Variation in Generative Grammar

Marc.van.Oostendorp@Meertens.KNAW.nl
Royal Netherlands Academy of Arts and Sciences, Amsterdam

April 21, 2005

Contents

Introduction

Thinking about the relation between generative grammar and language variation implies thinking about the value of empirical evidence. Most phonological theories deal, either implicitly or explicitly, with the Chomskyan notion of 'I-language' – phonology is seen as part of the individual's knowledge of language. Yet most facts about language variation are facts about E-language, about the way language functions in the world. How can we bring this evidence to bear on our theories? What data are relevant and what can we leave aside? This will be the main topic of study in this course, in which we will look at the three most important types of variation: social variation, geographic variation and diachronic variation (language change).

1 Variation and Phonological Theory

1.1 Why study variation if you are interested in grammar?

From its inception, Optimality Theory (OT, Prince & Smolensky, 1993) can be seen as a theory of language variation: it describes how languages can differ within the limits imposed on them by Universal Grammar. As a matter of fact, classical OT defends a very strong hypothesis about this:

(1) Systematic differences between two languages can only be the result of different rankings of the same constraints.
Two languages L_1 and L_2 are different *iff* there are constraints A, B such that A$\gg$B in L_1, and B$\gg$A in L_2.

The set of rankings {A$\gg$B, B$\gg$A} is called the *factorial typology* (FT) of the constraints A, B; it is assumed that every element of the FT constitutes a possible human language.

In accordance with a general tendency in generative phonology, the empirical scope of most work within OT is on macrotypology: the description of languages which are genetically and typologically distant. The early success may be due at least in part to the fact that many macrotypological facts could be described elegantly with the device of FT. For instance, we know that there are languages which do and languages which do not allow syllables to end in a coda consonant. Of the languages which do not allow codas, there are two different types: some avoid them by deleting final consonants (CVC $\rightarrow$ CV), others avoid them by inserting vowels (CVC $\rightarrow$ CVCV). The factorial typology of this involves three constraints, NoCoda, NoInsertion and NoDeletion. This gives the following factorial typology:

(2) a. Coda's are allowed: NoDeletion ≫NoInsertion ≫NoCoda
= NoInsertion ≫NoDeletion ≫NoCoda

b. Offending consonants are deleted: NoCoda ≫NoInsertion ≫NoDeletion
= NoInsertion ≫NoCoda ≫NoDeletion

c. Epenthetic vowels are inserted: NoCoda ≫NoInsertion ≫NoDeletion
= NoInsertion ≫NoCoda ≫NoDeletion

The question now arises to what extent also *microvariation*, i.e. variation between language systems which are genetically and typologically close can be represented within OT. This is the topic of this course.

We will distinguish here between the following types of variation:

- *Intraspeaker variation*. Variation 'within' a speaker. Every human being can use his language in more than one way. This type of variation can be further subdivided into two subtypes:
  - *Pragmatic variation*: this is variation which is somehow 'meaningful'. If the speaker uses variant A, this has a different meaning than if he uses B. Examples: formal vs. informal speech; fast vs. slow speech; etc.
  - *Free variation*: this is all the variation which we find and to which we cannot assign any meaning. Two forms which count as equally optimal internally, andwhich are not dependent on external factors
- *Interspeaker variation:* Two speakers speak differently, even though they can still be said to speak the 'same language' (in some informal sense).
  - *Geographical variation*: variation between 'dialects' of the o same 'language'
  - *Sociolinguistic variation*: variation between men vs. women; young vs. old people; people of different social background; etc.
  - *Temporal variation*: Language change; people speak differently than their ancestors did x years ago.

Obviously, it is not always very easy to determine where a given phenomenon belongs in this classification. From the point of view of the study of Chomskyan I-language (Chomsky, 1986), the most interesting of these obviously are types of intraspeaker variation. If we can detect how people are able to produce different forms, we can learn from this about the organisation of the grammar. The interest of interspeaker variation at first sight lies more in what it can tell us about Chomskyan E-language. But there are still several reasons why this is worth studying if you are interested in grammar:

i. Every language variety (dialect, sociolect) is obviously interesting in itself, and can potentially provide crucial evidence for or against some linguistic theory.

ii. Comparison of closely related systems can shed light on the right analysis for a given system.
iii. The issue of variation itself is interesting: how does grammatical variation correspond to external factors? E.g. what characterizes the differences between formal and informal grammars? Or how does the notion of grammatical distance correspond to geographical distance?

I will exemplify each of these points with examples from Dutch dialectology below (we will return to these and similar examples in chapter 3).

i. In Aalst Dutch nasals are assimilated to the place of an adjacent consonant (Taeldeman, 1980):

(3) /ɣryn/ 'green' [ɣrym bumkə] 'green tree-DIM'

In rule-based terms this can be described as follows:

(4) *n-Assimilation* n → [αlab βcor γback] / ___ [αlab βcor γback]

Aalst Dutch also has a rule of schwa apocope, which interacts with assimilation in an interesting way:

(5) schwa-apocope ə→ ∅ / { morphophonologically defined context}

(6)

	ɣryn+ bum + ən 'green trees'
n-Assimilation	d.n.a.
schwa-apocope	[ɣryn bumən]

This process can be described in rule-based terms, albeit using an order which is usuallyconsidered 'unnatural'. This process poses very difficult problems, on the other hand for ananalysis (such as OT or GP) that is phrased in terms of outputs only. The number of intriguing phenomena that could be found in human language is immensely large.

ii. The literature on Germanic contains many ideas about the proper representation of thevelar nasal /ŋ/. One idea is that this velar nasal is a cluster (e.g. /ng/) underlyingly.

This seems supported at first sight by Dutch and German dialects which say [dɪŋk - dɪŋə] 'thing-things'. In other dialects we find this [k] only before an

[s]. We could wonder whether this velar nasal is intrusive or underlying. If if we draw maps of the relevant areas, we discover that there are differences between monomorphemic *hengst* 'stallion' and polymorphemic *bangst* 'most afraid':

- dialects with [hɛŋst] and [bɑŋst]
- dialects with [hɛŋkst] and [bɑŋst]
- dialects with [hɛŋkst] and [bɑŋkst]
- very rare, and possibly absent, are dialects with [hɛŋst] and [bɑŋkst]

This corresponds with a finding of English dialectology:

- dialects with *finger* [fɪŋər] and *singer* [sɪŋər]
- dialects with [fɪŋgər] and [sɪŋər]
- dialects with [fɪŋgər] and [sɪŋgər]
- very rare, and possibly absent, are dialects with [fɪŋgər] and [sɪŋgər]

This may give us a clue as to how the interaction between morphological and phonological structure should be modelled. Apparently 'intrusive' segments are less likely to occur at a morpheme boundary than in morpheme-internal positions.

iii. The next class will be devoted to a large extent to modelling the relation betweenl and internal factors of language variation in the study of style levels (formal vs.informal speech). Here we concentrate on the study of geographical variation within OT.

Different ranking of universal constraints (the way all variation is modelled within thistheory) gives us a simple way to define the notion of 'linguistic' distance between two language systems:

(7) The linguistic distance between two dialects is the minimal number of minimal rerankings needed to get from one grammar to the other.

A minimal 'rerankingÕ in this definition is the reranking of two adjacent constraints; aranking a≫b≫c can thus be minimally reranked in two different ways (b≫a≫c or a≫c≫b); these have a distance of 1. Other rankings of these same constraints can only be attained by more than one minimal reranking. Thus the ranking b≫c≫a is at a linguistic distance of 2 from theoriginal ranking, and c≫b≫a at a distance of 3. We could now postulate that geographical variation may be modelled as in (8):

(8) The linguistic distance between two dialects of a language system equals the geographical distance between those dialects in a topological way

Topology is the branch of mathematics that is concerned with the geometrical properties ofobjects without being concerned with absolute distances.

It would be absurd to try to predict the exact number of kilometers (for instance) between two villages on the basis of the grammatical differences between the dialects spoken in those villages, but we do predict thatif dialect x and y are at a linguistic distance 2 to one another, and dialect z is at distance 1 to both, that z should be between x and y. In a few cases, problems arise, however. In these cases, the insights of traditional dialectology concerning the relation between cultural and geological boundaries and linguistic isoglosses help to refine the hypothesis in (8). (On the other hand, this idea about a direct relation between grammatical and geographic distance is certainly not uncontroversial. We will return to this later in the course.)

1.2 Models of intraspeaker variation in OT

Mutliple vs. variable grammars

What is the proper way of describing language variation from a grammatical perspective? In the literature we find roughly two approaches:

- *Multiple grammars*: Every individual grammar has exactly one output for every individual input; but an individual may command more than one grammar.
- *Variable grammars*: An individual has one grammar. Variation is represented within that grammar, for instance in the form of 'variable rules'

Within classical OT, it is hard to draw an absolute distinction between the two approaches. Let us take a (heavily simplified) example from Faetar (a Franco-Provençal dialect spoken by a small community in southern Italy (Nagy & Reynolds, 1997).***. In this dialect a word such as *brokele* '' can be pronounced as [brokələ], [brokə],[brok] or [brokl]. Of these, the first is the most frequent pronunciation by far.

***

The other forms display some amount of deletion. Acccording to Nagy & Reynolds (1997), the reason for this is that word-final stress is preferred. Responsible for this is a constraint ALIGNPRWD, stating that the edge of the main stressed foot should be at the (righthand) edge of the word. This constraint is counteracted by constraints such as NOCODA (because deletion of unstressed vowels may bring into being extra closed syllables) and PARSE (against deletion of vowels and consonants). Ranking ALIGNPRWD in different ways *vis à vis* the other two, gives different results:

(9) a. ALIGNPRWD≫NOCODA≫PARSE: Deletion of last two syllables: [brok]

b. NOCODA≫ALIGNPRWD≫PARSE: Gives [brokə] or [brokl̩]: as much as possible is deleted, but without creating a closed syllable.[1]

c. NOCODA≫PARSE≫ALIGNPRWD: No deletion at all: [brokələ]

One way to look at (10), is as a list of three different grammars, each of which gives a deterministic output. Another way of looking at it is as the extension of one more complex grammar in which ALIGNPRWD is 'floating', i.e., it can be ranked variably in any position from above NOCODA to just below PARSE (or actually the constraints which will make the choice between [brokə] and [brokl̩].

It is hard to draw a distinction between these two interpretations. We could say that the first one is less restrictive, it does not capture the intuition that the three grammars are still very similar – we have to assume that all of the high-ranking and low-ranking constraints are in the same order in these three grammars. If an individual has complete freedom in manipulating individual grammars, why don't we ever find a speaker who randomly chooses between pronouncing a word with a Chinese phonology and pronouncing the same word with an English phonology? We could answer this by assuming that different grammars should always be stored in a parsimonious way: redundancies between different grammars will be removed. But then we actually have something which is suspiciously similar to the intensional 'floating constraint' interpretation.

On the other hand, the 'one grammar' solution has as its problem that it can describe variation that is truly 'free': the constraint is floating and it can randomly land on one of the positions that is available to it. But the number of examples of real free variation is very limited; very often it turns out on closer inspection that there is some social or grammatical factor which influences the choice between form α and form β; e.g. the speaker uses α when she is talking to a woman, and β when talking to a man. It is very hard to see however how this choice could be described without referring to the constraint ranking which has α as its output as opposed to the ranking that has β as its output. But in that case, we are back at the extensional interpretation. It thus seems that we need to have *both* interpretations – that they are indeed interpretations of the same formalism, and sometimes it is more useful to think about them in one way, and at other times it is more useful to think about them in another way.

Before we continue, we have to note another point about the Faetar example that was made by Nagy & Reynolds (1997): the distribution of surface variants for /brokələ/ is not random. Rather, the (faithful) form [brokələ] occurs far more often than the other three forms. The reason for this is, according to Nagy & Reynolds (1997), that there are a few more constraints at

[1] Strictly speaking, the optimal candidate here would be [bro], which never occurs. This should be due to independent factors, which we will ignore here.

play at the bottom of the hierarchy. In the first place there are two constraints which decide between [brokə] and [brok̩]: HNUC is a constraint in favour of schwa heads (it says that the nucleus should be a vowel), and *SCHWA favours the syllabic nucleus. (These constraints are also floating with respect to one another, but that need not concern us here.) There also is a third constraint, ONSET, that needs to be placed somewhere in this part of the constraint hierarchy.

If we now assume that ALIGN-PRWD floats over these low-ranking constraints as well, we get the following result:

(10) a. ALIGNPRWD≫NOCODA≫PARSE≫ONSET≫HNUC, *SCHWA: Deletion of last two syllables
b. NOCODA≫ALIGNPRWD≫PARSE≫ONSET≫HNUC, *SCHWA: As much as possible is deleted, but without creating a closed syllable.
c. NOCODA≫PARSE≫ALIGNPRWD≫ONSET≫HNUC, *SCHWA: No deletion at all: [brokələ]
d. NOCODA≫PARSE≫ONSET≫ALIGNPRWD≫HNUC, *SCHWA: No deletion at all: [brokələ]
e. NOCODA≫PARSE≫ONSET≫HNUC≫ALIGNPRWD*SCHWA: No deletion at all: [brokələ]
f. NOCODA≫PARSE≫ONSET≫HNUC≫*SCHWA≫ALIGNPRWD: No deletion at all: [brokələ]

We thus have four rankings which give the same output. Assuming that every ranking is equally likely (the distribution is really random) this should mean that we find this particular form four times as often as the other forms. This is indeed what Nagy & Reynolds (1997) try to establish, and indeed find (to some extent):

Their actual calculation is a little bit more complicated, but these details need not concern us here. (Cf. Guy, 1997, for a critical evaluation of Nagy & Reynolds (1997)'s approach).

Stratified ranking

The most influential proponent of this 'classical' model of intraspeaker variation within OT is Arto Antilla (Antilla, 1997a,b, 2002). His approach is in certain ways very similar to the one of Nagy & Reynolds (1997), sketched above. A key example comes from Finnish. There are (at least) two endings

for the genitive plural in this language. One, /ien/ or /jen/, is called *weak*, and the other, /ien/, is called *strong* (example from Antilla, 2002, p. 226).

(11)

	Stem	Variants	Gloss
a.	/puu/	pu-iden	'tree'
	/maa/	ma-iden	'land'
b.	/lasi/	las-ien	'glass'
	/margariini/	margariin-ien	'margarine'
	/sosialisti/	sosialist-ien	'socialist'
c.	/naapure/	naapure-iden naapur-ien	'neighbour'
	/ministeri/	ministere-iden minister-ien	'minister'
	/aleksanteri/	aleksantere-iden aleksanter-ien	'Alexander'

The examples in (11a) shows that all monosyllabic stems choose the strong ending invariable; (11b) shows that all disyllabic forms as well as all longer words in which the prefinal syllable is heavy and the final vowel high, choose the weak form invariable; and (11c) shows that all other words display variable behaviour. This variation seems to be 'free' in the sense that no known linguistic of extralinguistic factor can predict which variant will be used. On the other hand, speakers have an intuition that one ending may be more appropriate for a given word, even though the other ending is still allowed. Furthermore, a corpus study shows that the endings are not equally frequent. According to Antilla, this frequence conforms to the acceptability judgements (more frequent forms are considered more acceptable).

(12)

	Stem type	Strong	Weak	Gloss
a.	/ka.me.ra/	ka.me.roi.den (99.4%)	?ka.me.ro.jen (0.6%)	'camera'
b.	/sai.raa.la/	sai.raa.loi.den (50.5%)	sai.raa.lo.jen (49.5%)	'hospital'
c.	/naa.pu.ri/	naa.pu.rei.den (37.2%)	naa.pu.ri.en (62.8%)	'neighbour'
d.	/po.lii.si/	?po.lii.sei.den (1.4%)	po.lii.sen (98.6%)	'police'

Antilla (2002) points out that two generalisations emerge: (i) the process is sensitive to vowel height (low vowels prefer the strong variant, high vowels the weak variant), and (ii) stems with a light penultimate syllable prefer the strong variant, stems ending in a heavy penultimate syllable prefer the weak variant. These effects are furthermore cumulative: if a word has a light penultimate and a low final vowel, the strong ending is preferred "almost categorically", whereas forms with heavy penultimates and high final vowels almost always choose the weak ending. The real variability is in those words in which the final and antepenultimate syllables impose conflicting demands.

In order to understand what is going on, we first have to note that primary stress is on the first syllable of every Finnish word. Secondly, we have to take into account two naturally looking constraints:

- *Ĺ: No stressed light syllables
- *H: No stressless heavy syllables

These constraints are not ranked with respect to one another. As in the floating constraint model, this means that we randomly choose *Ĺ≫*H or *H≫*Ĺ. For monosyllabic and disyllabic words, this is all there is to the analysis. No matter in what way we order the constraints, we always get the right results (given appropriate background assumptions about the avoidance of stress clash, etc.) (Antilla, 2002, p. 228)

(13) a. /maa/ 'land'

TABLEAU I	*Ĺ	*H
☞mái.den		*
má.jen	*!	*
TABLEAU II	*H	*Ĺ
☞mái.den	*	
má.jen	*	*!

b. /lasi/ 'glass'

TABLEAU I	*Ĺ	*H
lá.sei.den	*	**!
☞lá.si.en	*	*
TABLEAU II	*H	*Ĺ
lá.sei.den	**!	*
☞lá.si.en	*	*!

However, in longer words, these constraints do not decide:

(14) /naapuri/ 'neighbour'

TABLEAU I	*Ĺ	*H
☞náa.pu.rèi.den		*
☞náa.pu.ri.en		*
TABLEAU II	*H	*Ĺ
☞náa.pu.rèi.den	*	
☞náa.pu.ri.en	*	

This means that in this case, other constraints have to decide. Among the relevant constraints are the following:

(15) a. *L.L: No adjacent light syllables

b. *H/I: Ho heavy syllables with a high vowel

c. *Í: No stressed high vowel

Again these constraints are unranked with respect to each other (but they all are ranked below the two constraints we have just seen at work), and all orders are possible. We now get the following very interesting result:

(16) /naapuri/ 'neighbour'

TABLEAU I	*H/I	*Í	*L.L
náa.pu.rèi.den	*!	*	
☞náa.pu.ri.en			*
TABLEAU II	*H/I	*L.L	*Í
náa.pu.rèi.den	*!		
☞náa.pu.ri.en			*
TABLEAU III	*Í	*H/I	*L.L
náa.pu.rèi.den	*!	*	
☞náa.pu.ri.en			*
TABLEAU IV	*Í	*L.L	*H/I
náa.pu.rèi.den	*	*	
☞náa.pu.ri.en			*
TABLEAU V	*L.L	*H/I	*Í
☞náa.pu.rèi.den		*	*
náa.pu.ri.en	*!		
TABLEAU VI	*L.L	*Í	*H/I
☞náa.pu.rèi.den		*	*
náa.pu.ri.en	*!		

Two out of six grammars (33%) thus choose the strong form, and the other four (66%) choose the weak form. Notice that this corresponds quite well to the statistics for this word in (12). (On the other hand, it would have been quite easy to model a situation in which the distribution would have been e.g. 50%-50%; in that case we could have left either *H/I or *Í out of the equation, as the reader may verify (technically, we would have awarded it a lower, absolutely dominated ranking). This is true more in general: we can get any distribution as long as we have the right number of constraints pro and contra a given form.

Antilla's model is called *stratified grammar* by Boersma (2001), because the grammar can be seen as a number of strata of constraints (at least under the intensional interpretation): the strata are ordered with respect to one another, but the constraints within an individual stratum are not. Antilla (2002)

adapts this terminology. Above, we have considered two strata of Finnish phonology:

{ *Ĺ, *H} Stratum I
≫
{ *L.L, *Í, *H/I } Stratum II

Continuous ranking scales

A more radical departure of standard assumptions can be found in the work of Paul Boersma (Boersma, 1998; Boersma & Hayes, 2001, and other works). In this model, ranking is no longer a matter of discrete ordering. In stead of this there is a continuous scale. Constraints occupy certain zones on this scale, in the following way:

Constraint C1 occupies a higher order in the hierarchy than constraint C2, but there is some overlap. Assuming that the constraints can be positioned randomly within their domain[2], C1 will usually be ranked above C2, but in some rare cases, C2 will inversely dominate C1.

Common

Rare

Boersma & Hayes (2001) applied this idea to Antilla's Finnish data as well. The precise ranking value for each constraint was determined by a ('learning') algorithm on the basis of a relatively large corpus of data. We display a few of these 'ranking values' by way of illustration:

(17)

Constraint	Ranking value
*H	288.000
*Í	207.892
*L.L	206.428

These constraints are the only ones from those cited above that seem to be necessary. In all, Boersma & Hayes (2001) cite 10 constraints, partly also to deal with cases which we have not discussed here; they demonstrate that

[2] As a matter of fact, Boersma's assumptions are more sophisticated; he assumes that the probablity that a position is chosen close to the center is always higher than the probablity that a position is chosen far away from the center. The 'domains' of every constraint can then be assumed to be infinitely large.

their analysis can handle the Finnish data at approximately the same level of adequacy as Antilla. The authors claim that "[m]ost of the constraints we omitted were constraints that have little support from phonological typology".

But there is another reason why this analysis can do with a smaller number of constraints: Antilla needed the extra constraints to derive the statistic effects. Rougly speaking, if the corpus distribution is such that in 25% of the cases we find α, and we find β in the remaining 75% of the cases, this means that we need at least four constraints in stratal phonology, 1 favouring α and/or disfavouring β, and 3 to the reverse effect. If it turns out that there is evidence for 2 constraints in favour of α in the same stratum, we even need 6 constraints for β. In Boersma & Hayes (2001)'s approach, the same effect can be modeled by two constraints: one in favour of α and another in favour of β, placed at such a distance that α will win in 75% of the cases.

From the point of view of theoretical insight, a stratal approach might be seen as preferable over a continuous ranking approach. The latter does not really offer an explanation of why the numbers should be 25%-75%. *Any* number is available, and could be extracted by the learning algorithm from the corpus. In other words, there is no prediction as to what constitutes a possible corpus as to the amount of variation. Within the stratal approach, on the other hand, we are bound by the universal set of constraints. If we know that there are three relevant constraints for a particular choice, the only possible number pairs are 100%-0% (one constraint ranked higher than the other two), 50%-50% (two constraints in a stratum, and one below it) and 67%-33% (all three constraints in a stratum). It remains to be seen whether this restrictiveness holds against the richness of the data, but it seems to be a good guideline. However, this comes with the proviso that at the moment there is no consensus as to what the universal constraint set is, or which constraints have sufficient "support from phonological typology". Some authors have even suggested that constraints could be made "on the fly" during the language acquisition process, so that there is no universal set of constraints at all.

Yet another issue is the modelling of variation that is known *not* to be free, arguably the most common type. What if we know that some extragrammatical fact influences the choice between two alternatives? This seems easier to model within a continuous ranking approach, where we could say that in certain circumstances the probability that some constraint is ranked somewhat higher in its domain becomes larger. In a continuous ranking, it is not so clear what we should do at first sight. It remains to be seen what things could be said about this, and we will return to this issue in the next chapter.

Finally, we need to mention that the continuous ranking approach obviously invites a variable grammar interpretation. But even in this case, there is an equivalent approach using multiple grammars. Given a set of constraints $\mathscr{C}$ we could assume that a linguistic system consists of every possible rank-

ing of this constraint set, with a probability assessment attached to it. This gives a very large but still finite number of grammars ($|\mathscr{C}|!$).

A rank-ordering model of Eval

Coetzee (2004) proposes a model which is the only one – at least the only one I know – which can only be interpreted as a variable grammar system. He presents this as an interpretation of the Eval function in OT. Usually, it is assumed that Eval distinguishes only between one candidate (the winner) and all other candidates (the losers), but according to Coetzee we should see the Eval function as generating an ordering over the total set of candidates.

Take for instance the following example from Latvian. Final unstressed vowels in this language are optionally deleted, as the following example illustrates:

(18) a. /spligti/ → [spligt] - [spligti] 'dazzling' (M.PL.)
b. /pele/ → [pel] - [pele] 'mouse'
c. /spligtas/ → [spligts] - [spligtas] 'dazzling' (F.PL.)

Deletion is statistically preferred, but it is not the only option. Let us suppose, following Coetzee (2004), that the following two constraints are involved:

(19) a. *V̆]$_\sigma$]$_\omega$: Do not allow a vowel in an unstressed prosodic word-final syllable.
b. MAX: Do not delete (vowels).

Now the ordering *V̆]$_\sigma$]$_\omega$≫MAX establishes an order over at least two of the candidates:

(20)

/pele/	*V̆]$_\sigma$]$_\omega$	MAX
[pel]		*
[pele]	*	

Under the classical interpretation of OT, the result of this evaluation would be that [pel] is the winner. But in a rank-ordering interpretation, the result of the evaluation is an ordering on the candidates [pel]≻[pele]. The speaker can now choose to pronounce one of these forms; but the higher a form is on the hierarchy, the more likely the candidate is to pronounce this particular form.

This analysis clearly depends on the assumption that we have one grammar, which has more than one output – it could not be seen as a multiple grammars model in any way. Several issues now arise. For instance:

- Why is there variation for some phenomena but not for others?

- How can we describe the influence of extralinguistic factors on the actual choice one makes in variation?

If the result of Eval is a total ordering of all candidates (i.e. all logically possible forms), we would expect every possible candidate to sometimes surface. One possibility to answer this, would be along the lines of Boersma & Hayes (2001): we could say that these forms are indeed possible, but they are simply extremely unlikely, so that we will never find them in actual fact. For some reason, this is not the option that Coetzee (2004) chooses. Rather, he defines a 'critical cut-off' point: those forms which violate constraints which are too higly ranked are still considered to be 'losers' and will not be generated by the grammar at all.

The second question seems more difficult to answer, and Coetzee (2004) does not say anything about it (the words 'social' and 'sociolinguistic(s)' occur in this dissertation only one time each, the word 'informal' occurs only in the sense of the author giving an informal definition of some concept). We would presumably have to say that a sociolinguistic module would operate after Eval, to determine which forms will be actually pronounced in a given situation. It seems very hard to model cases, then, in which sociolinguistic and grammatical factors *interact*.

2 Social Variation

2.1 Variation linguistics

The study of variation linguistics has been quite succesful in the past decades; but its successes were largely ignored by generative grammarians until quite recently (and vice versa). Gradually, a mutual understanding seems to be coming into being. As far as phonology is concerned, an important reason for this might be the advent of Optimality Theory, which seems more suitable for modelling variation than some of its competitors.

A useful distinction to be made here is the Chomskyan dichotomy between:

- *I Language*: 'internal/intensional' language, roughly (and among other things) the language system as it is represented in the individual speaker
- *E Language*: 'external/extensional' language, roughly (and among other things) the language corpus as it is produced by the language community

In spite of some statements to the contrary, there is no a priori reason why one object of study would be more 'coherent' than the other. The best thing seems to be to take an instrumental approach to these matters and see to what extent a certain point of view is succesful, i.e. whether it leads to new insights. Chomsky has expressed a very similar opinion himself:

> There has been much impassioned controversy about [...] the question of how languages should be studied. The controversy is pointless, because there is no right answer. If we are interested in how bees communicate, we will try to learn something about their internal nature, social arrangements, and physical environment.
>
> These approaches are not in conflict; they are mutually supportive. The same is true of the study of human language: it can be investigated from the biological point of view, and from numerous others. Each approach defines the object of its inquiry in the light of its special concerns; and each should try to learn what it can from other approaches.
>
> (Chomsky, 2000)

So, let us try to see what we can learn from the other approach. We will take variationist linguistics in the sense of Labov, as it is laid out in Labov (1993, 2001), but the basis was laid out already in Weinreich *et al.* (1968). (We obviously can only give a very sketchy overview; Labov has an interesting website, at `http://www.ling.upenn.edu/~wlabov/`.)

The difference between I-language and E-language for Labov is mainly a difference between fundamental research attitudes: 'idealist' and 'materialist'. Idealists base their research on native speakers' judgements on how they should speak, not on the way they actually speak. Materialists, on the other hand, base their work "on the objective methods of observation and experiment". This methodological difference is the result of a more fundamental difference in the view of the object of study.

> Since judgments of acceptability differ radically and unpredictably across individuals, it is normal [within the idealist approach] for any disagreement about data to be answered by narrowing the unit of description to the 'dialect' of an individual, usually the theorist. Since each individual derives the rule system from fragmentary data, it is generally held that the community is an inconsistent mixture of consistent individuals.
>
> The materialist position begins with the study of the heterogeneity of the speech community, and reduces this variation to a series of regular quantitative patterns controlled by social factors.Ê Early statements about the speech community emphasized this 'structured heterogeneity' as the fundamental feature of the speech community, maintained by a uniformity of social evaluation.Ê More recently, the uniformity of these variable patterns has been found to be also based on a structural homogeneity.Ê In cities of a million or more population, the basic categories ands rules that define the variables are almost constant across social class, sex and age.Ê This reinforces the position that the funda-

> mental unit of description should be the language of the speech community, and that the speech of an individual can only be understood against this background.
>
> (Labov, 1987)

The most important point is that language variation is structured in some way. We can predict to a certain extent which variety an individual will use if take into account the position he (or she!) takes in society. Idealism cannot explain this because it falls outside of its scope.

The term 'structured heterogeneity' played an important role in the 1960s and 1970s – it was opposed to the 'homogenuous speech community' in which the Chomskyan language learner was supposed to grow up. The criticism against the latter notion was not so much that it was an idealisation of reality (every science needs to idealize its object of study to some extent) but that it obscured the fact that the non-homogenous reality was not completely random, but can be subjected to analysis.

It is confusing that other elements also entered this debate, viz. a discussion on the validity of the methodology of generative linguistics (syntax in particular). The 'materialist' criticism can also be found in the phrase 'usually the theorist' above. According to Labov and his coworkers using your own judgements is not a good way of gathering data, in particular if there is disagreement among speakers. Sombody who is acquainted with the theoretical issues at hand, should abstain from giving judgements, because he can be led subconsciously by that extralinguistic knowledge. This criticism is based at least in part on something that Labov called the 'observer's paradox':

(21) Observer's paradox: when observing or interviewing people to find out about their spontaneous speech, researchers will, by their own presence and participation, influence linguitic behaviour that is observed

We could see giving judgements as one form of linguistic behaviour (and not as some clear window on competence). In this case, the theorist is influencing herself. This criticism is probably justified, but it is not clear that it has ever been very urgent. There do not seem to be many generally accepted results which are based exclusively on dubious or even controversial judgements. More importantly, this debate is not the same as the discussion about homogenuous vs. heterogenuous societies, or even the use of judgements vs. the use of corpora. One can be a fundamentalist idealist and never abandon the idealisation of a homogenous society and still criticize a methodology which is purely based on the idiolect of an individual researcher.

The following is also important in this connection:

> There is [...] a marked asymmetry between the two bodies of linguistic activity: those doing empirical analysis can use

> the formal, qualitative analyses developed under an idealist program, but not visa-versa.Ê The latter are satisfied to construct rule schema without testing for validity against the data of speech production, while the former are not.
>
> This transition from qualitative to quantitative analysis is a familiar one in the development of science.Ê But the qualitative model of linguistics is not easily displaced.Ê Many forms of linguistic behavior are categorically invariant.Ê Furthermore, the number, variety and complexity of linguistic relations are very great, and it is not likely that a large proportion can be investigated by quantitative means.ÊAt present, we do not know the correct balance between the two modes of analysis: how far we can go with unsupported qualitative analysis based on introspection, before the proposals must be confirmed by quantitative studies based on observation and experiment.

Labov has protested against applying the label 'sociolinguistics' to his work for a long time, and this quotation makes it clear why: in his view the 'other' form of doing linguistics is merely part of what he wants to do.

An example: Fronting of /aw/

I will now give a quick example of how a Labovian analysis work. Like in many other cities in North America, the vowels in Philadelphia are shifting ÐÐ American city dialects are slowly growing apart. The vowel /aw/ in *south, out, down, now* undergoes a process of 'fronting' through [æo] to [e:o]. This change is not abrupt but gradual:

(22)

The y axis represents F2 values. One of the puzzles becomes clear immediately: how is it possible that the phonetics does not just change from one generation to the next (so that we would have two straight horizontal lines connected by a 'jump'), but also the direction in which the change is taking place. To be more precise, this is true if we assume that these graphics do not just represent average numbers – in which case this would be the result of the fact that more and more people take the same single qualitative step.

We will see below that this assumption is justified. Furthermore, this age stratification can be correlated with a social stratification:

(23)

Even more spectacular is the next picture in which social and age information are both represented (and which furthermore shows that we are not just dealing with averages over individuals which are all taking or not taking some qualitative step):

(24)

We see here that the F2-value increases as the speakers become older, while at the same time the relative distance between speakers of the same age but different social background continues to exist (or even stays the same). This is what is meant by 'structured heterogeneity': if we know somebody's age and social class, we can predict pretty accurately what this person's F2 values will be. (There are a few social classes which behave in a bizarre way, but this does not change the fact that there is a clear structure in these data.)

Linguistic change (or age difference) is always accompanied by social stratification, according to Labov; and since we expect that these mechanisms are universal, we assume that the same has been true for all language changes of the past.

Another relevant facto is gender. The next figure is also very characteristic for a certain type of language change (viz. one in which age and class are correlated in the way we have seen for Philadelphia /aw/ fronting):

(25)

In the first place we notice that women are 'ahead' of men: women of about 60 years old talk approximately in the same way as 40-year old men, and women in their 30s talk like teenage boys. But apart from this it is striking that the distribution gives almost a straigh line for women: if we know the age difference between two women, we can predict the difference in F2 values fairly accurately. For men, the chance seems more like a generation issue: we can idealize the male line as a 'staircase': men between 20-50 speak in approximately the same way, just like all boys younger than 20 and all men older than 50.

How can we explain all of this. The core of the argument is that all children learn literally their mother tongue during the first years of their life – their vowel systems, and more concretely, their F2 values for /aw/ are those of their parents, especially their mother. When the children go to school, a difference between the sexes arises. While the boys stick to their mother's system, the girls start paying close attention to their peers, i.e. girls of their own age and a little bit older. Now suppose that at some moment older girls have a lower F2 value than their mothers (for whatever reason; we do not have a theory about how language changes is instantiated). Younger girls will notice this, and they will assume that a lower F2 gives a positive social signal. They will than be more likely to display a tendencey to make their F2s even lower than to make them slightly higher. This is the force that will give us a drift which can last for many generations, and which explains how language learners sometimes do not just pick up a new change, but they also seem to know the direction in which this change is going (and go further in that direction than previous generations).

How about the boys? They stick to their mother's system, and this explains why men between 20-45 speak like women of 60: those women are their mothers. Because the mothers of men of, say, 45 do not all have an equal age, but they all are of the same generation, we find an average in this case: this explains the staircase.

This model thus does not explain how language variation comes into being, but it can explain how it spreads once it has entered a community. Notice by the way that the parameter we are studying here is rather superficial-

phonetical and furthermore gradual/non-categorical: we are only talking about F2 values. Like most other work on E-language, quantitative sociolinguistics is often limited to measurable units of language. There is no principled reason why grammatical factors are not considered, except for the fact that these are more difficult to quantify.

2.2 The grammatical structure of the speech community

We will now turn to possible ways of using insights of variational linguistics in our study of grammatical structure. Specifically, we will try to understand what are the implications of the fact that e.g. social class *differences* stay the same over time, even if every class itself shifts.

Related to this is a hypothesis put forward by Gregory Guy of NYU (Guy, 2004):

(26) *Shared Constraints Hypothesis*:
The members of a speech community share common values for the probabilistic constraint effects on variable linguistic processes.

Here is an example Guy (2004) gives to illustrate this point. Philadelphia English and New York English both have a process of coronal stop deletion, like many other English dialects. Yet the conditions under which the process applies are different for the two dialects. This is true for instance if we consider the following consonant. In the following table (C>V) means: there is more deletion before a consonant than before a vowel.

(27)

	C > V	C >∅	V >∅	Preferred order
Philadelphia (N=19)	89	100	95	C>V>∅
New York (N=4)	100	50	0	C=∅>V

The table should be read as follows: of the (four) New Yorkers who have been participating in this study, 100% showed evidence that the context before a consonant has more deletion than the context before a vowel, 50% showed evidence that the preconsonantal position has more deletion than the one before a pause, and none showed evidence that the position before a vowel was more favourable than before a pause (so the reverse seems to be true). Thus, it seems to be true that preconsonantal and prepausal positions are more favourable for deletion than the position before a vowel.

It is important to realize that this is true, independent from the fact that some New Yorkers have (much) more deletion than others. Similarly, the Philadelphians (almost) all followed a somewhat different system (viz. one in which prevocalic positions are more favourable to deletion than postvocalic

positions), even though some showed far more deletion than others (which is something you cannot see in the table above).

We thus could say that the righthand column in (27) somehow represents a fragment of the grammars of Philadelphia and New York English respectively. The two speech communities thus have slightly different grammars. The variation within each community is restricted to the amount of deletion a speaker has (e.g. younger people have more deletion than older people).

Guy (2004) formulates his Shared Constraints Hypothesis in terms of 'probabilistic constraint effects' as in (26). Yet it is also possible to formulate a similar insight in terms of Optimality Theory. In order to see how this works, let us set up a grammar for these two languages. We use three constraints that are well known from the OT literature (we will abstract away from the fact that this process only involves coronal stops, no other consonants):

- NOCODA: Syllables should not end in a coda (Prince & Smolensky, 1993)
- FINALC: Syllables should not end in a coda (McCarthy, 2003a, and references cited there)
- NOHIATUS (Rosenthal, 1997): Two vowels should not occur adjacent to each other.

These three constraints together give us a sixfold factorial typology, which among other things makes sure that /t/ is always deleted before a consonant:

(28)

1.	FINALC≫NOHIATUS≫NOCODA	No deletion except before consonant
2.	FINALC≫NOCODA≫NOHIATUS	No deletion before pause; deletion elsewhere
3.	NOCODA≫FINALC≫NOHIATUS	Deletion everywhere
4.	NOHIATUS≫FINALC≫NOCODA	No deletion except before consonant
5.	NOHIATUS≫NOCODA≫FINALC	No deletion before a vowel; deletion elsewhere
6.	NOCODA≫NOHIATUS≫FINALC	Deletion everywhere

Interestingly, the systems 1-3 together form the system of Philadelphia: they generate exactly the types outputs that supposedly are possible for individual speakers of this dialect (in other words, speakers from Philadelphia choose one of these grammars). Similarly, the systems 4-6 form the system of New York (at least under the idealising assumption that in this system there still is a difference between preconsonantal and prepausal position). We can now characterize Philadelphia as (29a) and New York as (29b):

(29) a. Philadelphia: FINALC≫NOHIATUS, with floating NOCODA
b. New York: constraintNoHiatus≫FinalC, with floating NOCODA

We now have effectively a floating constraints analysis of the two dialects. Yet this still leaves certain questions unanswered. For instance, we could have divided up the six constraint rankings in (29) in different ways: is it a coincidence that here we have two dialects where it is NOCODA which is

floating? Unless we find an answer to such questions, the OT translation of the Shared Constraints Hypothesis (26) does not seem to have a lot of substance. It is not really informative to say that a certain speech community has certain constraint rankings in common.n

One observation we can make is that NOCODA on the one hand and FINALC and NOHIATUS play different roles in our analysis of this phenomenon: NOCODA is there to cause a change (deletion of /t/) whereas the other two constraints block this change from happening. It is not necessarily the case that these constraints are conservative in this way: in some languages FINALC might be satisfied by insertion of some consonant or something similar, but none of this is the case here. We thus reach the following generalisation:

(30) The conditions to which a phonological process is subjected are ranked in the same way in a speech community, even though the process itself may be variable.

Let us now see whether we can give more substance to this generalisation. In order to do this, we now turn to intraspeaker variation, assuming that social variation is reflected faithfully within the grammar of the individual.

Style levels

In order to account for linguistic phenomena related to the formality of speech, generative phonologists have traditionally taken recourse to the notion *style level* or *register* (Selkirk, 1972). Every register is a single, non-variable generative grammar and a language system consists of several registers. The language user selects his register within the system depending on the situation.

It is obviously necessary to restrict the ways in which two styles within one language system can differ from each other. In many rule-based theories of phonology two styles A and B can differ from each other because A has more rules than B, of because the forms of some of the rules in A is more general (contains a smaller number of specified feature values) than the form of those in B, or because A and B have the same rules ranked in different orders. All of these possibilities are attested in the literature. This means that in principle two registers could differ from each other in the same way as any two grammars can differ. If we do not limit these differences, we would expect languages in which for instance one informal style has the phonology of informal Chinese, involving tone, tone sandhi, and a fairly well-developed syllable structure, whereas a more formal style would resemble for instance formal Turkish, involving vowel harmony, stress and a simple syllable structure. It seems improbable that such a language system can actually exist; the

possible difference between any two registers within one language system seems much more restricted.

We are talking here about 'one language system' in the same (loose) way we talked about 'one language community' in the preceding discussion: it is not entirely clear where knowing more than one register ends and bilingualism starts, but it should be possible to give an independent definition.

Within OT, the only possible source of difference is the ranking of constraints. Yet this can still not be the complete solution since constraint ranking is a very powerful tool. If the assumptions of ('classical', 'traditional', 'conservative') OT are correct, also the differences between the phonologies of informal Chinese and formal Turkish can be characterised by differences in constraint ranking only. For this reason it is still useful to try and find a more restrictive interpretation of the notion 'register'. van Oostendorp (1997) proposes such an interpretation.

One of the most important results of the work on the universal set of constraints (Con) within OT, is that this set can be divided into two subsets. The first subset consists of well-formedness, such as the constraint against onsetless syllable ONSET or the constraint disallowing front rounded vowels *[-back, +round]. The latter constraint would be ranked high in languages, such as English, which disallow these vowels. The second subset consists of so-called faithfulness constraints requiring phonological output forms to be maximally similar to the input.

van Oostendorp (1997)'s suggestion is that this distinction plays an important role in the characterisation of style level differences, more specifically that these can be characterized as follows:

(31) The more formal the register, the higher ranked the faithfulness constraints.

Every language system has a fixed ranking of the faithfulness constraints and a fixed ranking of the well-formedness constraints. In addition, the hypothesis in (31) allows us to take any two registers in a language system and predict which of the two is the more formal. These are the reasons why I think the hypothesis in (31) is worth to be explored.

van Oostendorp (1997) gives examples from a variety of languages. Here, we will restrict ourselves to one, from Turkish. (Certain) colloquial varieties of this language show epenthesis: if a word (again, a loanword) starts with a cluster of consonants, an epenthetic vowel is inserted. This vowel is [ɪ] in slightly informal registers, but it is harmonic in the most informal registers (Clements & Sezer, 1982):

(32)

	formal form	*less formal form*	*informal form*
'fetters'	pranga	pıranga	pıranga
'prince'	prens	pırens	pirens
'test'	prova	pırova	purova
'announcer'	spiker	sıpiker	sipiker
'cruiser'	kruvazör	kıruvazör	kuruvazör

In order to analyse these facts we need four constraints. Two wellformedness constraints (e.g. and), requiring vowel harmony and vowel epenthesis respectively, and two faithfulness constraints blocking insertion of non-underlying material: (29c) blocks insertion of vowel roots and (29d) blocks insertion of association lines.

(33)
a. SPREAD-F. If a feature F is linked to one segment in a word, it should be linked to all segments in that word.; the relevant instances for this constraint scheme in Turkish are SPREAD-[front] and SPREAD-[round].
b. NOCLUSTER: *C_1C_2 in the onset
c. NOEPENTHESIS: A vowel in the ouput form should be present in the underlying form.
d. NOSPREADING: An autosegmental association between a feature and a segment in the output form should be present in the underlying form.

The relative ranking in the language system of the wellformedness constraints Spread and NoCluster in this case is as hard to establish as the relative ranking of the faithfulness constraints with respect to each other. This is not crucial however, since we can still see how the hypothesis in (1) gives the correct registers (plus one more, to be discussed below):

(34)
a. careful register:
NOEPENTHESIS≫NOCLUSTER
NOSPREADING≫SPREAD
b. less careful register:
NOCLUSTER≫NoEpenthesis
NOSPREADING≫SPREAD
c. colloquial register:
NOCLUSTER≫NOEPENTHESIS
SPREAD≫NOSPREADING

There should be put some upper and some lower limits on the variation: every style has vowel harmony in the case of stem-affix combinations, meaning that we should probably refine the definitions of the constraints SPREAD

and NOSPREADING in the appropriate way, perhaps adding some more constraints to be able to account for the full range of complexity of the harmony processes in the language.

One other possible ordering of the constraints would be the following:

(35) NOEPENTHESIS≫NOCLUSTER
SPREAD≫NOSPREADING

This would have approximately the same level of informality as the 'less careful' register: one faithfulness constraint dominates a wellformedness constraint and one faithfulness constraint is crucially dominated by a wellformedness constraint. Yet the result of NOEPENTHESIS≫NOCLUSTER is that no epenthetic vowels occur. This makes the actual ordering of SPREAD and NOSPREADING irrelevant.

Based on considerations such as this, van Oostendorp (1997) considers the possibility that something like (31) functions as a universal principle of the language system. The learnability problem which is one of the foundations of generative grammar obviously also arises in the case of registers. It seems unlikely that a child will learn three (or more) separate grammars like she would have to if there were no formal relation between the registers of one language system. If (31) were to be a universal principle, the child would only have to learn (i) the relative rankings among the faithfulness constraints, (ii) the relative rankings among the wellformedness constraints and (iii) the upper and lower limits for the (upper and lower) faithfulness constraints with respect to the hierarchy of wellformedness constraints. This is still a considerable task but in any case it is easier to fulfill than acquiring three completely different systems.

But now consider the fact that (31) is very similar to (30): faithfulness constraints obviously belong to the 'conditions to which a phonological process is subjected'. Furthermore there is an obvious relation between style variation within the individual and social stratification. We can now go one of two ways. Either we consider (31) to be a special case of (30) (one way in which the latter can be refined); or, inversely we reanalyse the Philadelphia and New York grammars in such a way that the relevant constraints are not FINALC and NOHIATUS (obvious markedness constraints), but faithfulness constraints (for instance, positional faithfulness constraints, ANCHORing the final consonant of the word or phrase and the first consonant of the syllable).

3 Geographical variation

Different from social variation, geographical variation is probably not reflected in an interesting way in the linguistic system of a language user. It is of course very common that somebody knows more than one dialect of a

language, but there is little reason to think that these dialects will be related to each other: bidialecticity is some form of microbilingualism (and obviously interesting as such).

It has been mentioned already in chapter 1 that there are three reasons to study geographical variation: individual dialects are interesting in their own right; comparison of closely related systems can shed more light on how one system is organized; and the existence of geographical variation itself poses certain questions. We will go into these reasons in more detail now. In this class, I will mainly discuss examples from dialects of Dutch (and Frisian), since these are my own chief object of study, but it should be clear that similar arguments can be made based on any region in the world.

3.1 Dialects as language systems

The Netherlands have two officially recognized 'state languages', Dutch and Frisian, of which the latter plays an official role only within the province of Fryslân. Both languages belong to the West-Germanic branch of Indo-European. There are spates of Frisian speakers in Germany as well, but their speech is hardly mutually understandable with those of the Dutch speakers of Frisian (the latter variety is called West Frisian, or Westlauwers Frisian in the frisian literature).

(36) Dutch province of Fryslân (with the capital city Leeuwarden/Ljouwert):

West Frisian has a number of dialects itself, among which is Klaaifrysk 'Clay Frisian'. Like most Frisian dialects, Klaaifrysk displays an alternation between schwas and syllabic sonorants. (One of the topoi of the Frisian phonology literature concerns the question whether these alternations are due to a process of schwa deletion or rather to schwa epenthesis.)

One generalisation we can make is:

(37) a. Generalisation I. We do not find schwa between any consonant and a liquid or /m/; in those positions we rather find that the sonorant is syllabic.

b. Examples for Gen. I.

passer 'pair of compasses'	*pɔsər	pɔsr̩
woartel 'carrot'	*vwɑtəl	vwɑtl̩
biezem 'broom'	*biəzəm	biəzm̩

Syllabic sonorants and schwa are in a form of Ôcomplementary distributionÕ: the former occur in exactly those environments where the latter do not occur. What can explain this specific distribution? The following constraint may provide part of the analysis:

(38) *CLOSEDSCHWASYLLABLE
Schwa occurs in open syllables only.

The cross-linguistic relevance of this constraint (or rather of a somewhat more sophisticated machinery having the same effect) has been established by van Oostendorp (2000a). There it is shown that *CLOSEDSCHWASYLLABLE has as an effect in French that schwas that end up in a closed syllable are ÔstrengthenedÕ to [ɛ]. This is something we can see in alternations such as those in (39):

(39) *appeler* 'to call' [ɑ.pə.le] *appelle* '(I) call' [ɑ.pɛl]

There is an obvious reason why Generalisation I mentions (a subset of the) sonorants. These segments are allowed to surface as syllabic, for reasons of sonority. In terms of Optimality Theory, we could say that *CLOSEDSCHWASYLLABLE is subordinated to another constraint, SONOROUSNUCLEUS:

(40) SONOROUSNUCLEUS
The head of the syllable should be a sonorant consonant or a vowel.

We may now wonder whether other sonorant consonants, in particular /n/, display the same behaviour. We know that /n/ in syllable rhyme has a tendency to assimilate to obstruents in its vicinity. Here we can state the following generalisation:

(41) a. *Generalisation II*: We do not find schwa between a non-fricative and tautosyllabic /n/; in this context /n/ assimilates to the preceding plosive and becomes syllabic.
b. *Example for Gen. II*: *iepen* 'open' [iəpm̩] /iəpən/

An interesting aspect of this generalisation is that its left-hand context is more restricted than that of the generalisation we have discussed before. In the former case, the consonant could be preceded also by fricatives, but apparently, this is not the case if the sonorant is /n/. In the context before fricatives, we have a different generalisation:

(42) a. *Generalisation III*: A fricative cannot be followed by a syllabic nasal. Such a cluster is always split up by a schwa.
b. Examples for Gen. III:
even 'even; just a while' [evən] /evən/ *[evn̩]
te dragen 'carrying (gerund)' [draɣən] /draɣən/ *[draɣŋ̩]

Why do fricatives behave differently from non-fricatives. This must be due to the assimilation behaviour of coronal nasals, combined with a generalization that has been introduced into the phonological literature by Padgett (1994) (cf. Visser, 1997; van Oostendorp, 2000b).

(43) *Padgett's Generalisation*
If [+nas, +cons], then [-cont]
'A nasal consonant cannot be linked to the feature [+continuant]'

If a nasal is next to an obstruent, it has to assimilate. But assimilation to a fricative is forbidden by Padgett's Generalisation. The only solution in this case is to have a schwa between the fricative and the nasal, so that assimilation is not necessary. That this idea is on the right track is confirmed by the following facts:

(44) a. *Generalisation IV*: We do not find schwa between a coronal fricative and /n/; in this context /n/ is syllabic.
b. *Examples for Gen. IV*:
tassen 'bags' [tɔsn̩] /tɔsən/
huzen 'houses' [hyzn̩] /hyzən/

Clearly in this case there is no need for assimilation, because the two segments are already specified underlyingly for [coronal] (or because both remain underspecified).

Although Padgett (1994) lists a number of consequences of the generalisation in many different languages, this particular effect (of interaction with vowel deletion/insertion) was not among them. On the one hand, an apparently idiosyncratic point in Klaaifrysk phonology can thus be understood in terms of universal grammar; on the other hand, these data fill a gap in our knowledge (and may pose a puzzle for some theories of spreading).

3.2 Comparing adjacent dialects as a heuristic means

It is completely legitimate to restrict ourselves in our study of dialects to one single variety, but things become even more interesting once we start comparing closely related systems. Such a comparison can help us, for instance, to understand the system we are interested in even better. One example

is that we can get a better understanding of the vowel system of Standard Dutch if we compare it to some dialects of the language.

This vowel system can be divided into two (largely) disjoint subsets which can be distinguished primarily by their phonotactic distribution (taking [aː]-[ɑ] as an example pair):

(45)	A.	r[aː]	r[aː]m	r[aː]p	*r[aː]mp
		'yard'	'window'	'turnip'	—
	B.	*r[ɑ]	r[ɑ]m	r[ɑ]p	r[ɑ]mp
		—	'ram'	'quick'	'disaster'

A-vowels can occur before 0 or 1 consonant, B-vowels before 1 or 2 vowels (at the end of the word). Phonetically, the two groups can be generally distinguished in two ways: A vowels are usually longer than B vowels, and als A vowels are considered to be [tense] or [+ATR], while B vowels are [lax] or [-ATR].

The easiest way of understanding the facts in (45) is by taking the length as 'phonological'/underlying and deriving the tenseness. If A vowels occupy two positions in the rhyme, and B vowel occupy one position, we can make the following claim about syllable structure:

(46) a. A syllable rhyme has to occupy exactly two positions
b. At the end of the word, a syllable rhyme can be followed by one additional consonant

Because an A vowel already has two positions, it can be followed by at most one consonant (but it does not have to); because a B vowel has only one position, it has to be followed by at least one and at most two positions.

The alternative theory, based on Tenseness, seems much less attractive:

(47) a. A tense vowel has to be in an open syllable, a lax vowel has to be in a closed syllable
b. The syllable rhyme contains at most two positions
c. At the end of the word, a syllable rhyme can be followed by one additional consonant

In particular the claim in (47a) looks suspicious: why would there be this relation between tenseness and syllable structure? Despite its relative attractiveness, there are quite a few problems with the theory in (46). For instance, it makes us assume that Standard Dutch does not have the syllable type CV, in spite of strong claims that this is universal. It also causes problems in our analysis of Dutch word stress, since 'long' vowels do not make a syllable heavy (different from closed syllables and from diphthongs). On the other

hand, the typological objections against (47) have become less clear, because analyses along these lines have been proposed for related languages such as English (e.g. Hammond, 1997), German (Féry, 1997) and French (van Oostendorp, 2000a; Féry, 2001). For the latter language, the difference may be the clearest: the difference between [ɛ] and [e], or between [ɔ] and [o] clearly corresponds to a difference between closed and open syllables, but there is no reason to assume that French distinguishes between 'long' and short vowels. We thus seem to need something like (47a) in our description of this language independently.

Now a sensitive blow to the theory in (46) comes from the study of Brabant dialects of Dutch.

(48) Brabant Dutch area (with Tilburg in the north, Antwerp, Hofstade in the south)

Most dialects which are spoken in the Brabant area have a threeway (and sometimes even a fourway) distinction between vowels. Tilburg Dutch, for instance, has the following vowels (Swets, 2004):

(49) *Tilburg vowel system*

a. tense vowels

i	y	u
e	ø	o

b. short lax vowels

ɪ	ʏ	ɔ
ɛ	ʌ	ɒ

c. long lax vowels

ɪː	ʏː	ɔː
ɛː	ʌː	ɒː

Yet even in this case, the tense vowels behave as long (they do not allow more than one consonant). Now one could still argue that tense vowels are

long by default (in other words, what is missing in Tilburg are the short tense vowels). Yet the Antwerp dialect shows that this relation is not a necessary one: this dialect probably has a real underlying length distinction, but here all lax vowels are long and almost all tense vowels short (Nuyts, 1989):

(50)	st[i]pt	'prompt'	g[ɪː]r	'scream'
	sp[e]l	'play'	b[ɛː]k	'brook'

And what is more, the Hofstade dialect (Keymeulen & Taeldeman, 1985) has a complete cross-classification of tenseness and length (so both tense and lax long and short vowels). All in all this dialect has 25 distinctive vowels, some of which are listed below:

(51)	w[i]t	'white'	w[iː]l	'wheel'
	b[e]lt	'image'	v[eː]l	'much'
	[ɛ]mme	'skirt'	w[ɛː]t	'wide'

Yet even in this dialect, the short tense vowels do not seem to occur in a context before more than one consonant. We need something like (47a) to describe this. But if a statement to this effect is included in Universal Grammar, there is no reason why it could not be referred to in the analysis of Standard Dutch, and this in turn lifts the main argument against the analysis in (47).

3.3 The implications of geographic distribution of phonological facts

The geographic distribution of facts can at least be seen as indication of the way in which language systems have changed ("Aus dem räumlichen Nebeneinander ein zeitliches Nacheinander").

Lloret (2004) applies the Optimal Paradigms framework of McCarthy (2003b) to certain facts of the Catalan verb. Her point of departure is a comparison between Alguerese (A) and Balearic (B) on the one hand, and Central (C) Catalan on the other. All three dialects have vowel epenthesis in certain illicit clusters in the noun (52a); but the dialects A and B do not have the same restrictions in the first person singular form of verbs (which is without an ending), whereas the C dialect does (52b).

(52) a.

A	*B*	*C*	
sófra	sófrə	sófrə	'sulphur MASC'
séntra	séntrə	séntrə	'center MASC'
ra.táw.la	rə.táw.lə	rə.táw.lə	'altarpiece MASC'

b.

A	*B*	*C*	
éntr	ənt*ɾ*	ént*ɾ*u	'I enter'
úmpr	úmpl	ómplu	'I fill'
res.táwr	rəs.táw*ɾ*	rəs.táw*ɾ*u	'I restore'

We thus see a difference here between the nominal and the verbal paradigm. Lloret (2004) assumes that the three dialects have the following verbal conjugations (for the so-called Conjugation I verbs — the other two classes we will disregard here):

(53)

	A	*B*	*C*
1SG	∅	∅	u
2SG	as	əs	əs
3SG	a	ə	ə
1PL	ém	ám	ɛ́m
2PL	áw	áw	ɛ́w
3PL	an	ən	ən

All suffixes except for the 1SG in A and B Catalan are vowel-initial in this table. Because of OP, 1SG then will also behave as if it is followed by a vowel as well, and specifically, it will not allow vowel epenthesis.[3] The stem of Balearic 1SG séntrə+∅ would be to different from 2SG séntr+əs or 1PL sentr+ám, which do not have a vowel in the stem (and 2SG séntrə+əs is excluded because of a constraint against two unmarked vowels in a row).

It is not true that the stem of the 1SG verb is exactly identical to stems in other parts of the paradigm; Lloret (2004) lists several segmental process which distinguish it in different dialects. One of them is final devoicing, which makes this position perfectly comparable to other final positions, but different from other forms in the paradigm:

(54) a. 1SG: *aca[*p*]* 'I finish' (cf. *aca[*b*]e* 'he finishes')
b. *Other final positions*: *tu[*p*]* 'tube' (cf. *tu[*b*]et* 'tube DIM')

Obviously, this fact does not necessarily constitute a problem: the OP Faithfulness constraints against vowel insertion may be high ranked even though the OP Faithfulness constraints on consonant identity are low ranked. We might even expect this kind of phenomenon to happen if we assume that there are many OP Faithfulness constraints, and they are ranked independently.

[3] With Lloret (2004), we will assume that the vowel is the result of epenthesis. With some complications, the analysis would actually also work if we assume that the schwa can also be underlying, as we are forced to do, given Richness of the Base.

On the other hand, Lloret (2004)'s analysis makes apparent some of the more problematic aspects of the OP paradigm. We already noted in connection to McCarthy (2003b)'s analysis that it needs to assume that affixes are given, and cannot be changed. This is crucial for Lloret (2004) as well: the analysis breaks down as soon as we postulate that the first person singular could be séntr+ə rather than séntrə+∅, i.e. that the epenthetic vowel could be part of the suffix rather than of the stem. Nothing would block the former structure within the OP paradigm.

Notice in this connection that the vowel in the unstressed suffixes is exactly the same as the epenthetic vowel ([a] in Alguerese and [ə] in Balearic and Central Catalan) in (53). This seems an unlikely state of affairs, and it is not a priori clear why we would not postulate that the consonants on their own are the suffixes, and the vowels are indeed epenthetic. Yet, we cannot say this within this theory, since this would force OP to be violated, and since the majority of suffixes would now actually be consonant-initial (the exceptions would still be the stressed suffixes, which have a vowel of unpredictable quality), the stem might be attracted to go with the majority, i.e. to force epenthesis rather than to block it.

In brief, the problem of this approach is that we have to abandon Richness of the Base. Lloret (2004) briefly discusses an alternative option, which she attributes to Mascaró (1983); Dols (2000), among others. In this view, the 1SG is not absent, but it is an empty vowel, licensing the preceding complex onset. Therefore, vowel epenthesis is not necessary in this case. Lloret (2004) points out that there is a problem with this approach: final devoicing (and other segmental phenomena) seem to clearly show that the last segment in a word belongs to a coda, not to an onset. Notice, however, that this argument only holds if we assume that the devoicing in question is *syllable final devoicing* only. Now the evidence shows that the syllable is indeed the domain of devoicing normally. However, constraints on word final devoicing presumably should also exist, as we know from the study of other languages (Steriade, 2001). Its effects may not be particularly visible in other circumstances, where its effects could always be due to other factors, but that does not mean that it could not show up here.

Interestingly, there is some evidence from Dutch dialects which seem to point exactly in the direction of this representational approach rather than paradigm uniformity. Coincidentally it also involves the first person singular *and* final devoicing. All dialects of Dutch have final devoicing (just like Catalan). However, in some dialects, the 1SG form of verbs are exceptions to this. This gives us again a difference between verbs and nouns:

(55) a. *Verb*: *ik geleuv* 'I believe' [ɪkxəløv]
b. *Noun*: *geloof* 'belief' [xəløf] (cf. *geloven* 'beliefs PLUR.' [ɣəlovən])

As in the Catalan case, there are two ways of describing the difference between the verb and the noun. A *structural* analysis, on the one hand, assumes that the first person singular has some property which blocks final devoicing, e.g. an phonetically empty suffix vowel. A *paradigmatic* analysis has it that the first person singular should resemble 'related' forms as much as possible; apllication of final devoicing would increase the differences between forms in the paradigm to an unacceptable level. van Oostendorp (forthcoming) claims that there are several problems connected to the paradigmatic (OP) analysis in this case.

i. The *dialect geography* seems to point in a completely different direction. Whereas in most Dutch dialects, the 1SG is not pronounced, in three independent areas it *is*. And the phenomenon shown in (55) is something we find *on the border* of these three areas.
(Goeman, 1999, 216-217) lists a large number of dialects where this phenomenon may be found; furthermore such dialects can be found in quite a large part of the Dutch language area (cf. map 1). The reason Goeman gives for this, is a historical one: the first person singular schwa has been deleted 'recently' and therefore the final devoicing has not yet taken place. We could say that this statement depends on the opacity of diachronic language change: the final devoicing process proceeds as if the historical ending were still there.[4]

(56)

[4]The data are from Goeman (1999) and the Goeman-Taeldeman-Van Reenen Database; the latter is available at http://www.meertens.knaw.nl/; look for 'Morfologische Atlas'/'Morphological Atlas'.

It is of interest that once again fricatives are the main focus of this exceptional behaviour, as is to be expected if we assume that fricative voicing is primarily an issue of syllable positions and those positions can be used to express morphological structure.
It is reasonable to assume that the dialectgeography mirrors language change in these cases. But within a paradigmatic account, there is no particular reason why this geographical patterning should happen: paradigm effects could show up anywhere.

ii. The structure of the paradigm: in at least some of the Dutch cases, it is not clear at all what the voiced segments are faithful to. In some of these dialects *none* of the other endings in the present tense paradigm have a vowel (they are -/st/ and -/t/ respectively). The infinitive does have a vowel, but this raises the question which other forms exactly should be included in the paradigm. Further, even if we would include the infinitive, it is not clear why the other forms in the paradigm would not be attracted to it: if 1SG is pronounced with a [v] because the infinitive is [ɣələøvən], why is the 1PL pronounced as [ɣələøft] and not as [ɣələøvd]?

Notice that the structural approach suffers from neither of these problems. It needs to assume that the 1SG is an 'empty' vowel of some sort. Loss of inflectional schwa (reduction) may then be assumed to go through a stage where the vowel is completely empty: this explains the geographical distribution. Furthermore, the fact that other forms in the paradigm are not affected follows from the fact that these other forms do not have a 1SG ending.

We would have to assume that a relevant difference between Catalan and Dutch dialects is that the former show effects of word-final as well as syllable-final devoicing, but otherwise both systems seem to be adaptable to a structural analysis. There is a trade-off, however: we have to accept an analysis which is to some extent abstract: it postulates phonological categories which are not directly visible in the phonetics, i.e. a requires a view of phonology where the output is not directly the same as the phonetic form. The representations in OP (and other OO Faithfulness accounts) can be and should be more concrete. This does not necessarily mean that the latter are theoretically more parsimonious, however, or even less 'abstract', since they have to postulate many (invisible) correspondence relations between individual segments in all representations, and they have to assume that all these representations are available in the production of one form, even when they are not pronounced. In terms of computational extras, it might well be the 'concrete' approaches which face the most serious problems.

4 Language Change

The topics of language variation is closely connected to that of language change: there can be no variation without change, and inversely, all change

seems to be accompanied with variation. All modern theorizing on language change still has to deal with the ideas of the Neogrammarians ('Junggrammatiker', sometimes also called 'Leipziger Schule') in one way or another. In particular, the work of the most well-known generative theorist on language change, Paul Kiparsky, has been clearly inspired by this school. This is the reason why we will study their ideas in this class, before going into the question how language change should be understood in terms of modern phonological theory.

4.1 'Beyond explanatory adequacy'

Consider the following Optimality Theoretic markedness constraint (Prince & Smolensky, 1993):

(57) NoCoda: Syllables do not have codas.

Like most OT constraints (or, as a matter of fact, most phonological generalisations proposed in any framework) this is a markedness tendency rather than a true linguistic universal in the sense that every language obeys it completely. This constraint serves several purposes at the same time. Most importantly, it expresses:

1. the fact that open ('CV') syllables are universal in human language (there is no language which disallows them), while closed syllables are allowed only in a subset
2. the fact that even those languages which do have open syllables, tend to avoid them: E.g. (tautomorphemic) VCV presumably is universally syllabified as V.CV, not as *VC.V.

Every modern theory of phonology uses some mechanism which expresses these generalisations:

- In rule-based theory (**?**), it is assumed that a rule syllabifying CV is basic and universal, whereas rules syllabifiying postvocalic consonants are language-specific and apply in a later module
- In purely representational approaches, like (certain versions of) Government Phonology (**?**), it is assumed that CV is the only available syllable type. Something that phonetically looks like CVC phonologically really is $CV_1.CV_2$, where V_2 is an empty vowel, subject to a number of specific constraints and (therefore) marked.

In this course, however, we will concentrate on Optimality Theoretic constraints such as the one in (57). If we use NoCoda in some analysis, we may say that this constraint 'explains' a certain set of facts. For instance, some OT analyses of French liaison will use this constraint to 'explain' those facts. The following tables are an example (from **?**)

(58) a.

/pəti(t)/ pinson 'finch'	NoHiatus	NoCoda
a. ☞[pəti] pinson		
b. [pətit] pinson		*

b.

/pəti(t)/ aigle 'eagle'	NoHiatus	NoCoda
a. [pəti] aigle	*	
b. ☞[pətit] aigle		

The constraints NoHiatus and NoCoda are part of an explanation of liaison in these tableaux to the extent that they (i) are independently motivated, and (ii) help to derive the observed results.

At the same time, constraints such as these are obviously themselves in need of an explanation: why are our constraints (or our rules, or our representations) the way they are and not otherwise? In the case at hand, why is there constraint NoCoda and not, alternatively, a constraint Coda? In the terminology of **?**, we are going 'beyond explanatory adequacy' if we try to answer these questions: we try to explain why UG is shaped the way it is.

It turns out that the answer to this question is very much dependent on our idea of the place of phonology within linguistics, or its relation to phonetics:

1. We might assume that these constraints are 'grounded' in the phonetics. E.g., we know that obstruents, and more specifically stops, are harder to articulate and perceive after a vowel than before it. This gives a motivation for NoCoda, whereas Coda is quite absurd.
2. Alternatively, we might try to find an explanation in the way in which cognitive structures are realized. For instance, we may try to relate the fact about syllable structure to the idea of **?** that the syntactic structures of all languages are SVO. If both subjects and syllable onsets are linguistic 'specifiers', we have discovered some similarity to the two. Obviously, still the question needs to be answered *why* specifiers occur on the left-hand side. Under this view, it might even be possible that coda's are more difficult to perceive, because human beings know that they are less prominent **?**.

If we take the 'grounded' position, we have to deal with the question how exactly the phonetics can influence phonology. Here there seem to be roughly three positions:

1. Constraints such as NoCoda are part of Universal Grammar. The problem with this account is one of *duplication*: we have to assume that NoCoda somehow is part of the 'outside world' — the speech signal, the auditory system — and at the same time of the 'inside world' — the innate capacities of human beings. A reason for this might be evo-

lutionary: the language system has adapted over time to the way in which language is used. But it is unclear that there has been enough evolutionary time to get to this point. There is no clear representative of this position ('nativists' seem to usually prefer a cognitive point-of-view).

2. The second option is to assume that the language-learning child constructs constraints such as NoCoda on the basis of what she observes in her own speech and speech errors. The child thus acts as a small experimental phonetician (or 'laboratory phonologist'). This approach has been defended by **?**, among others (see below).
3. The third option is that these constraints are not part of grammar at all. Their explanation has to be sought elsewhere, and the most obvious place to look is the diachrony: languages change because of misperception or misarticulation, and when children acquire the language, they simply pick up whatever centuries of phonetically initiated change have made out of the system.
   Presently, there seem to be two paths to reach this conclusion. One is by assuming that phonology is only about 'hard universals', hence not about markedness. Phonology is a pure cognitive symbolic system in which there is no place for statistical tendencies. Since virtually no principle in phonology is 'hard', this means we void the theory from many modules that used to be part of it. This is the position defended most forcefully in **?**Hale (2003) and related works.
   More or less the same conclusion has been reached by authors such as **?**, starting from the assumption that "language is a self-organizing system, and grammar, including both morphosyntax and phonology, is an emergent property of that system" (p. 190). In this view, phonology is all about statistical tendencies, and symbolic systems, if they exist at all, are seen as epiphenomenal. The child just acquires whatever is available, and this material will have been largely subjected to the principles of language use.
   The two positions converge in the sense that most traditional objects of study for the phonology are delegated to a component of statistics. They may differ as to the role of language acquisition: the cognitive view will suppose that the phonetic facts still have to go through a filter of UG (be it one which is much leaner than traditional grammar has it), whereas **?** basically assumes a blank slate model of the human mind.

All in all, we can see that the current discussion on markedness has connections to many very basic questions in linguistic theory, such as: what is the division of labour between synchronic and diachronic explanation? What is the role of language acquisition in linguistic change? And how are phonetic and phonological explanations of phenomena to be related?

4.2 The life cycle of sound patterns

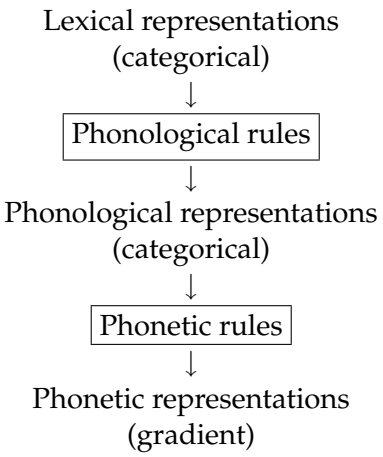
The traditional point of view of the phonology-phonetics interface can be summarised as follows.[5] We assume that language change has its origins in phonetics; this origin will most likely (or in some models exclusively) be in the direction of greater ease of articulation and/or ease of perception. After a while, the results of this phonetic change may first become *phonologized*, and later *morphologized* or even *lexicalized*. **?** traces this idea back to **?** and summarizes the 'life cycle of sound patterns' as follows:

- *Phase I*
  The life cycle begins when, by Neogrammarian sound change, some physical of physiological phenomenon gives rise to a new cognitively controlled pattern of phonetic implementation. This development, known as *phonologization* (**?**), involves the addition of a new phonetic rule to the grammar.
- *Phase II*
  Subsequently, this gradient sound pattern may become categorical. [...][S]uch a change would involve the *restructuring* of the phonological representations that provide the input of the phonological implementation, with the concomitant development of a new phonological counterpart for the original phonetic rule.[...]
- *Phase III*
  Reanalysis can also cause categorical patterns to change. Over time, phonological rules typically become sensitive to morphosyntactic structure, often with a reduction in their domain of application [...] Phonological rules may also develop lexical exceptions [...]
- *Phase IV*
  At the end of their life cycle, sound patterns may cease to be phonologically controlled. Thus a phonological rule may be replaced by a morphological operation (*morphologization*), or may disappear altogether, leaving an idiosyncratic residue in lexical representations (*lexicalization*). [...]

As **?** points out, this view of sound change fits very well into the standard generative view of the synchronic relation between phonology and phonetics, as it is exemplified in models such as Lexical Phonology and Stratal OT. amd which can be summarised as follows:

[5] This section relies heavily on **?**.

(59)



Under this view, then, sound change moves 'bottom-up' in the grammar: a change which originates in the phonetics may in the course of time end up having an effect only in certain lexical representations.

Although **?** does not discuss this point, notice that this view as a matter of fact implies that the explanation of markedness is essentially in the realm of phonetics, because this is where every rule or process will start its life cycle. On the other hand, the process of phonologization (which turns gradient phonetic facts into categorical phonological ones) will be in part the product of phonology. The phonology will then be responsible for the 'universal' aspects and the phonetics for the 'markedness aspects'.

This might be easiest to see within a rule-based framework. Let us suppose that a language L at some point in its history will be subject to a phonetic change by which word-final consonants are gradually reduced. At some point, this might become phonologized to something like:

(60) $C \rightarrow / \_\_\_ \# C$

What has happened, at this point is that a gradient reduction has turned into something categorical: now word-final consonants are deleted completely. he same consonant may still show up in some other environment, e.g. before a vowel-initial word. The fact that this rule is a 'natural' rule (an implementation of NoCoda, so to say) is a consequence of the fact that it has originated in the phonetics, however. The only reason why rules such as (60) are (much) more frequent than rules such as (61) is that the latter does not have a plausible phonetic origin; from a purely phonological point of view, there is nothing wrong about it.

(61) $C \rightarrow / \_\_\_ \# V$

At the same time, the effects of the phonetics may become obscured in due course, because new rules might follow this one. And then at some point,

the rule may become lexicalized: it just happens that some words alternate word-final contexts according to context. (This may be the case of French liaison, which does not affect new words.)

As simple and elegant as this picture may be, there also are a few problems connected to it. Most importantly, it implies that the grammar of every generation is built on the basis of that of previous generations by addition of rules at the end of the grammar. One conceptual problem for this is that this means that every generation of language learners has to be able to see into the heads of their parents directly in order to see the grammars represented there (Hale, 2003).

Also, it is not very clear how this idea can be made compatible with Optimality Theory, precisely. On the one hand, **?** shows that the idea of phonologization/lexicalization can be explained more elegantly in OT than it could in rule-based theory, because of a principle of Lexicon Optimization — we will not go into that here. On the other hand, the only thing that can be manipulated in (classical) OT is constraint rankings. This then leaves the source of the constraint NOCODA still unexplained: if it is in the universal constraint set Con, how did it get there in the first place? If this constraint mirrors a phonetic generalisation, how does it do that? The only possibility would be, in fact, to say that Con contains all kinds of constraints, including NOONSET and CODA, and that the only reason why we do not see the latter is that they are unlikely to ever take effect. (And obviously there is always the alternative of rejecting the thought that the origin of sound change should always be phonetic in nature.)

In recent years, an alternative to the traditional view of sound change and the phonetics-phonology interface has been proposed under the rubric of 'exemplar theory'. In this view, lexical items are not categorical — let alone underspecified. Rather, language users store individual phonetic soundshapes of tokens into their memory. These tokens, which are often referred to as 'exemplars', are associated to each other because they are of course phonetically very similar. But in the extreme versions of this theory, they are not categorized in any way. There is no independent phonological representation of a given word, there is only a network of individual tokens and 'emergent generalisations' (cf. the quote from **?** above). Actually semantically, phonetically and morphologically related words will also be connected, albeit with looser ties.

One consequence of this theory is that it predicts that there is no independent phonology: if generalisations can be made, they are due to phonetics, or processing, or other considerations. There is no grammar for individual languages, so by extension there can be no Universal Grammar, and indeed there can be no phonological universals (apart maybe from a few hard restrictions imposed on us by the vocal tract etc.) Also the existence of productive phonological alternations is effectively denied. Final devoicing in Dutch for instance is presumably 'represented' by the fact that all singulars of nouns

end in voiceless obstruents, and some plurals have a voiced obstruent in a corresponding position.

Another consequence is that language change can only be gradient and lexically diffusing. The reason why it can only be gradient is because there is no categorial phonology, everything is gradient phonetics. The reason why it can only be lexically diffusing (i.e. affecting item by item, not taking one sound in the language and change it in every place where it occurs) is because there is no such concept of 'a consonant following a vowel': there only are individual occurrences of consonants following vowels in individual tokens of words. This means that the network of words may change in the direction of less and less prominently pronounced coda consonants, but there is no particular reason why the networks of other words should move in the same direction at the same time.

All of this obviously means that the whole idea of the life cycle of phonological rules is completely abandoned, which is a little bit too radical for many scholars, as may be the idea that there are no truly phonological alternations. For this reason, more moderate versions of this approach have been proposed, e.g. by Pierrehumbert. In any case, all of this shows that the studies of markedness, historical phonology and the phonology-phonetics interface are strongly intertwined. By studying them together, we may get a better view on each of them individually. This is what we hope to achieve in this course.

4.3 Neogrammarians and grammatical theory

It is no exaggeration to say that the relative 'scientific' prestige that linguistics has among the humanities is due in large part to Neogrammarians. They introduced 'hard' methods into the study of language, bringing rigid logical thinking to bear on it; and they have influenced almost all important linguists of the 20th century and beyonf (including Saussure, Chomsky, Labov). 19th Century linguistics became a prestigious field also outside the humanities. Darwin, for instance, stated in his book *The Descent of Man* that:

> The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel.

And after this he quoted 15 parallels between language change and biological evolution (Labov, 1993). Like many successful branches of humanities in the 19th century, neogrammarian linguistics was primarily historical. The principles that were discovered were historical ('diachronic') principles, as we will see. We will also see what might have been the largest problem for this type of linguistics: the fact that the ontological status of its object of study,

'language', was not clear. One could describe how language changed, but not what it was, exactly, that changed when language changed.

When the neogrammarians arrived on the scene, scholars already knew that there were correspondences between sounds in different language families — (*qu-* /kw/ in Romance (*quod, qui,* etc) corresponds to /hw/ in Germanic (*where, waar, wo*). But the Neogrammarians drew a radical conclusion from facts like these: sounds correspond to each other because languages were derived from a common origin by way of a systematic sound change. This sound change followed universal, exceptionless and phonetically motivated laws and was predictable, as it were. If we did find apparent exceptions, these could be caused by one of three factors:

1. Other sound laws, which had not been discovered before.
2. Analogy with other forms; e.g. the similarities between *feówer* and *fif* in Old-English or between *quattuor* and *quinque* in Latin cannot be explained by some systematic development from IE *kwetwer, *penkwe; in both languages, one of the numerals has developed a sound 'by analogy' to the other numeral. There is an internal force within the language system to be really systematic.
3. Loanwords from other (IE) languages which were borrowed after a systematic sound change had applied could make things look as if there are exceptions.

A well-known example of a sound law explaining apparent 'exceptions' to another sound law is Verner's Law. The original sound law was formulated by Jakob Grimm around 1820 (so before the Neogrammarian period) and described a development in Germanic:

(62) *Grimm's Law*
   a. PIE *b^h, *d^h, *g^h >PGerm. *b, *d, *g (e.g. PIE *bher (cf. Latijn *fer-*) >PGerm. *ber)
   b. PIE *b, *d, *g >PGerm. *p, *t, *k (bijv. Latijn *dens* - Du. *tand)
   c. PIE *p, *t, *k >PGerm. *f, *þ, *x (e.g. Latin *frater-* - Eng. *brother)

There were a few exceptions to this law, but for a long time these were considered to be merely random. The Danish linguist Karl Verner (1846-1896)[6] showed, however, that many of these 'exceptions' were the result of another law, which would become known as Verner's Law.

The exceptions concerned the last part of Grimm's Law. The PIE word for *father* was *pə$_2$teːr; the PGerm. word was *faðeːr. The first plosive obeyed Grimm, but the second was an exception (it should have been þ). The PIE word *bhraːteːr 'brother' on the other hand had followed Grimm's Law, and

[6] Verner (1875). An English translation can be consulted online at `http://www.utexas.edu/cola/depts/lrc/iedocctr/ie-docs/lehmann/reader/Chapter11.html`.

turned into PGerm. *broːþeːr. Even more remarkably, the same sound in the same stem seemed to have differently in different contexts:

(63) PGerm. *werþ 'to turn', *warþ ('he turned'), *wurd (past part.)

Verner argued that PGerm. had known a rule which had voiced all fricatives in certain positions; the voiced fricatives had subsequently merged with the voiced plosives (this was in accordance with Grimm). This context was intervocalic, after an unstressed syllable. In Germanic, the lexical stress differences had disappeared subsequently (all words were stressed on the first syllable), but in Sanskrit, for example, you could still see it (pitáː vs. bhráːtaː). The differences between different forms of the verb 'turn' could be explained by the fact that these had hosted different affixes. We get a completely regular explanation for the 'exceptions' by assuming chronological ordering between these three historic processes (Grimm, Verner, stress shift):

(64)			
	PIE	*pə$_2$téːr	*bhráːteːr
	Grimm	*faþéːr	*bróːþeːr
	Verner	*fadéːr	
	Stress shift	*fádeːr	*bróːþeːr

It is necessary to assume an abstract historical stage (abstract because we do not have any direct evidence) before stress shift and after Grimm. Verner's article showed that it is possible (and desirable) to explain all facts of natural language within linguistic theory, without putting some facts apart as 'exceptions'.

It is usually assumed that according to the Neogrammarians, sound change was (i) exceptionless/systematic/'konsequent', (ii) phonetic.

Paul (1880) gives a slightly more sophisticated picture:

> Wenn wir daher von konsequenter Wirkung der Lautgesetze reden, so kann das nur heissen, dass bei dem Lautwandel innerhalb desselben Dialektes alle einzelnen Fälle, in denen die gleichen lautlichen Bedingungen vorliegen, gleichmässig behandelt werden. Entweder muss also, wo früher einmal der gleiche Laut bestand, auch auf den späteren Entwickelungsstufen immer der gleiche Laut bleiben, oder, wo eine Spaltung in verschiedene Laute eingetreten ist, da muss eine bestimmte Ursache und zwar eine Ursache rein lautlicher Natur wie Einwirkung umgebender Laute, Akzent, Silbenstellung u. dgl. anzugeben sein, warum in dem einen Falle dieser, in dem andern jener Laut entstanden ist. Man muss dabei natürlich sämtliche Momente der Lauterzeugung in Betracht ziehen. Namentlich muss man auch das Wort nicht isoliert, sondern nach seiner Stellung innerhalb des Satzgefüges betrach-

> ten. Erst dann ist es möglich die Konsequenz in den Lautveränderungen zu erkennen.

The Neogrammarian theories have survived unto the present day, and furthermore they had their influence both directly and indirectly on linguists working on language variation and change until the present day.

There seems to have been one major problem with Neogrammarian theories: it was not always clear where in reality language change took place. Implicit in their work is the notion of E-language in stead of grammar (I-language); it was assumed that the existence of 'languages' was real, so that one did not bother to define it explicitly. This was the source of a number of clear conceptual problems, as Kiparsky (2003) observes. Sound change is not as blind as it would appear from the Neogrammarian theories.

a. it cannot create languages which violate universal principles. (All languages have voiceless plosives. Spirantisation of plosives — the third case of Grimm's Law in (62) — is a familiar process, but it does not lead to a language without voiceless plosives.)
b. this is true even for implicational universals ('if a language has property α, it will also have property β')
c. in some languages or language families, we can observe phenomena of 'long-term drift' (e.g. a tendency in Slavic to create open syllables; vowel shift in North American English)

Phenomena of this type are better understood if sound changes do not operate blindly on randomly chosen segments, but rather operate in the context of a language *system*, which in turn is sensitive to universal principles.

4.4 Two views on diachrony vs. synchrony

In this section, we will compare two radically different views of language change based on modern phonological thinking: Kiparsky (2003) (K) and Hale (2003) (H).

It seems that for K the most important factor in explaining 'exceptions' to Konsequenz is in the way phonetic generalisations may be treated inside Lexical Phonology. For H, the most important factor is 'imperfect' learning: because the data for a new learner are different, she may construct a different grammar. K stresses the importance of synchronic linguistics for understanding diachrony; H of diachronic linguistics for understanding synchrony. But the most important difference is this: for K, there is a clear relation between the grammar of an early stage S1 and that of a later stage S2 (S2 has either added a rule, or generalized one); for H the relation between subsequent stages of evolution is not clear at all.

K therefore is closer to the Neogrammarians in the sense that he virtually discusses language/grammar as if it were an independent organism (the

learner merely takes the existing grammar and somehow changes it mildly), which grows and develops; while H is much more radical in taking the role of acquisition as most central.

Diffusion across speakers seems a problem for both K and H. The latter explicitly denies its relevance for phonological theory. Furthermore, one could debate whether acquisition (i.e., in some sense, 'error') really is the only source of change.

K mentions potential problems for eacht part of the neogrammarian hypothesis. We will discuss these problems in turn:

1. Lexical diffusion (Changes which do not seem to be systematic)
2. Structure dependence (Changes which do not seem to be blind)

Ad 1. Lexical diffusion

An example of lexical diffusion (p. 316-317): shortening of /ū/ is generalised on a word-by word basis to other contexts. This is a problem, since it can hardly be seen as 'konsequent', at least during the 'intermediate stages of development' (such as right now).

Accordig to K, lexical diffusion should be understood formally as analogy. An example of 'normal' analogy: cow-kine -> cow-cows. If no separate plural form is available, the plural is formed 'by analogy' to other plurals ending in -s. Extending this to lexical diffusion, requires two steps:

1. The context of a rule is 'generalised' (e.g. the shortening rule is generalised from [-anterior]___[-anterior, -coronal] to [-anterior]___). Words which still have a long vowel now count as exceptions (are marked underlyingly for being long).
2. Forms start losing their underlying markings on a one-by-one basis (= by analogy).

H criticises both points:

1. What is the rationale for this change? In what sense is (1) 'an optimisation'? H assumes that all words which do not undergo the rule at first have to acquire some diacritic marking; in that case the new system would not count as optimal, indeed. But the marking is not necessarily diacritic; if it is phonological (assuming that long vowels have an underlying marking for two moras) there is no real way in which they become more complex. On the other hand, in this case, it is not clear what is the relation between the fact that a rule is generalised (step 1), and the fact that forms loose their underlying marking (step 2); the last step might be taken independently as a way of implifying the lexicon, even if there is no rule.
2. H notes that the conceptual relation to morphological analogy is not clear. The analogical form cows comes into being if for some reason kine is no longer available. But one cannot say that the underlying form good (with a short vowel) comes into being because the underlying

form with a long vowel is no longer available. Because why would it no longer be? (In a sense, this amounts to the same problem as (1), I suppose.)

H also mentions a few other problems:

- 'K derives the change in surface forms from a change in underlying process (the extending of a rule), but this is putting the cart before the horse: the rules are posited by the grammar constructor on the basis of the analysis of surface forms, not vice versa.' I do not follow this criticism, which is based on the assumption that somehow the grammar constructor should build a grammar which as faithfully as possible the attested facts (overgeneralisation is not allowed). It is not clear to me why we need to postulate this.
- Much more important is the criticism on p. 355, where a language is cited in which word-final consonants are dropped on a word-by-word basis. It is not clear how this process could be described in terms of 'underspecification'.

H's own solution is that lexical diffusion is the result of imperfect learning. In the case of /ū/ shortening, there are phonetic reasons that the vowel length may be hard to pronounce or perceive in certain contexts (it is not clear to me what they are, but let us assume they are present); therefore speakers are likely to make mistakes, or the child may mishear certain words, and therefore construct different underlying representations.

But there are several problems with this approach as well:

- K argues that only certain kinds of phonological features are subject to diffusion (viz. those available to the lexical phonology of the language); his model of Lexical Phonology actually predicts this. H makes no such prediction, and his model therefore is less constrained.
- It is not clear how something could ever turn into a phonological rule in this way. E.g. if a rule starts to automatically also affect loanwords, etc., how could we ever explain this, if lexical diffusion is only a matter of changing underlying representations?

Ad 2. Structure-dependence

Sound change often does not appear to be blind, as we saw on page 5.3. In order to account for this, Kiparsky proposes 'a two-level model' of language change, a Darwinian model of variation and selection.:

Level 1. Phonetic change is indeed exceptionless/blind/konsequent;

Level 2. But it is filtered by UG once a child acquires the language.

An example of a UG filter is the following (p. 328):

(65) *Priming effect.* Redundant features are likely to be phonologized if the language's phonological representations have a class node to host them.

For instance: loss of voicing contrast on obstruents can only give rise to tonal contrast on vowels if the language already has tones.

H basically agrees with K (and the Neogrammarians) that phonetics is the 'blind' part of language change. He presumably also agrees that UG could act as a filter, turning a collection of random data into a coherent grammar. But H nevertheless contests (c) in the list on page 5.3, and also the priming effect. ((a-b) are neglected; presumably because H agrees with this in principle, even though he dismisses most candidates for those universals as 'extraphonological' (i.e. phonetic)).

The criticism against (c) is that it is unclear (i) where in the grammar a 'tendency' to open syllables (which is violated untill the very last stage of the change) would be represented, (ii) why a language learner would ever decide that such a tendency is operative in the initial stages of the change, where there are massive amounts of closed syllables. This is a very important point: if cases of drift exist, this is a very important problem for everybody, including K and H.

☞ Remember that some cases of drift may be reducible to sociolinguistic factors, as Labov has shown for the Northern Cities vowel shift.

The priming effect is contested by H on empirical grounds: he claims that there are counterexamples of changes which do not satisfy the requirements of this effect. Such arguments are never very strong, if they are not backed up by an analysis of the data which are supposed to be counterarguments. K's theory clearly is more restrictive than H's in this respect.

A more important criticism of H against priming is: how does the child know that ' the language's phonological representations have a class node to host' a feature, given that she is in the process of acquiring the language. Furthermore, it has to be noted that the priming effect could not explain why a language would ever develop tonal contrast. An additional hypothesis has to be invoked (e.g. language contact).

H's alternative hinges (in part, at least) on the figure on p. 363. Certain changes are implausible or even impossible because of the way diachrony (i.e. grammar transmission) works, not because of UG, but because of a separate theory of 'possible misunderstandings' Ñ in essence this is the theory of diachrony. This basically makes a thorough understanding of diachronical mechanisms a prerequisite for work on UG; I think it is very hard to say anything of interest about it given our present poor state of understanding diachrony.

Unfortunately, Hale also does not seem to have very much to say about restrictions on 'diachronically possible grammars', except that these restrictions should exist. In all, his own proposal therefore seems rather unrestricted.

5 Language Contact

Language contact situations are often seen as the driving force of language change and language variation. It is often assumed that in such situations a language may 'borrow' a construction or a grammatical rule from another language. Here we will look at one well-known case of a language contact situation (dialects of Greek spoken in Cappadocia, i.e./ 'Asia Minor'), and discover that matters are actually more complicated.[7]

The Cappadocian dialects of Greek display a pattern of vowel assimilation that looks superficially like the vowel harmony that is familiar from Turkish. In this talk, we discuss these patterns and show how these are not to be analysed as vowel harmony from the Turkic type, but rather as extensions of data patterns that also exist in other (Southern) Greek dialects. In particular, we argue that two bisyllabic domains can be identified, one at the beginning of the word and one at the end. 'Harmony' within these two domains obeys different principles. Consider the following examples from a number of Cappadocian dialects (in our examples 'Standard Greek' refers to the Greek standard language as it is currently spoken):

(66)

	Standard Greek form	Dialect form	Gloss	Dialect name
a	ónoma	ónama	'name'	Silli
b	koskin-ó	koskun-ó		Silli
c	evðomáða	ovdomája		Axo
d	é-θe-k-a	éθaka		Farasa
e	zerv-á	zavrá		Livisi

Most of the speakers of this dialect have now died. It is not completely clear whether this 'harmony' was still an fully active phonological process at the moment at which these data were recorded, or whether it reflects a diachronic process which had applied at an earlier stage. We assume that, even if the latter is the case, this change of underlying forms still needs an explanation in terms of phonological theory.

Vowel harmony processes are not as widespread in Greek dialectology as they are in Turkish.

(67)

	nom.sg.	*gen.sg.*	*nom.pl.*	*gen.pl.*
'rope'	ip	ipin	ipler	iplerin
'girl'	kız	kızın	kızlar	kızların
'face'	yüz	yüzün	yüzler	yüzlerin
'stamp'	pul	pulun	pullar	pulların
'hand'	el̂	el̂in	el̂ler	el̂lerin
'stalk'	sap	sapın	saplar	sapların

[7] This work has been carried out in cooperation with Anthi Revithiadou from the University of Rhodes.

At first sight, it may therefore seem plausible to assume that the Cappadocian forms have simply adopted the Turkish process and added it to their otherwise Greek phonology. This is indeed the standard view in contact linguistics at least since the work of Thomason & Kaufman (1988) (see also Winford, 2003, for an alternative view of the Asia Minor contact situation).

5.1 Properties of Cappadocian 'Vowel Harmony'

Comparison to Turkish

In spite of appearances, there are numerousdifferences between the Cappadocian pattern and Turkish Vowel Harmony (VH):

1. The Cappadocian pattern does not always involve features; in the usual case the whole vowel is copied:

(68)

	Standard Greek form	Dialect form	Gloss	Dialect name
a.	ánem-os	ánomos		Axo
	ðáskal-os	ðáskol-os		Farasa
b.	ónoma	ónama		Silli
	pandeleímon-as	pandeleímanas		Silli
c.	ektóte	ektéte		Axo
	fílak-s-e	fílekse		Axo
	erxó-maste	erúmeste		Axo

2. Sonority plays an important role, much more than in Turkish, where there is only an effect of vowel height on labial harmony:

(69)

	Standard Greek form	Dialect form	Gloss	Dialect name
a.	kamilafk-i	kamalafki		Axo
	kateváz-i	kataváz		Axo
	meɣaríz-o	maɣaríz-o		Axo
b.	velón-i	volón-i		Axo
	embrós	ombrós		Axo
	meθópor-o	moxópor-o		Axo
c.	lizmon-ó	zolmonó		Axo
	evðomáð-a	ovdomád-a		Axo
	fover-ós	fovor-ós		Axo
d.	miruð-já	murudjá		Axo
	pipér-i	pepér-i		Axo

3. It is not sensitive to morphological structure:

(70)

	Standard Greek form	Dialect form	Gloss	Dialect name
a.	*within the stem*			
	tésera	tésara		Farasa
	ékso	ókso		Ulaghatsh
	ónoma	ónama		Silli
	ektóte	ektéte		Axo
b.	*between stem-suffix*			
	petsét-a	petʃáta		Silli
	ánem-os	ánom-os		Axo
	fílak-s-e	fílekse		Axo

4. Stressed final vowels are *not* triggers; in this case, the 'default' sonority-driven harmonic process takes place:

(71)

	Standard Greek form	Dialect form	Gloss	Dialect name
a.	kerat-ás	tʃaratás		Farasa
b.	monax-ós	manaxós		Axo, Silli
	orfan-ós	arfanós		Livisi
	perpat-ó	parpató		Farasa
	aðelf-ós	áðarfós		Livisi
	elin-ik-ó	elen-ik-ó		Farasa
c.	kirek-í	kerekí		Axo
d.	alep-í	alapú		Livisi

Comparison to Southern Greek

The importance of sonority in the Cappadocian harmony process reminds us of a vowel copying pattern found in other (southern) dialects of Greek. The following examples show that Karpathos Greek has a vowel copying pattern in which two adjacent vowels are assimilated:

(72) *initial vowel assimilation in Karpathos Greek*

a.	orfan-ós	arfanós
	árotr-on	áratron
	kalo-póð-i	kalapói
	pano-fór-i	panafóri
b.	elafr-ís	alafrís
	ená-mis-i	anámisi
	eryá-t-is	argátis
	ðrepán-i	ðrapa$_n$i
c.	irakl-ís	araklís
	ipako-í	apakoí
d.	velón-i	volóni
	embrós	ombrós
	pepón-i	popóni
e.	igr-ós	ogrós
	siróp-i	sorópi
f.	stomúx-i	stumúxi
	korúp-i	kurúpi
g.	ésθi-ma	éstema
	éksi	ékse
i.	kukíð-i	kukúi
	e-vréx-umin	evréxumun

Vowel-assimilation obeys a sonority hierarchy. If two vowels are adjacent, the less sonorous one assimilates to the more sonorous one, according to the following hierarchy:

(73) a >o >u >e >i
(there are a few problems with respect to the ordering of /o/ and /u/ which we will ignore here)

This hierarchy is obviously more generally known in the phonology of Greek, since it also guides vowel deletion in hiatus context. There are, however, differences between Karpathos vowel copying and Cappadocian harmony:

1. In the first place, unlike Karpathos Greek, Cappadocian does not obey the sonority hierarchy, especifially at the end of the word:

(74) a. *Karpathos*

orfan-ós	arfanós
elafr-ís	alafrís
velón-i	volóni
igr-ós	ogrós

b. *Cappadocian* (Axo)

ektóte	ektéte	Axo
fílak-s-e	fílekse	Axo

2. In the second place, Karpathos vowel copying only seems to happen within a stem (or possibly within a bisyllabic morpheme). More specifically, assimilation in Karpathos seems to be limited to:
   a) the stem:
      ésθi-ma éstema
      an-ésqi-t-os anéstetos
   b) disyllabic suffixes:[8] e-vréx-umin evre$_x$umun
   c) except if the stem is monosyllabic:
      élk-os órkos
      érg-on órgon
      igr-ós ogrós

   Cappadocian, however, does not obey this restriction:
   a) between stem-suffix
      petsét-a petʃáta
      ánem-os ánomos
      perðik-ó-θir-a perðikóθara
   b) within a suffix
      erxó-maste erúmeste
   c) within a stem
      meta-káno matakáno
      moná-ðipl-os manáðiplos

5.2 Two domains of harmony

Theoretical background

In order to describe the pattern of Cappadocian we assume the following:

a. A harmonic span of two syllables is constructed at the end of the word and at the beginning of the word (for various implementations of the notion of harmonic span)
b. The two spans obey different requirements:
   - The span at the end of the word is more like Turkish vowel harmony . It concerns mainly spreading from roundedness and backness. There is no spreading from stressed vowels, or such spreading is very limited in this position.
   - The span at the beginning of the word is less restricted. It copies one vowel to the other vowel, along the lines of sonority à la Karpathos and other Southern Greek dialects (e.g. Symi, Rhodes, etc.).
c. Since the span at the end is more restricted, it takes precedence over the one at the beginning in the case of possible conflict.

In two syllable long words, the harmonic domains coincide:

[8] this could be due to the labial /m/.

(75)

	Standard Greek form	Dialect form	Gloss	Dialect name
a.	fáɣo	fóɣo		Ulaghatsh
b.	ékso	ókso		Ulaghatsh
c.	pu θá	paá		Livisi
d.	kíθe	kéxe		Axo
e.	ő̆oken	édeken		Ulaghatsh

☞ Examples such as fáɣo - fóɣo show that final-domain harmony takes precedence over initial-domain harmony.

In words which are longer than two syllables, harmony domains do not coincide. Provided there is a harmony-triggering vowel, namely a vowel from the set { a, o, e }, the domain is at the end of the word; otherwise, the harmony span is formed at the beginning of the word:

(76)

	Standard Greek form	Dialect form	Gloss	Dialect name
a.	tésera	tésara		Farasa
	petsét-a	petʃáta		Silli
	ónoma	ónama		Silli
b.	ánem-os	ánomos		Axo
c.	ektóte	ektéte		Axo
	fílak-s-e	fílekse		Axo
d.	kamilafk-i	kamalafki		Axo
e.	kateváz-i	kataváz		Axo
f.	megaríz-o	magarízo		Axo,

Stressed final vowels do not create a harmonic span:

(77)

	Standard Greek form	Dialect form	Gloss	Dialect name
a.	kerat-ás	tʃaratás		Farasa
b.	monax-ós	manaxós		Axo; Silli
c.	kirek-í	kerekí		Axo
	elin-ik-ó	elenikó		Farasa
d.	alep-ú	alapú	Livisi	

The reason for this presumably is that this would create a mismatch between the structure of the harmonic span and that of the metrical foot, which is a trochee:

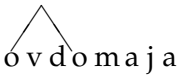
(78)

```
          ⇓                  mismatch
          (          )       harmonic span
    (           )             metrical foot
    m o n a x ó s
```

Formalisation

In this section we will attempt a formalisation of the chief insights presented in the previous section. First we assume that we need a notion of a harmonic span, consisting of two syllables (Halle & Vergnaud, 1978; Harris & Lindsey, 1995; Hulst & Weijer, 1995). In accordance with at least some of these authors, we assume that these spans are congruent with metrical feet, more specifically, trochees (McCarthy, 2004, presents a different approach).

(79) a. Initial domain



b. Final domain

Within these feet, different principles apply, as we have seen. We propose to formalise these using the notion of positional markedness (Kiparsky, 1997; Zoll, 1998; Smith, 2004, and others): certain markedness constraints hold only (or hold more forcefully) in prominent positions than in others. Prominence may be defined either in terms of stress, or of absolute position: word-initial positions are considered more prominent than others.

We propose the following positional markedness constraint is in effect at the beginning of the word

(80) HNUC/FIRSTFOOT: Syllable nuclei should be maximally sonorous (within the first foot of the word)

One way to make the nuclei within the first foot maximally sonorous, would be by simply turning them all into the most sonorous vowel, /a/, e.g.:

(81) elin-ik-ó >*alanikó

The reason why this does not happen, is that (80) interacts with a faithfulness constraint to the following effect:

(82) MAX-VFEAT: Do not insert (vocalic) features

(83)

/elin-ik-ó/	MAX-VFEAT	HNUC/FIRSTFOOT
a. elinikó		ei!
☞ b. elenikó		ee
c. elanikó	*!	e
d. alanikó		**!

This part of the system thus has nothing to do with harmony, from a purely formal point of view. Both vowels want to be as sonorous as possible, without adding new material. Spreading of the full vowel is the best way to get this effect.

The domain at the righthand edge of the word obeys a different type of positional markedness constraint. In this case, we propose a constraint which is more in conformity with the proposals of Walker (2004) for metaphony in Romance (specifically Italian) dialects, in which features seem to move to stressed (i.e. head) positions in the word. In order to analyses these, Walker uses constraints of the following type:

(84) LICENSE(F, S-Pos): Feature [F] is licensed by association to strong position S. Let:
   i. f be an occurrence of feature [F] in an output O (optional restrictions:
      (a) f is limited to a specification that is perceptually difficult,
      (b) f belongs to a prosodically weak position,
      (c) f occurs in a perceptually difficult feature combination),
   ii. s be a structural element (e.g. σ, μ, segment root) belonging to perceptually strong position S in O,
   iii. and sðf mean that s dominates f. Then (∀f) (∃s) [sðf] .

Simply put, LICENSE(F, S-Pos) requires that a feature be affiliated with a perceptually strong position. In the case of Cappadocian dialects, the relevant features F are (possibly) [round] and [back]. The S-Pos is the head of the domain-final harmonic foot:

(85) LICENSE([round, back], HeadHarmony): Features [round, back] are licensed by association to the head of a harmonic domain.

Because of the nature of this constraint, spreading will only go from a less prominent position to a more prominent position. This is the reason why we do not find forms such as **monoxós* from *monaxós*, spreading from a prominent (main stressed) position

This constraint also interacts with a faithfulness constraint, in this case of the following type:

(86) IDENT([round, back]): If an input segment A and an output segment B are in a correspondence relation, they should have the same specification for features [round, back].

The interaction between these two constraints gives us the required pattern:

(87)

/anemos/	LICENSE	IDENT
a. anemos	*	
☞b. anomos		*
c. onomos		**!

If we assume that LICENSE≫HNUC/FIRSTFOOT, we can also describe how the domain at the end of the word takes precedence over the domain at the beginning of the word.

Discussion

Our analysis of the data in the preceding section just scratches the surface of the complicated data we find in the Cappadocian dialects. Even though our generalisations made above seem to cover a large majority of data, it also is possible to find problematic cases, which do not conform to what we have said. We may see these forms either as lexical exceptions, as indications that more fine-grained analysis is necessary or as indications that other (diachronic) processes have interfered. In either case, we believe that the basis of our analysis will stand to scrutiny.

Some further issues:

- We could wonder why Cappadocian dialects have developed these intricate patterns of harmony. Even though we have shown that they do not really have a truly Turkic type of vowel harmony, it stands to reason that these patterns have still developed under the influence of language contact with Turkish. Possibly, this contact has brought Greek language learners to extend the patterns they already found in the (Southern) Greek of their parents so that they would look more like vowel harmony.
- Another question is why the 'Greek' pattern shows up at the beginning of the word, while the 'Turkish' pattern shows up at the end. Our guess is that the language learner will have more opportunity to observe the Turkish pattern at the beginning of the word. First, vowel harmony patterns in Turkish are most easily observed at the edge between stems and suffix (because this is where the real alternations are). Second, the end of the word is where the main stress usually is (in these dialects), so that these positions are more prominent. We speculate that adoption of something similar to the foreign language is more likely in these prominent positions than in non-prominent positions.

Bibliography

Antilla, Arto (1997a). 'Deriving variation from grammar'. In: Hinskens, Frans, Roeland van Hout, & Leo Wetzels (eds.), *Variation, Change and*

Phonological Theory, pp. 35–68. Amsterdam: John Benjamins.

——— (1997b). *Variation in Finnish Phonology and Morphology*. Ph.D. thesis, Stanford University.

——— (2002). 'Variation and Phonological Theory'. In: Jack Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, pp. 206–243. Oxford: Blackwell.

Boersma, Paul (1998). *Functional Phonology*. Ph.D. thesis, University of Amsterdam.

——— (2001). 'Review of Antilla (1997a)'. *GLOT International*, 5, 1: 33–40.

Boersma, Paul & Bruce Hayes (2001). 'Empirical tests of the gradual learning algorithm'. *Linguistic Inquiry*, 32, 1: 45–86.

Chomsky, Noam (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

——— (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Clements, Nick & Engin Sezer (1982). 'Vowel and Consonant Disharmony in Turkish'. In: van der Hulst, Harry & Norval Smith (eds.), *The Structure of Phonological Representations*, pp. 213–255. Dordrecht: Foris.

Coetzee, Andries W. (2004). *What it Means to be a Loser: Non-optimal Candidates in Optimality Theory*. Ph.D. thesis, University of Massachusetts at Amherst.

Dols, Nicolau (2000). *Teoria fonològica i sillabificació. El cas del català de Mallorca*. Ph.D. thesis, Universitat de les Illes Balears.

Féry, Caroline (1997). 'The Mora as a Measure of Weight and as a Syllabic Constituent'. In: Bertinetto, P.M., L. Gaeta, G. Jetchev, & D. Michaels (eds.), *Certamen Phonologicum III. Papers from the Third Cortona Phonology Meeting*, pp. 91–110. Torino: Rosenberg & Sellier.

——— (2001). 'Markedness, Faithfulness, Vowel Quality and Syllable Structure in French'. In: en Ruben van de Vijver, Hans-Martin Gärtner (ed.), *Phonology in Potsdam*, pp. 1–31. Potsdam: Potsdam University.

Goeman, Ton (1999). *T-deletie in Nederlandse Dialecten. Kwantitatieve analyse van structurele, ruimtelijke en temporele variatie*. Ph.D. thesis, Vrije Universiteit Amsterdam. URL `http://www.meertens.nl/uitgeverij/extern/tonpdf/download.html`.

Guy, Gregory R. (1997). 'Competence, Performance, and the Generative-Grammar of Variation'. In: Hinskens, Frans, Roeland van Hout, & Leo Wetzels (eds.), *Variation, Change and Phonological Theory*, pp. 125–143. Amsterdam: John Benjamins.

——— (2004). 'Dialect unity, dialect contrast: the role of variable constraints'. Talk presented at the Meertens Institute, August 2004.

Hale, Mark (2003). 'Neogrammarian Sound Change'. In: Joseph, Brian & Richard Janda (eds.), *The Handbook of Historical Linguistics*, pp. 344–368. Oxford: Blackwell.

Halle, Morris & Jean-Roger Vergnaud (1978). 'Metrical structures in phonology'. MIT.

Hammond, Michael (1997). 'Vowel Quantity and Syllabification in English'. *Language*, 73, 1: 1–13.

Harris, John & G. Lindsey (1995). 'The Elements of Phonological Representation'. In: Durand, Jacques & Francis Katamba (eds.), *Frontiers of Phonology*, pp. 34–79. London and New York: Longman.

Hulst, Harry van der & Jeroen van de Weijer (1995). 'Vowel harmony'. In: Goldsmith, John (ed.), *The Handbook of phonological theory*, pp. 495–534. Oxford: Basil Blackwell.

Keymeulen, Lut & Johan Taeldeman (1985). 'Tussen fonologie en morfologie. De vokaalverkorting in een Brabants dialekt'. *Taal & Tongval*, 37: 124–164.

Kiparsky, Paul (1997). 'The Rise of Positional Licensing'. In: van Kemenade, Ans & Nigel Vincent (eds.), *Parameters of Morphosyntactic Change*. Oxford: Oxford University Press.

——— (2003). 'The Phonological Basis of Sound Change'. In: Joseph, Brian & Richard Janda (eds.), *The Handbook of Historical Linguistics*, pp. 313–343. Oxford: Blackwell.

Labov, William (1987). 'Some Observations on the Foundation of Linguistics'. Available at `http://www.ling.upenn.edu/~wlabov/Papers/Foundations.html`.

——— (1993). *Principles of Language Change. I: Internal Factors*. Blackwell: Oxford.

——— (2001). *Principles of Language Change. II: Social Factors*. Blackwell: Oxford.

Lloret, Maria-Rosa (2004). 'The phonological role of paradigms: The case of insular Catalan'. Appeared as ROA 646-0404.

Mascaró, Joan (1983). 'Three Spanish Variations and a Majorcan Theme'. Ms., Universitat Autònoma de Barcelona.

McCarthy, John (2003a). 'Comparative Markedness'. *Theoretical Linguistics*, 29: 1–51.

——— (2003b). 'Optimal Paradigms'. Available on ROA 482-1201.

——— (2004). 'Headed Spans and Autosegmental Spreading'. ROA 685-0904.

Nagy, Naomi & Bill Reynolds (1997). 'Optimality theory and variable word-final deletion in Faetar'. *Language Variation and Change*, 9, 1: 37–56.

Nuyts, Johan (1989). 'Het Antwerpse vokaalsysteem: Een synchronische en diachronische schets'. *Taal en Tongval*, pp. 22–48.

van Oostendorp, Marc (1997). 'Style levels in conflict resolution'. In: Frans Hinskens, Roeland van Hout & Leo Wetzels (eds.), *Variation, Change and Phonological Theory*, pp. 207–229. Amsterdam: John Benjamins.

——— (2000a). *Phonological Projection*. Berlin/New York: Mouton de Gruyter.

——— (2000b). 'Syllabic Sonorants in Clay Frisian and Padgett's Generalisation'. *Philologia Frisica*, pp. 225–240.

——— (forthcoming). 'An Exception to Final Devoicing'. In: van der Torre, Erik Jan & Jeroen van de Weijer (eds.), *Voicing in Dutch*. Amsterdam: John Benjamins.

Padgett, Jaye (1994). 'Stricture and Nasal Place Assimilation'. *Natural Language and Linguistic Theory*, 12: 465–513.

Prince, Alan & Paul Smolensky (1993). 'Optimality Theory: Constraint Interaction in Generative Grammar'. Manuscript, Rutgers University and University of Colorado at Boulder.

Rosenthal, Sam (1997). 'The distribution of prevocalic vowels'. *Natural Language & Linguistic Theory*, 15: 139–79.

Selkirk, Lisa (1972). *The Phrase Phonology of English and French*. Ph.D. thesis, MIT.

Smith, Jennifer (2004). 'Making constraints positional: Toward a compositional model of CON'. *Lingua*, 114, 12: 1433–1464.

Steriade, Donca (2001). 'The Phonology of Perceptibility Effects: The P-map and its consequences for constraint organization'. Ms., UCLA.

Swets, Francine (2004). *The Phonological Word in Tilburg Dutch. Government Phonology and a City Dialect*. Ph.D. thesis, University of Amsterdam.

Taeldeman, Johan (1980). 'Inflectional aspects of adjectives in the dialects of Dutch-speaking Belgium'. In: et al., Wim Zonneveld (ed.), *Studies in Dutch Phonology*, pp. 265–292. Den Haag: Martinus Nijhoff. URL `http://www.dbnl.org/tekst/tael002infl01/`.

Thomason, Sarah Grey & Terence Kaufman (1988). *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.

Verner, Karl (1875). 'Eine Ausnahme der ersten Lautverschiebung'. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen*, 23, 2: 97–130.

Visser, Willem (1997). *The Syllable in Frisian*. Ph.D. thesis, Vrije Universiteit Amsterdam.

Walker, Rachel (2004). 'Weak Triggers in Vowel Harmony'. *Natural Language and Linguistic Theory*.

Weinreich, Uriel, William Labov, & Marvin Herzog (1968). 'Empirical Foundations for a Theory of Language Change'. In: Lehmann, W. P. & Yakov Malkiel (eds.), *Directions for historical linguistics. A symposium*, pp. 95–188. Austin: University of Texas Press.

Winford, Donald (2003). 'Contact-induced changes - Classification and processes'. In: Dawson, Hope C., Robin Dodsworth, Shelome Gooden, & Donald Winford (eds.), *OSU Working Papers in Linguistics 57*. Columbus, Ohio: The Ohio State University.

Zoll, Cheryl (1998). 'Positional asymmetry and licensing'. 'Expanded handout' of a talk presented at the LSA meeting. ROA 282-0998.