a schwa, and that the word is treated as such in umlaut on the diminutive.

Non-coronal final consonants namely /χ,p,k,m/, do not display absorption of umlaut. Only in B9 variation is attested and in B10 umlaut is totally blocked. Therefore we suggest for B1-8 the constraint ranking LICENSE[COR]/PLACE » LOCALITY, but in B10 the reversed ranking holds.

In this section, we introduced a number of constraints which explain application and blocking of umlaut on diminutives. Constraint pairs were set up to indicate differences between dialects for a certain phonological context. Although the data look very chaotic at first, it turned out that a relatively small number of constraints and constraint rerankings can explain the major part of the dialectological variation. In the next section, these constraint pairs will be combined to full constraint rankings for each dialect, concerning umlaut on diminutives.

4.5 Phonological distance

Full constraint rankings are necessary to define phonological distances. The grammars consist of two blocks of two constraints each. First, we have two types of Licensing constraints: the strong constraint LICENSE[COR]/HDFT and its weaker counterpart LICENSE[COR]/PLACE. As

we have already indicated, these two constraints will be ranked together by a language learner unless there is a reason to tease them apart, in particular if there is another constraint separating the two. The reason for this is that the two constraints have the same goal.

(21)

Dialect	*constraint ranking*	*minimal ranking difference*
B1	LICENSE[COR]/HDFT, LICENSE[COR]/PLACE » ANCHOR-R, LOCALITY,	
B2-3	**ANCHOR-R»** LICENSE[COR]/HDFT, LICENSE[COR]/PLACE » LOCALITY	(2)
B4-8	ANCHOR-R » **LICENSE[COR]/PLACE » LOCALITY »** LICENSE[COR]/HDFT	(1)
B9	ANCHOR-R » **LICENSE[COR]/PLACE , LOCALITY »** LICENSE[COR]/HDFT	(0,5)
B10	ANCHOR-R, **LOCALITY »** LICENSE[COR]/HDFT, LICENSE[COR]/PLACE	(0,5)

When the constraint rerankings are calculated, the phonological distances from B1 to B10 can be represented as follows:

(22)

B1	-	B1	0	cum	0
B1	-	B2-3	2	cum	2
B1	-	B4-8	1	cum	3
B1	-	B9	0.5	cum	3.5
B1	-	B10	0.5	cum	4.0

In this way a close correlation between phonological distance and geographical distance can be expressed and this provides evidence for the topological relation which we hypothesized in section 2.

In the next sections we will briefly investigate the results for the Gelderland and Overijssel dialects.

5 Overijssel

The villages in the province of Overijssel which we investigated in the GTRP database are, from east to west: De Lutte (O1), Ootmarsum (O2), Tubbergen (O3), Vriezenveen (O4), Hellendoorn (O5), Heino (O6), Hasselt (O7), Genemuiden (O8), Vollenhove (O9) and Urk (O10). In Urk, although it is the westernmost village in our sample, umlaut is also attested albeit in a restricted set of circumstances. Urk was an island in the IJssel lake (the former Zuiderzee) for a long time, and after the land to its east was reclaimed, it is nowadays situated at the eastern border of the IJssel lake; there are no neighbouring villages to the west.

(23) Umlaut in Overijssel results in the following alternations:

u	→	y
o	→	ø
ɔ	→	ʏ
a/ɑ	→	ɛ

Low front vowels do not occur in this region, so *æ is undominated. In O1,2,3, the low vowel is umlauted to ɛ, but in the other dialects umlaut does not occur on the low vowel, indicating that a markedness constraint against low front vowels (*[cor, low]) is active in the regiolect and that in O1,2,3 LICENSE[COR]/PLACE » *[cor, low] whereas in the other dialects *[cor, low] » LICENSE[COR]/PLACE.

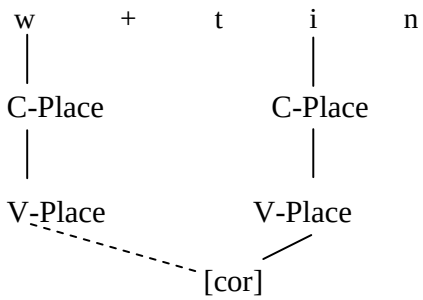
In several Overijssel dialects, the diminutive suffix is *–tin*, so that umlaut is not just a matter of linking a floating feature, but by licensing of a [cor] in the root that is already underlyingly linked to the suffix. What are the

licensers for [cor]? Umlaut occurs on the stressed vowel in O1-3. In O1-3 we find m/ɔu/w ~ m/ʏu/wke 'sleeve', however, in O4-10 the stem ends in schwa and then umlaut does not occur. Several reasons could be adduced for this blocking effect. First, licensing to the full vowel would involve violation of LOCALITY. In this case, it would be expected that this word would display the same patterning as disyllabic words, which turns out to be false, as can be seen in (27). A second reason for the underapplication of umlaut on m/ɔu/w might be that the target results in a marked diphthong. However, in Vriezenveen (O4): sxɑun 'shoe' is umlauted, but mɔu is not. Further, Nijen Twilhaar (1984:44) mentions that umlaut never occurs when the stem ends in a glide. An example is provided from the Hellendoorn dialect (O5), which does not have a diphthong:

(24) vrɔwə ~ vrɔwχi/* vrʏwχi 'wife'

We assume that it is the sonority of the glide which causes absorption of [cor]. Furthermore we assume that in the dialects where the final glide blocks umlaut, Licensing of [cor] occurs to the V-place node of the glide:

(25)



The relevant constraints are repeated from (8) and (19):

(26) a. LICENSE[COR]/ HDFT
[cor] is licensed by the head of a foot.
b. LICENCE[COR]/PLACE
[cor] is licensed by a Place node in the root.

Some contexts in which umlaut does not surface in the Overijssel dialects are the same as in Zuid-Limburg/Noord-Brabant: disyllabic words, final

open syllables and low vowels followed by /-r/. A noticeable difference is that final consonants seem to have less influence on umlaut than in Zuid-Limburg /Noord-Brabant. Instead, it is the height of the stem vowel which plays an important role, as can be seen in table (27):

(27) Umlaut blocking contexts in Overijssel

	De Lutte	*Ootmar-sum*	*Tub-bergen*	*Vriezen-veen*	*Hellen-doorn*	*Heino*	*Hasselt*	*Gene-muiden*	*Vollen-hove*	*Urk*
Code	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Kloeke Code	G209	G177	G174	G171	G168	F107	F085	F084	F066	F077
-VCə(C)	*var*	*var*	*var*	*var*	*var*	*var*	*var*	*var*	*var*	✘
-w/j (ə)	✓	✓	✓	✘	✘	✘	✘	✘	✘	✘
-V:r(C)	✓	✓	✓	-	-V□$_{[+low]}$r(t)	-	-	-V□$_{[+low]}$r(t)	✘	✘
-V$_{[+low]}$C$_{\#}$	✓	✓	✓	*var*	*var*	*var*	*var*	*var*	*var*	*Var*
-V$_{\#}$	✓	✓	-	✓	✓	✓	*var*	*var*	*var*	✘
-V$_{[+high]}$ (C)(əC)	✓	✓	✓	✓	✓	✓	✓	✓	✓	*Var*

✓ = umlaut
✘ = no umlaut
var = variably umlaut
- = context does not occur

As can be seen in (27), low vowels are always umlauted in O1-3 and variably umlauted in the other dialects. But non-low vowels are always umlauted, except for O10, where variation occurs.

In (27) three regiolects can be distinguished: O1-3, O4-6, O7-8. The constraint rankings for the Overijssel dialects are as follows:

(28)

Dialect	*constraint ranking*	*minimal ranking difference*
O1-3	*æ » LICENSE[COR]/ HDFT, LOCALITY » *[COR, LOW], LICENSE[COR]/PLACE, ANCHOR-R, IDENT-IO(V)	
O4-6	*æ » **LICENSE[COR]/PLACE, *[COR, LOW], LOCALITY** » LICENSE[COR]/HDFT » ANCHOR-R, IDENT-IO(V)	(5)
O7-8	*æ » LICENSE[COR]/PLACE, *[COR, LOW], LOCALITY, **ANCHOR-R** » LICENSE[COR]/HDFT, IDENT-IO(V)	(1)
O9	*æ » LICENSE[COR]/PLACE, *[COR, LOW], LOCALITY, ANCHOR-R » LICENSE[COR]/HDFT, IDENT-IO(V)	(1)
O10	*æ, **LOCALITY, ANCHOR-R** » LICENSE[COR]/PLACE, *[COR, LOW**], IDENT-IO(V)** » LICENSE[COR]/ HDFT	(1,5)

The difference between O1-3 and O4-6 is that*[COR, LOW] and LICENSE[COR]/PLACE are promoted two strata, since they dominate LICENSE[COR]/HDFT in O4-6 (and are not in variation with it). It concerns two constraints, so the phonological distance is 2*2 = 4. LOCALITY is promoted one stratum so the sum of the phonological distance is 4 + 1= 5.

(28) shows that, even though the phonological processes in Overijssel are somewhat different from Zuid-Limburg/Noord-Brabant, the minimal constraint rerankings again correspond to the geographical distances. Therefore, a topological relationship between geographical and phonological distance is also established for the Overijssel dialects.

6 Gelderland

In our last strip, the Gelderland dialects, the diminutive suffix is often -əkə, similar to Limburg/Noord-Brabant. Floating [cor] is attached to the rightmost full vowel of the stem, so the relevant constraint is LICENSE[COR]/PLACE. From the most eastern Winterswijk, the data of nine villages in north-western direction are investigated for umlauted diminutives; Winterswijk (G1), Lichtenvoorde (G2), Zieuwent (G3), Zelhem (G4), Hengelo (G5), Steenderen (G6), Brummen (G7), Loenen (G8) and Hoenderloo (G9). Successive decline of umlaut is attested in these villages until G9, in which umlaut (almost) no longer occurs. The vowel alternations are:

(29)
	u	→	y
	o	→	ø
	ɔ	→	ʏ
	ɑ	→	ɛ
	a	→	æ

For some reason, the front low vowel cannot occur when followed by -/r/ or -/x/ in the rhyme, which is shown in (30):

(30) Umlaut blocking contexts in Gelderland

	Winters-wijk	*Lichten-voorde*	*Zieuwent*	*Zelhem*	*Hengelo*	*Steen-deren*	*Brum -men*	*Loenen*	*Hoen-derlo*
Code	*G1*	*G2*	*G3*	*G4*	*G5*	*G6*	*G7*	*G8*	*G9*
Kloeke Code	M13	M7	G279a	G278	F206	F205	F178	F175	F173
-a:rt	✗	✗	✗	✗	✗	✗	✗	✗	✗
-a:x	✗	✗	✗	✗	✗	✗	✗	✗	✗
-Cə(C)	✓	*var*	✓	*var*	*var*	✗	*var*	✗	✗
- $V_{[+low,front]}CC$	✓	✓	*var*	✓	*var*	*var*	*var*	✗	✗
- $V_{[-low,\ front]}CC$	✓	✓	✓	✓	✓	✓	✓	✓	✓
-au -	✓	✓	✓	✗	✗	✗	✗	✗	✗

✓ = *umlaut*
✗ = *no umlaut*
var = *variability*
-x = *context in which no umlaut is attested*

In Standard Dutch, the velar fricative and the rhotic are (presumably) differently specified, but in the Gelderland dialects, a crucial similarity between these two sounds exists. The /r/ is pronounced as a glottal fricative when followed by a coronal consonant (Deunk 1977: II, 3-5), which explains the relationship between /rt/ and /x/ in non-transparency to umlaut. Both consonants display the same behaviour too, in a different phonological process, namely shortening. Long vowels are shortened in diminutives, but -/i:r/ and -/i:x/ cannot be shortened (Deunk 1977: II, 71). So in the region of Winterswijk (G1) /r/ and /x/ seem to form a natural class and apparently the sequence æ$C_{[+cont,+back]}$ is disallowed in all dialects investigated.

Another context which blocks umlaut (G4-9) is the diphthong /au/. We assume the front diphthong to be very marked, so in these dialects *[ʏu] dominates the licensing constraints.

A complexity arises with the unrounded low vowel, which has to be umlauted if followed by a single coda consonant, but cannot be umlauted if it is followed by a consonant cluster. So we find the following alternations in G8:

(31)	a.	kɑt	~	kɛt-jə	‘cat’
	b.	hɔnt	~	hʏnt-jə	‘dog’
	c.	kɑnt	~	kɑnt-jə	‘side’

It has been proposed that ‘monosyllabic’ words ending in consonant clusters actually behave as a disyllabic unit for a variety of Dutch dialects (Van Oostendorp 2000, Swets 2004). The final consonant behaves as if it were in the onset of an otherwise empty syllable. In this way *kant* is parsed as $(kan)_{\sigma}(t\emptyset)_{\sigma}$. In this representation, floating [cor] has to pass the final syllable, thereby minimally violating LOCALITY. This is not impossible, as shown in (31). Moreover, it is also not impossible to violate *[COR, LOW]. However, violation of both LOCALITY, as well as *[COR, LOW] seems to be avoided. This is expressed by the conjoined constraint (LOCALITY & *[COR, LOW]):

(32) (LOCALITY & *[COR, LOW])
Movement should be local *or* the low vowel in the input should be identical to its output correspondent

The constraint rankings for the Gelderland dialects are provided in (33):

(33)	*Dialect*	*constraint ranking*	*minimal ranking difference*
	G1	*æ$C_{[+cont,+back]}$ » LICENSE[COR] » IDENT-IO(V), Loc, *[ʏu], *[COR, LOW], (Loc & *$_{[COR, LOW]}$)	
	G2-3	*æ$C_{[+cont,+back]}$ » LICENSE[COR]/PLACE, **Loc, (Loc & *$_{[COR, LOW]}$)** » IDENT-IO (V), *[ʏu], *[COR, LOW]	(1)

...to be continued on next page

…	*Dialect*	*constraint ranking*	*minimal ranking difference*
	G4-7	*æC$_{[+cont,+back]}$, ***[YU]** » LICENSE[COR], Loc, (Loc & IDENT-IO(V)$_{[LOW]}$)» IDENT-IO(V), *[COR, LOW]	(2)
	G8	*æC$_{[+cont,+back]}$, *[YU], **(Loc & *[COR, LOW]), LOC, *[COR, LOW]**» LICENSE[COR]» IDENT-IO(V)	(1)
	G9	*æC$_{[+cont,+back]}$, *[YU], Loc, IDENT-IO, (Loc & *$_{[COR, LOW]}$), ***[COR, LOW] »** LICENSE[COR]	(2)

Summarizing, the results for Gelderland dialects are similar to the results of Overijssel and Zuid-Limburg/Noord-Brabant dialects, in the sense that geographical distance is, again, reflected in the phonological distance for umlaut on the diminutive.

7 Conclusion

This paper presents the data of an explorative study implementing the idea of doing dialectometry based on comparing OT grammars rather than products of grammars (such as words or sentences). We have shown how to define a notion of grammatical distance within OT, and how this notion can be successfully used to describe the geographical landscape of three areas in which umlaut is gradually lost.

Obviously, our data are idealized in the sense that we only consider three one-dimensional geographic objects here: the dialects are placed on a line moving from east to west. One of the next challenges would be to make

an OT map of a truly two-dimensional space, i.e. of some geographical area. Similarly, in this case we have considered strips where we knew that a particular phenomenon was going on: the disappearance of umlaut. There obviously may be even stretches of dialects where things are more chaotic: a phenomenon might be present in the east and in the west, but not in the centre, for instance. However, the methodology to be followed would be the same in all of these more complicated cases, and the prediction would be that our definition of grammatical distance would still hold. The present study, working on a relatively simple sample merely shows how it works.

Even though our data are relatively simple, however, we should point out that the fact that our method works at all is significant. There is no *a priori* reason why OT grammars should change gradually when we are travelling through a geographic area. The loss of umlauting could also have been a purely lexical issue, where fewer and fewer words undergo umlaut, but not because the phonological restrictions on umlaut become stronger and stronger, but because e.g. the more frequent words are the first (or the last) to lose it. We have seen that there is no evidence for these frequency effects. Phonological conditioning, on the other hand, seems rather strong.

The fact that it is possible to draw these grammatical landscapes at all, then, might be taken as an argument in favour of writing OT grammars such as these in the first place. Apparently, they convey some hidden regularities behind the apparent chaos of the initial pattern.

A more complete model of the phonological landscape would obviously also take other phonological phenomena into account, and the mutual interaction between them. At some point, when we start considering larger fragments of larger numbers of grammars, arranged in more complex geographic spaces, we will no doubt need to invoke the more sophisticated statistical measures which were developed by the dialectometrists we mentioned in the introduction. We hope to have shown in this article; however, that insight into the structure of grammar is essential for the understanding of dialectological space as well.

References

ANTTILA, A.
(1997). 'Deriving variation from grammar'. In: Frans Hinskens, Roeland van Hout & Leo Wetzels (eds.), *Variation, change and phonological theory*. 35-68. John Benjamins, Amsterdam.

BOTMA, B. & TORRE E.J. VAN DER.
(2000). The prosodic interpretation of sonorants in Dutch. In: de Hoop, Helen & Ton van der Wouden (eds.), *Linguistics in the Netherlands* 17–29. Amsterdam and Philadelphia: John Benjamins.

CLEMENTS, G. N. & HUME, E.V.
(1995). The Internal Organization of Segments. In: *J. Goldsmith (ed.) The Handbook of Phonological Theory. Blackwell Handbooks in Linguistics*. Blackwell Publishers, 245-306.

DEUNK, G.H.
(1977). *Het dialect van Winterswijk. Deel II. Grammatica van het Winterswijks.* Sasland, Groningen.

GUNNINK, J.
(1908). *Het dialect van Kampen en omstreken.* Dissertatie. Kampen, Kok.

HAAR, D. VAN DER.
(1969). *Gaellemuun en 'et Gaellemunegers*. Nijmegen, Thoben offset.

HEERINGA, WILBERT.
(2004). *Measuring Dialect Pronunciation Differences Using Levenshtein Distances.* PhD Dissertation, University of Groningen.

HERMANS, B.
(1994). Palatalisatie in een Limburgs dialect: dissimilatie als gevolg van OCP. In: G.Booij & J. van Marle (eds.), *Dialectfonologie*. Amsterdam (Cahiers van het P.J. Meertens-Instituut), 60-82.

HINSKENS, F. & OOSTENDORP, M. VAN.
(2004). Coronale nasaal-plosief clusters in de Nederlandse dialecten. Een verkenning. In: Johan De Caluwe, Georges De Schutter, Magda Devos, Jacques Van Keymeulen (eds.), *Taeldeman, man van de taal, schatbewaarder van de taal*. Gent, Academia Press.

IVERSON, G.K. & SALMONS, C.
(1996). The primacy of primary umlaut. *PBB/BGDSL.* Beiträge zur Geschichte der deutschen Sprache und Literatur.
www.uwm.edu/~iverson/primacy.pdf

KAGER, R.
(1999). *Optimality Theory.* Cambridge University Press.

LEVENSHTEIN, V.
(1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* 19:707-710.

NERBONNE, JOHN, AND WILLIAM KRETZSCHMAR JR.
(2006). Progress in dialectometry: Toward explanation. *Literary and linguistic computing* 21(4):387-397.

NIJEN TWILHAAR, J.
(1990). *Generatieve fonologie en de studie van Oostnederlandse dialecten.* Dissertatie. Amsterdam, Meertens Instituut.

NIJEN TWILHAAR, J.
(1984). *Morfonologische aspecten van Hellendoornse diminutiva.* Doctoraal Scriptie/MA thesis. Universiteit van Utrecht.

OOSTENDORP. M. VAN.
(2000). *Phonological projection: a theory of feature content and prosodic structure.* Mouton de Gruyter, New York.

PRINCE, ALAN AND PAUL SMOLENSKY .
(1993). *Optimality Theory: constraint evaluation in generative grammar.* MIT Press, Massachusetts. Published at Blackwell, London, 2004.

SPA, J.J.
(2000). *De dialecten van Groot-IJsselmuiden. Klank- en vormleer.* Kampen, IJsselacademie.

SPA, J.J.
(1996). *Het dialekt van de Stad Vollenhove. Klank- en vormleer.* Kampen, IJsselacademie.

SPRUIT, MARCO RENÉ.
(2008). *Quantitative perspectives on syntactic variation in Dutch dialects.* PhD Dissertation, University of Amsterdam.

SWETS, F.
(2004). *The Phonological Word in Tilburg Dutch.* Utrecht: LOT.

WALKER, R.
(2004). Vowel feature licensing at a distance: Evidence from Northern Spanish language varieties. In: Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez, and Benjamin Schmeiser (eds.), *Proceedings of WCCFL 23.*787-800.