Tone and segmental structure

Marc van Oostendorp

August 28, 2006

Contents

1　Phonological models of tone

1.1　Autosegmental phonology

This course will be devoted to the phenomenon of lexical tone, and its relation to segmental structure. Although it will be assumed that students have a basic knowledge of phonological theory, in the first two classes we will set out to describe the basics of phonological theory.

　When I was an undergraduate student, the following example — which I found in Goldsmith (1990) — for me was a convincing reason to want to do phonology for the rest of my life.

　The facts are from Kikuyu, a Bantu language spoken in Kenya. The language is a tone language, but the way in which tones are distributed in the word looks rather messy at first. In order to see this at all, we first have to briefly consider the morphological structure of the Kikuyu word, which can be described by the following template:

(1)

Subject	(Object)	Root	Tense
to 'we'	mo 'him'	rɔr 'look at'	
ma 'they'	ma 'them'	tom 'send'	irɛ

Now if we combine these words and we look at the resulting patterns, it looks at first as if (almost) any morpheme can occur both with a low tone (marked à) or with a high tone (marked á):

(2)

	Subject 'to'	Subject 'ma'
rɔr	tò rɔ̀r ìrɛ́	má rɔ́r ìrɛ́
	tò mò rɔ̀r ìrɛ́	má mó rɔ̀r ìrɛ́
	tò mà rɔ́r ìrɛ́	má má rɔ́r ìrɛ́
tom	tò tòm írɛ́	má tóm írɛ́
	tò mò tòm írɛ́	má mó tòm írɛ́
	tò mà tóm írɛ́	má má tóm írɛ́

It is not exactly true that any morpheme shows any tone: when the subject marker 'to' always comes with a Low tone, while the subject marker 'ma'

always comes with a High tone. Furthermore, the morpheme immediately following the subject marker always bears exactly the same tone as the subject marker itself. In some sense, the subject thus determines the tone of the following morpheme.

Similarly, we may observe that the final tone of the tense suffix is always high, but that the first syllable has a varying tone: if the stem is rɔr, we find a low tone, if it is *tom*, it is a high tone. Thus it seems to be the stem which determines the first tone of the suffix.

Goldsmith proposes that we can best understand the Kikuyu tone system if we generalise these observations: the tone of every morpheme shows up on the following morpheme. Every morpheme in Kikuyu thus consists of two separate parts: segmental material on the one hand, and completely independent of that, a tone.

The underlying representations thus look as follows:

(3) | to | ma | mo | ma | rɔr | tom | irɛ |
     |----|----|----|----|-----|-----|-----|
     | L | H | L | H | L | H | H |

On the surface, every tone needs to be linked to some vowel, due to the so-called Association Convention:

(4)   Association Convention: No 'floating' tones are allowed on the surface, every tone needs to be linked to a vowel.

The Association Convention for tones is part of a more general set of requirements on phonological structure, requiring every element in a phonological representation to be linked to the other parts of the phonological structure.

In many languages, the tones would be linked to the vowel in their own morpheme, but in Kikuyu there apparently is a different requirement which is more important:

(5)   ALIGN-Tone: All tones want to be as close to the right edge of the word as possible, given other conditions of the language.

In many tone languages of the world, we see the effect of ALIGN-Tone: tones tend to move to the right ('spread').

When ALIGN-Tone would be on its own in the world, it would choose to have the following representation as the best one for *ma mo tom irɛ*:

```
        ma   mo   tom  irɛ
                        |
(6)  H    L    H    H
```

All tones are linked to the final syllable, and thus maximally aligned. This comes at the cost, however, of creating a very complex tonal configuration on

this final syllable, and apparently, this is not a price which Kikuyu is willing to pay. In particular, the relation between tones and vowels in this language is very transparantly one-to-one. In other words, the Association Convention above can be refined to the following:

(7) WELLFORMEDNESS CONDITION (WFC): Every tone in the output representation should be linked to exactly one vowel, and vice versa.

Given the absolute nature of the WFC in Kikuyu — it is not absolute in all languages, as we will see later — the best we can do to maximally satisfy ALIGN-Ton, is the following:

(8)
$$
\begin{array}{cccc}
\text{ma} & \text{mo} & \text{tom} & \text{ir}\varepsilon \\
& \diagup & \diagup & \diagup \mid \\
\text{H} & \text{L} & \text{H} & \text{H}
\end{array}
$$
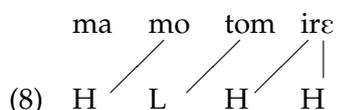
Every tone is now linked as much to the right as possible, without creating illicit 'contour' tones. Notice, however, that there is still one problem: the very first vowel (the one of the subject marker) does not bear a tone at all. There is no way we can solve this problem, given the requirements of the WFC, and some Bantu languages would leave it like this in similar situations, creating a toneless syllable.

However, notice that the WFC expresses several requirements at the same time, e.g. 'no tone should be linked to more than one vowel', and 'no vowel should be toneless'. Apparently, the former counts as a stronger violation in Kikuyu than the latter and therefore the following repair is made:

(9)
$$
\begin{array}{cccc}
\text{ma} & \text{mo} & \text{tom} & \text{ir}\varepsilon \\
\mid & \diagup & \diagup & \diagup \mid \\
\text{H} & \text{L} & \text{H} & \text{H}
\end{array}
$$
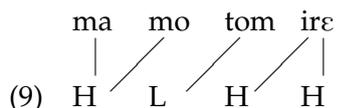
1.2 Contour tones

As we have seen, Kikuyu is very strict in its requirement that vowels can be linked to at most one tone. Another application of the idea of autosegmentalism which has proved to be very useful, is the analysis of so-called contour tones. For instance, the Chadic language Margi (Hoffmann, 1963; Williams, 1976; Kenstowicz, 1994) has three tones: a low tone, a high tone, or a rising tone. In principle, there are two ways of dealing with a situation such as this. We can either have a three way featural distinction (e.g. a feature Tone which has as values High, Low and Rising); or we can describe the rising tone as a combination of Low followed by High. Autosegmental analysis advises us to

take the latter route, so that we can minimize the number of primitives in our theory (there are only high and low tones, and autosegmental association):

$$
\begin{array}{ccc}
\text{a} & \text{a} & \text{a} \\
| & | & \diagdown \\
(10) \quad \text{H} & \text{L} & \text{L H}
\end{array}
$$
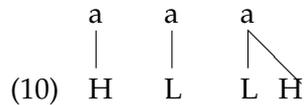
For Margi, the advice that autosegmental phonology gives us, turns out to be good advice. In the first place, this representation helps us to understand what is going on with tones. First, look at the following facts concerning the definite suffix *-ári*. The left-hand column represents the underlying shape of the stems to which this suffix is added (ˇa represents the rising tone):

(11) a. sál   sál-árì   'man'
          kùm   kùm-árì   'meat'

      b. ʔímí   ʔímy-árí   'water'
          kú   kw-árì   'goat'

      c. tì   ty-ǎrí   'morning'
          hù   hw-ǎrì   'grave'

(11a) shows that nothing happens if the suffix is attached to a consonant-final stem. Unlike in Kikuyu, every morpheme keeps its own home base; apparently the tone of the suffix is high.

(11b) shows that if the stem ends in a high vowel with a high tone, this turns into a glide. Since glides, like all consonants, cannot carry their own tone, it looks as if the high tone disappears.

(11c) shows that something does happen if the stem ends in a high vowel with a low tone. Again, the vowel turns into a glide, but now the tone of the suffix changes to a rising tone. Under autosegmental assumptions, it is very easy to understand this process: the rising tone is a combination of the original low tone of the stem and the high tone of the suffix:

$$
\begin{array}{cccc}
\text{t} & \text{i} & \text{a} & \text{r} & \text{i} \\
 & | & | & & | \\
(12) \quad \text{a. Input:} \quad \text{L} & \text{H} & & \text{H}
\end{array}
$$

$$
\begin{array}{cccc}
\text{t} & \text{y} & \text{a} & \text{r} & \text{i} \\
 & \diagup\!\diagdown & & | \\
\text{b. Output:} \quad \text{L} & \text{H} & & \text{H}
\end{array}
$$
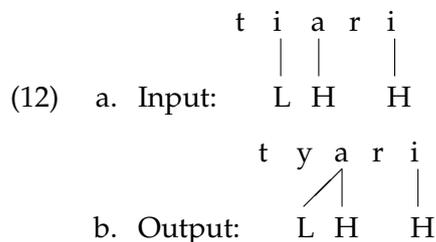
The reason why this happens can be seen as an interaction of the impossibility of the glide to carry the tone, and the wish of the tone to be linked to *some* vowel. Notice, by the way that this is always the vowel which is closest to

the tone in some intuitive sense. In particular, we will not find the following structure (the representation for *tyárĭ*):

(13)

$$t \quad y \quad a \quad r \quad i$$
$$\quad\quad\quad L \quad H \quad\quad H$$
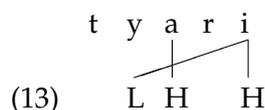
The reason why we do not find this, is because there is a very hard constraint on autosegmental representations of the following kind:

(14)   NoLineCrossing: Association lines may not cross

Different from all other constraints we have seen so far, NoLineCrossing is hard-wired into every known grammar: languages cannot fiddle with it. The reason for this presumably has to do with the interpretation of autosegmental representations. We are dealing in this case with two lines (traditionally called *tiers* in the theory): one line on which we have the tones, and another line on which we have our x-slots — in our example, we have given these x-slots the names of the sounds they carry, by way of abbreviation.

Each of those tiers represent in some sense a timeline: if element A stands before element B on a tier, this means that the pronunciation of A precedes the pronunciation of B. Thus, in (13), the realisation of the low tone will always precede that of the high tones.

If we think about our representations in this way, it stands to reason that association of an element X to an element Y means that the realisation of X overlaps with that of Y in time. Thus the pronunciation of the low tone in (12a) will happen during the pronunciation of the /i/.

But given all of this, (13) defies all logics: the low tone precedes the first high tone, but it is also realised during the pronunciation of an [i] which follows the [a] with which the low tone is associated. In other words, the pronunciation of the low tone will also follow the pronunciation of the high tone. This is logically impossible: α cannot at the same time precede and follow β (except if they overlap, but that is not the case here).

We can thus conclude that grammars can entertain all kinds of representations, including those which are not completely well-formed (because they display contour tones, or floating tones, or toneless vowels); but they will never entertain possibilities which do not make any sense at all.

Another remark to be made with respect to (12), is that this raises the question what is exactly the output representation for e.g. *kwárí*. We may assume that the high tone of the stem is deleted, but it is also logically possible to assume the following:

```
              k   u   a   r   i
                  |   |       |
(15)  a.  Input:  H   H       H

              k   w   a   r   i
                     ╱|       |
      b.  Output:  H   H       H
```

This would make the derivation of high vowel and low vowel stems exactly parallel. Whether or not we accept this, seems to be a matter of taste. Scholars who like the parallelism will readily accept this; others will point out that there is no empirical difference between a segment linked to two tones and one linked to one tone, and that we should therefore go for the simplest representation. The matter is hard to decide.

We quickly look at yet another argument in favour of the representation of rising tone as a sequence LH. We get this if we look at the underlying structure of stems in Margi. Bisyllabic stems in Margi come in three flavours: some of them have two low tones, some of them have two high tones, some of them have a low tone followed by a high tone. Monosyllabic stems similarly exist in three variants: some have a high tone, some a low tone, and some a rising tone. Under the autosegmental assumption, we can unify these by assuming that there are only three tonal templates in Margi: H, L, and LH:

(16)

	H	L	LH
bisyllabic	*ndábyá* 'touch'	*gə̀rhù* 'fear'	*pə̀zú* 'lay eggs'
	tə́dú 'fall down'	*dzà?ù* 'pound'	*ngùrsú* 'bend'
monosyllabic	*tsá* 'beat'	*dlà* 'fall'	*hǔ* 'grow up'
	sá 'go astray'	*ghà* 'reach'	*vǔl* 'fly'

Notice that this means that, even though Margi allows (rising) contour tones, it still only does this by way of a last resort: only because otherwise a tone would be lost (as in the gliding cases just discussed) or because it is the only way to express a tonal template. A bisyllabic word **gə̀rhǔ* is still not allowed, since it contains an 'unnecessary' rising tone. We thus cannot say that the wellformedness condition does not play a role at all; it just seems to be less stringent in Margi.

1.3 Place in the feature geometry

Although the ideas of autosegmental phonology have been extended to cover all other phonological features, it should be observed that autosegmental phonology seems most succesful with tones, which tend to behave as more independent from other phonological features than, say, [±coronal] or [±continuant].

Yip (2002, p. 40) offers a list of 'desiderata for a feature system for tone':

(17)   a. Characterize all and only the numbers of level tone contrasts (at least four, possibly five)

b. Characterize contour tones, and their relationship to level tones (rising, falling, convex, concave)

c. Characterize all and only the numbers of contour tone contrasts (two, possibly three, of a given shape)

d. Allow for simple statements of common tonal alternations (assimilation, dissimilation, contour formation and simplification, downstep)

e. Allow for simple statements about tonal markedness (in a two tone system, low is usually unmarked, in a three tone system, mid is usually unmarked)

f. Characterize the relationship between tonal and non-tonal features, particularly laryngeal features, both synchronically and historically (low tone associated with voicing, and especially breathiness, high tone associated with voicelessness)

The desiderata (17b) and (17d) are the job of autosegmental phonology, and to some extent we have already seen how this fulfills the job. The other items on the list will have to be taken care of by an appropriate theory of phonological primitives (for instance, features or elements) which are organized into (17f) is going to be the main focus of the rest of this course, but we will concentrate here on (17a), (17c) and (17e).

Let us start with (17a). An influential early analysis of four-way tonal contrasts is from Yip (1980), who uses two features to describe the tonal phonology of Chinese: [±Upper] and [±high]. The former is called a 'Register' feature and the latter a 'Tone' feature. Together they can define four levels of tone:

(18)

Register	Tone	descriptive label
[+Upper]	[+high]	extra-high
	[-high]	high
[-Upper]	[+high]	mid
	[-high]	low

Yip (1980)'s proposal was set in a so-called 'bottle-brush' model of autosegmental phonology: the two features should spread independently. On the one hand, this makes some desirable predictions: they certainly sometimes seem to spread independent from each others, indeed. On the other hand, they also sometimes spread together, which is not something that Yip's model could explain.

Let us start with the independent behaviour. In Ewe, mid vowels become extra-high when they are neighboured by high vowels (Clements, 1978;

Odden, 1995). From a purely phonetic point of view, this is quite unexpected: if these vowels just assimilate to their context, why would they not just become high? From the point of view of our feature theory, we can describe this property on the other hand quite elegantly as the assimilation of the Register feature, leaving the Tone feature untouched.

(19)

$$
\begin{array}{ccc}
+U & -U & +U \\
\sigma & \sigma & \sigma \\
| & | & | \\
-h & +h & -h
\end{array}
$$
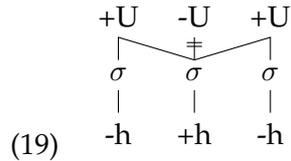
This analysis can count as a success for a formal model of tone which treats mid and extra-high tone as a natural class. Similarly, Chaozhou Chinese (Bao, 1990) has a three-level tone system. At first sight, mysterious alternations go on in the first parts of compounds:

(20) a. *hue* 'goods'
  b. i. hue   lung | 'cargo ship, freighter'
         HM   H |
     ii. hue   ts'ng | 'warehouse'
         ML   L |

(21) a. *hue* 'fire'
  b. i. hue   ba | 'torch'
         MH   HM |
     ii. hue   tsi | 'rocket'
         LM   LM |

Again, under autosegmentalist assumptions, these alternations are easily understood, if we assume the following:

- mid tones are represented as either [+U, -h] or as [-U, +h] (since this is a three-way height system).
- the word for 'goods' is underlyingly [+h, -h] and the word for 'fire' is [-h, +h]
- the register feature spreads from the second part of the compound to the first part

We thus conclude that Register can spread independent of Tone. The reverse is also true, and exemplified in yet another dialect of Chinese, Zhenzhai (Chen, 2000).

Again, the evidence comes from tonal sandhi in compounds. The word for 'room' (kɛ) is pronounced with HM in isolation, but if it occurs after a

low-register word such as 'sleep' (fã, with LM; the compound means 'bed-room'), the tonal contour turns into MH; it is as if the rising tonal contour has moved from 'sleep' to 'room', but the latter word has retained its original high Register. The contour *moves* rather than *spreads*: the first syllable shows up with a default contour:

$$
\begin{array}{cc}
\text{-U} & \text{+U} \\
| & | \\
\text{fã} & \text{kɛ} \\
\end{array}
$$

(22)   -h+h +h-h
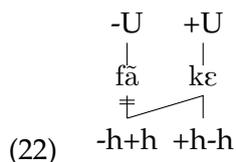
The problem with bottle-brush models such as (19), on the other hand, is that we do not explain why the two parts of the contour spread together. A feature geometric solution, in which the two tones are dominated by an extra node seems preferable:

$$
\begin{array}{cc}
\text{-U} & \text{+U} \\
| & | \\
\text{fã} & \text{kɛ} \\
& \\
\text{o} & \text{o} \\
\end{array}
$$

(23)   -h +h +h -h
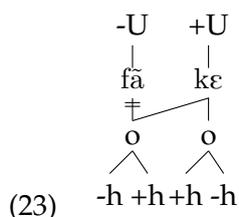
However, given the same line of reasoning, we should also device a common node for the Register and the Tone features, since they can sometimes also spread together. The diminutive suffix /tə(ʔ)/ of Changzhi does not have its own tones, but assumes a copy of the tone of the stem, regardless of its values for Register or Tone (Yip (2002); Duanmu (1994) ; tone is noted here according to the Chinese tradition, where '5' denotes the highest pitch and '1' the lowest):

(24)   a. tsə^{213} tə^{213} 'cart'
        b. paŋ^{535} tə^{535} 'board'
        c. xæ^{24} təʔ^{24} 'child'
        d. ɕiaŋ^{53} təʔ^{53} 'fillings'

For this reason, it has been suggested that the geometric structure of (contour) tone is as follows (Bao, 1990):

$$
\begin{array}{c}
\text{o} \\
\diagup \diagdown \\
\text{±U} \quad \text{o} \\
| \\
\text{±h}
\end{array}
$$

(25)
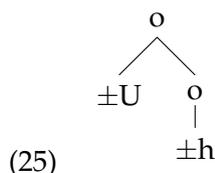
Since we can now determine the difference between a Register tone and a Tone tone by their position in the hierarchy, there is no longer any need to encode them by using different labels like Upper. We will here use the privative features H and L everywhere:

(26)

```
        o
       / \
    H, L   o
           |
          H, L
```
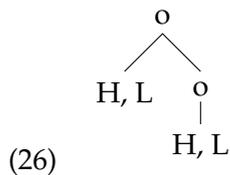
Finally, we have to give this tonal node its place in the feature hierarchy. Here, there are two options. On the one hand, there are reasons to believe that the tonal node leads a life independent from the rest of the segmental structure: tone is the most autosegmental of features, and it is sometimes assumed that it is linked directly to moras or syllables rather than internally to the root node of the segment, as in (**??**) (e.g. Piggott, 1992, for discussion). On the other hand, however, tone sometimes seem to interact directly with segmental features, most famously with laryngeal features, and therefore the tone features have sometimes been placed under the laryngeal node, as in (27b).

(27) a.

```
       μ
      / \
   Root  Tone
```

b.

```
         Root
         /  \
     Place  Laryngeal
                / \
         [±voice]  Tone
```
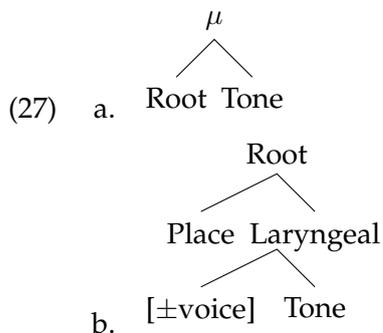
Notice that the arguments which are given for the two positions are of a different nature. The argument for (27b) is about interaction and mutual visibility; however, this is not a classical type of argument within a feature geometry approach. Mutual visibility is not an issue in such an approach, but relative (in)dependence of spreading phenomena is. For this reason, the stronger argument at this point seems to be in favour of (27a).

2   Grounding

2.1 'Beyond explanatory adequacy'

In the following classes, we will see that there are important phonological relations between tone on the one hand and consonantal voicing and vocalic length or syllable weight on the other hand. Now there also are phonetic relations between these different aspects of sounds of natural language. This raises the issue to what extent we unnecessarily duplicate our analytical tools if we adopt an 'abstract' account of these phenomena. This is going to be an important point of discussion in all of this course, in which we will also have to take into account that there are certain phenomena which are never or hardly ever phonologized; concretely, this will be the relation between vowel height and tone.

In order to get there, it is useful to study in some more detail the relation between phonetics and phonology in present-day phonology, and in particular in Optimality Theory.

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

(28)   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 (Clements & Keyser, 1983), 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 (Scheer, 2004), 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 (28). 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 Féry, 2003)

(29) 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 Chomsky (2004), 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 Kayne (1994) 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 Carstairs-McCarthy (1999).

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 evolutionary: 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 Hayes (1999), 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 & Reiss (2000); Hale (2003) and related works.

   More or less the same conclusion has been reached by authors such as Bybee (2001), 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 Bybee (2001) 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?

2.2   Synchrony and diachrony

The traditional point of view of the phonology-phonetics interface can be summarised as follows.[1] 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*. Bermúdez-Otero (2005) traces this idea back to Baudouin de Courtenay (1895) 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* (Hyman, 1976), 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,

---

[1]This section relies heavily on Bermúdez-Otero (2005).

leaving an idiosyncratic residue in lexical representations (*lexicalization*). [...]

As Bermúdez-Otero 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:

(30) Lexical representations
  (categorical)
  ↓
  Phonological rules
  ↓
  Phonological representations
  (categorical)
  ↓
  Phonetic rules
  ↓
  Phonetic representations
  (gradient)
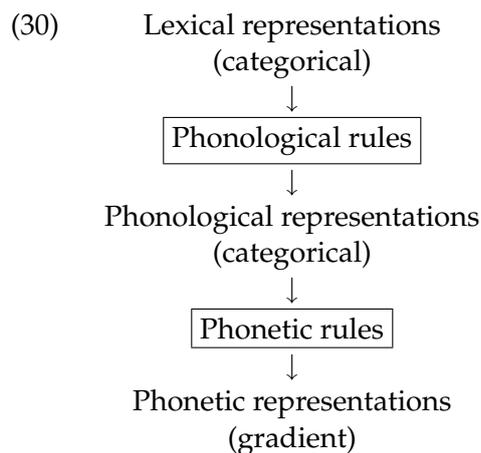
Under this view, then, sound change moves 'bottom-up' in the grammar: a change which originates in the phonetics may in the course of time end up having an effect only in certain lexical representations.

Although Bermúdez-Otero (2005) 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:

(31) C → / ___ # 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 (31) are

(much) more frequent than rules such as (32) is that the latter does not have a plausible phonetic origin; from a purely phonological point of view, there is nothing wrong about it.

(32)   C → / ___ # 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, Bermúdez-Otero (forthcoming) 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 Bybee (2001) 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.

3   Tone and sonority

3.1   Two groundings in phonetics

If contour tones are complex tones, i.e. if there representation is more complex than that of level tones, we expect their distribution to be more limited as well. This prediction turns out to be correct. In many languages, there is a limitation on the occurrence of contour tones. For instance, in Kiowa (Watkins, 1984; Gordon, 2001; Hyman, 1985) contour tones may only occur on syllables which are characterized by a long vowel or by a short vowel followed by a sonorant:

(33)   a.   kʰûːl 'pull off', hâː 'arise' REFL., kʰûːltɔ: 'pull off' FUTURE

b. k^hút 'pull off' PERFECTIVE, *k^hût, *k^hû

It can be shown that there is a typological hierarchy of the following shape:

(34)   CV:(C) ≻ CVR ≻ CVO ≻ CV

This should be read in the following way: C denotes consonants, V denotes vowels, R denotes sonorants and O denotes obstruents. The relation A≻B implies that if B is able to carry a tonal contrast in a language, so will A (A is a better tone carrier than B).

Individual languages allowing tonal contrast will thus usually have a cut-off point at some '≻' sign: those below it will allow for contour tones, those above it will only allow for level tones. Kiowa has the cut-off point between CVR and CVO; Shan (Morev, 1983) has a cut-off point between CVO and CV; etc. The same typology can also be found for pitch-accent languages, which will typically only allow contrast on syllables which are sufficiently 'heavy'. This weight is determined by the hierarchy in (34).

In this class, we will discuss two papers which both aim to show how this hierarchy is grounded in the phonetics Gordon (2001) and Zhang (2004). It is interesting to compare the two approaches, also since they are partly incompatible.

It is of course crucially important for both authors to be able to find a phonetic grounding. Gordon (2001, p.16) finds this in the following:

> The physical correlate of tone is fundamental frequency which is only present in voiced segments. In fact, the property which defines a voicing contrast is the fundamental frequency: voiceless segments lack a fundamental, voiced segments have one. Thus the only type of segment on which tone may be *directly* realized is a voiced one.

Notice that from this distinction alone, the scale in (34) does not follow: we can distinguish between CV:(C) and CVR on the one hand and CV on the other, but within the class of CVO, we should distinguish voiced from voiceless obstruents. However, rhymal obstruents hardly ever seem to be distinguished according to voicing in terms of weight.[2] At first sight, the predictions made by this phonetic criterion thus is simply wrong.

Gordon (2001) therefore refines his assumption by pointing out that the harmonics of a segment give good cues as to the value of the fundamental frequency. Harmonics occur at frequencies which are multiples of the fundamental: if the fundamental is at 400Hz, harmonics will occur at 800Hz, 1200Hz, 1600Hz, etc. In such a structure of harmonics, we could even take

---

[2]Interestingly, it has been argued recently that for onset obstruents, the voicing criterion is relevant (Topintzi, 2006) in the calculation of the attribution of weight. This proposal is couched in the (not uncontroversial) assumption that onsets may be 'moraic'.

away the fundamental from the signal; it would still be detectable because of the harmonics.

The key observation now is that vowels have a richer and more energetic harmonic structure than any consonants, and sonorants have stronger, more easily discernable harmonics than obstruents. Even voiced obstruents, although they have a fundamental have very weak harmonics:

> This fact, taken together with the inability of voiceless obstru-
> ents to carry tone, means that the class of obstruents considered
> as a whole is quite poorly suited to supporting tonal information.
> Gordon (2001, p. 17)

The following spectrograms display the differences in harmonic structure between the different types of consonantal segments:

(35)
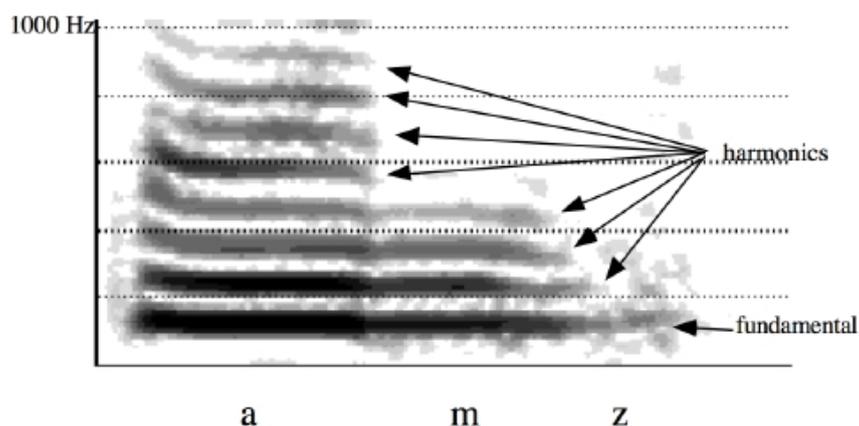
Gordon (2001) also provides a way to calculate the differences which are visible in this picture, viz. by adding up the intensities of the first five harmonics in the different types of segments in three languages, Cantonese, Navajo and Hausa (the first lacks obstruents and the second lacks voiced obstruents in coda, so that for these only the differences between vowels and sonorants could be measured). We then get the following table (sum of the first five harmonics in dB; standard deviations in parentheses):

(36)

Language	Vowel	Sonorant	Obstruent
Cantonese	255.8 (23.7)	146.0 (28.5)	
Hausa	246.5 (24.2)	219.0 (13.0)	162.3 (11.3)
Navajo	149.2 (27.2)	107.6 (21.9)	

The scale in (34) should now follow from this. Notice, however, that certain assumptions are underlying this line of reasoning. One is that apparently

the fundamental frequency itself apparently does not seem to matter: there is no language which distinguishes between CV[+voice] vs. CV[-voice] and there is no principled explanation for that. Although indeed the calculation of the first five harmonics gives us a scale which resembles the one we need in phonology, there is no principled *a priori* reason why we should choose these five harmonics. This (I believe) is a problem inherent in many 'functional' or 'phonetic' explanations of phonological phenomena; it will not be hard to find something in phonetic reality which *corresponds* to the phonological reality, but the issue is whether this correspondence can count as an explanation if the choice of the phonetic correspondent is completely free.

The goal of Zhang (2004) is partly different from that of Gordon (2001). He observes that contour tones are not only often restricted to long vowels, but also to stressed syllables (for instance in Xhosa; Lanham (1958)). The issue is:

> What is the link between these two contexts? Should they be accounted for by independent mechanisms, based on contrastive vowel length and stress respectively, or by some unified mechanism?

This is a rhetorical question of course, and Zhang (2004) aims to find this unifying context. Zhang (2004) is a little more precise as to the reason why we need to take into account certain harmonics and not others: phonetic studies (Plomp, 1967; Ritsma, 1967) have shown that the spectral region containing the second, third and fourth harmonics is especially important in the perception of fundamental frequencies in the range of speech sounds.

Notice, however, that this still does not really give us an explanation. Suppose that it was somehow given (let us say by UG, for lack of something better) that vowels and sonorants are good tone bearers, whereas obstruents are not. The listener would then be well-advised to pay close attention to those aspects of sounds which reflect this difference; the influence could therefore also be the other way around.

But next to sonority, another aspect plays an important role in Zhang (2004)'s explanation as well: duration, which is proposed to play a role both in articulation — it takes more time to produce a contour tone than a level tone; furthermore, "because the muscles responsible for pitch falls are both more numerous and more robust than those that execute a rise, it takes longer to implement a pitch rise than a pitch fall of the same extent" Zhang (2004, p. 158) — and in perception — "given the same pitch excursion, the longer the duration of the vowel, the more 'contour-like' the tone is perceived by the listener."

Based on these considerations, Zhang (2004) proposes the following formula for tonal complexity:

(37)   a. $C_{Contour} = a \cdot \text{Dur(V)} + \text{Dur(R)}$

b. Tonal complexity scale
    For any two tones T_1 and T_2, let C_1 and C_2 be the minimum $C_{Contour}$ values required for the production and perception of T_1 and T_2 respectively. T_1 is more tonally complex than T_2 iff $C_1 > C_2$.

Notice that the sonority difference has been built as a primitive into the definition of $C_{Contour}$. All in all, the definitions do not say much more than that *tonal complexity* depends on the duration of both vowel and sonorant consonant (and a little more on the former one, hence the constant a). Furthermore, Zhang (2004) assumes that contour tones have a larger tonal complexity than level tones, that drastic tonal changes (from extra high to extra low) are more complex than smaller changes, that HLH is complexer than HL and that rises are more complex than falls. All of these should follow from the phonetic observations just made.

From these primitives, Zhang (2004) derives:

- the relation with stress, since 'duration is often one of the key phonetic correlates of stress'; therefore, "it is reasonable to assume that all else being equal, a stressed syllable has a greater $C_{contour}$ value than an unstressed one."
- a relation with phrase-final position, since such positions are also often lengthened. "We thus expect that, all else being equal, a final syllable in a prosodic unit has a greater $C_{contour}$ value than a non-final syllable".
- the prediction that a syllable in a shorter word will have a greater $C_{contour}$ than an otherwise comparable syllable in a longer word, since words have a tendency to be of equal overall length.
- the segmental composition of the syllable rhyme: "all else being equal, VV has a greater $C_{contour}$ value than V; a VR [. . .] has a greater $C_{contour}$ value than VO [. . .]; and VV has a greater $C_{contour}$ value than VR, provided they have comparable duration."

Several things may be noted here. In the first place, the phrase 'all else being equal' recurs, but it is not very clear what exactly should make 'all else equal' in every case, if it would not be for phonology (e.g. faithfulness relations in Optimality Theory). Purely from the point of phonetics, there is no apparent reason, as far as I can see, why we could not lengthen the vowel in every case ad libitum just to create enough space to host the tone, especially in order to derive the segmental effect. (A possible explanation will probably invoke the notion 'contrast' in one way or another — a phonological notion.)

In connection to this, notice that neither Zhang (2004) nor Gordon (2001) is able to derive the difference between CV and CVO at this point (although we will see an attempt by Gordon below).

Finally, notice that it is implicit in both approaches that we consider syllable rhymes: the following consonant is supposedly tautosyllabic with the

preceding consonant. Yet the notion 'syllable' is notoriously hard to define in purely phonetic terms.

3.2 Turning phonetic constraints into a grammar

Gordon (2001) is satisfied with giving a phonetic background to the relevant typology, Zhang (2004) presents a more ambitious programme and aims to build a phonological grammar within the framework of Optimality Theory. He distinguishes two approaches to this:

- the *direct approach*, which assumes that positional licensing is contrast-specific. For instance, it is known that the first syllable of the word is a special licenser for many types of contrast, but this should not play a role in the case of tonal contrasts (it is a little unclear why the first position of the word licenses other contrast but not tone, but we will leave this aside).
- the *structure-only approach* which singles out certain phonological positions as special or strong; these positions should then license all kinds of contrasts.

One specific prediction of the direct approach is that language-specific implementation of certain positions may influence the contrast-bearing quality of those positions: if a language chooses to have only very minimal phrase-final lengthening, it may not be able to host contour tones there. Such a prediction is not possible in a structure-only approach.

Suppose that in a language syllables can have one of two distinct properties P_1 and P_2, and that the $C_{contour}$ value induced by P_1 is greater than that induced by P_2. In a structure-only approach we now predict that these two positions will correspond to two constraints, which will be freely rankable:

(38)   a.   *CONTOUR($\neg P_1$): no contour tone is allowed on syllables without property P_1 .

b.   *CONTOUR($\neg P_2$): no contour tone is allowed on syllables without property P_2 .

The crux is that these constraints should be freely rankable. If we also allow a 'disjoint' constraint *CONTOUR($\neg P_1$)∪*CONTOUR($\neg P_2$) (violated only when both are violated; hence satisfied when either one is satisfied, we get a six-way factorial typology — including a general *CONTOUR constraint, and a constraint on faithfulness:

(39)

Ranking	Type of language	
1.	*C(¬P$_1$), *C(¬P$_2$), *C≫FAITH	No contour on any syllable
2.	*C(¬P$_1$), *C(¬P$_2$) ≫FAITH≫*C	Contour only on σ with both P$_1$ and P$_2$
3.	*C(¬P$_1$) ≫FAITH≫*C, *C(¬P$_2$)	Contour only on σ with P$_1$
4.	*C(¬P$_2$) ≫FAITH≫*C, *C(¬P$_1$)	Contour only on σ with P$_2$
5.	*C(¬P$_1$)∪*C(¬P$_2$) ≫FAITH≫ *C, *C(¬P$_1$), *C(¬P$_2$)	Contour only on σ with either P$_1$ or P$_2$
6.	FAITH≫ *C, *C(¬P$_1$), *C(¬P$_2$), *C(¬P$_1$)∪*C(¬P$_2$)	No contours

On the other hand, within a direct approach, the constraints referring to P$_1$ and P$_2$ get a natural ranking order, due to the differences in phonetic weight between the two factors:

(40)   *C(¬P$_2$)≫*C(¬P$_1$)≫*C(¬(P$_1$&P$_2$))

This implies that the factorial typology is (somewhat) more restricted:

(41)

	Ranking	Type of language
1.	*C(¬P$_2$), *C(¬P$_1$), *C(¬(P$_1$&P$_2$)),*C≫Faith	No contour on any σ
2.	*C(¬P$_2$), *C(¬P$_1$), *C(¬(P$_1$&P$_2$))≫Faith≫*C	Contour tone only on σ with both P$_1$&P$_2$
3.	*C(¬P$_2$), *C(¬P$_1$)≫Faith≫*C(¬(P$_1$&P$_2$)),*C	Contour tone only on σ with P$_1$
4.	*C(¬P$_2$)≫Faith≫*C(¬P$_1$),*C(¬(P$_1$&P$_2$)),*C	Contour tone on σ with P$_1$ or P$_2$
5.	Faith≫*C(¬P$_2$), *C(¬P$_1$),*C(¬(P$_1$&P$_2$)),*C	Contour tone on all σ

This typology is slightly more restrictive, which Zhang (2004) considers to be an advantage in itself. Indeed, restrictiveness seems a desirable property of any theoretical account of some phenomenon, but note:

- that this restrictiveness should be balanced by the number of axioms in the theory; just ruling out some possibility by some blind stipulation does not make a theory more elegant or desirable. It remains to be shown that Zhang can develop a formula from which it really *follows* in every case that C(¬P$_1$) and C(¬P$_2$) are different in exactly the way that is required by the theory.
- that the lack of restrictiveness of the structure-only approach is partly due to the fact that constraint disjunction is allowed; were it not allowed, it would also only allow five possibilities, albeit five different ones from the direct approach. It may of course be the case that this prediction of the 'restrictive' structure-only approach is wrong — this could be demonstrated by showing analyses of languages in which only e.g. stressed syllables *or* long vowels can bear tone. Zhang does not provide such evidence, as far as I can see.

Zhang's implementation of the 'direct' approach relies on the following principles (p. 177-178):

1. *Canonicality*. It is assumed that there is a 'canonical' speech rate, and that the grammar makes calculations over this speech rate. Furthermore, there is:
2. *Normalisation*, so that the constraint rankings do not change at faster speech. The point of these two assumptions is that phonological systems tend to be rather stable over speech styles and circumstances, etc. There are no known cases of languages which divide up the scale in (34) in a different way in fast speech rather than in slow speech.
Notice that these two assumptions really fall outside of Zhang (2004)'s programme proper, and even seem to undermine the whole enterprise to some extent, since normalisation involves some type of abstraction: the 'direct' approach turns out to be less direct than is assumed.
3. *Contrast constraints*. Again, these constraints are needed to get some phonology into the system: not all of the (infinitely many) phonetic differences are possible. For instance, we will need to somehow get out of our system that there are no languages which phonologically contrast more than two phonological lengths. It is actually not clear how this particular effect can be obtained. Zhang (2004) refers to Flemming (1995); Kirchner (1997), but it is not clear (to me) how these could be applied to the problem at hand, without assuming that at some level of representation there is an object of which there are two in long vowels and one in short vowels (i.e. almost the definition of a mora). In any case, again this seems a weakening of the theory, introducing 'indirectness' through the back door.

We thus create a theoretical model in which there is no modularity between phonology and phonetics; these levels are heavily mixed. (A priori, this is neither a good thing nor a bad thing, but it is something to keep in mind.)

3.3   Two typologies

Gordon (2001)

Both Gordon (2001) and Zhang (2004) present extensive typological surveys to support their argumentation.

Gordon (2001) studied a sample of 105 languages which used tone or pitch accents contrastively; these were taken from a larger typologically balanced corpus of 399 languages. These 105 were exactly those languages within the relevant corpus which possessed the relevant property, although 12 did not allow any contour on a single syllable at all. Gordon (2001) distinguished five types of language:

(42)

	Type 1 (12)	Type 2 (25)	Type 3 (29)	Type 4 (3)	Type 5 (36)
CVV	No	Yes	Yes	Yes	Yes
CVR	No	No	Yes	Yes	Yes
CVO	No	No	No	Yes	Yes
CV	No	No	No	No	Yes

All these languages seem thus to conform with the scale in (34). Gordon (2001) analyses his data also from a purely typological point of view; although some language families show some bias in one direction or the other, nothing special can be noted however. The generalisation expressed by (34) thus seems to be a typologically valid one.

A few notes need to be made, however:

- 4 languages classified as Type 2, 5 languages classified as Type 3 and 11 languages classified as Type 5 lacked closed syllables. Furthermore, of the languages in Type 5, 3 do not have obstruent codas, and 11 lack syllables containing phonemic long vowels.

- For 5 languages in Type 5, "tonal restrictions could not be established with confidence based on available sources".

- No language has been found in which onsets count for the assignment of complex tones. According to Gordon (2001), this means that "the domain of tone is the syllable rime, parallel to many other prosodic phenomena such as weight-sensitive stress and weight-sensitive poetic metrics". Notice that this means, again, a retreat to a more abstract phonological account.

  By the way, Topintzi (2006), a dissertation dealing with 'moraic' behaviour of onsets, argues that although we would logically expect such onsets also to be functioning as TBUs. The reason for this is that moraic onsets preferably are voiceless obstruents — Topintzi (2006) assumes that moraic consonants are typically the most well-formed in their respective syllable position, so that moraic coda's are sonorants and moraic onsets are voiceless obstruents, which gives good predictions for e.g. the interaction with stress. But voiceless obstruents are of course very bad tone bearers, so that the optimal moraic consonants are bad tone bearers. A small minority of languages allowing moraic onsets also allows sonorants to be moraic; only those languages are therefore potentially of the right type, and Topintzi (2006) cites Kpelle as a potentially relevant case Hyman (1985); Welmers (1962) (noting however, and interestingly, that in this language also voiceless obstruents then can be argued to be tone bearers at least underlyingly).[3]

---

[3]Topintzi (2006)'s data are not directly of the type which is relevant for Gordon (2001),

The upshot of this argumentation is that we seem to be only be able to describe the facts properly if we refer to abstract syllabic positions such as onset and rhyme. If Topintzi (2006) is right, the moraicity of those positions is calculated independent from their potential tone-bearing status — exactly the opposite of what a 'direct' approach would predict. As far as I can see, such a direct interpretation suffers from severe problems interpreting the observation that onsets do not contribute to the tone-bearing quality.

- The three languages of Type 4 in (42) are the problematic cases, since they have a split between CVO and CV, whereas such a split should not be there from a phonetic point of view, if obstruents have nothing to contribute to the tone-bearing quality of a rhyme. These three problematic languages are Hausa, Luganda and Musey.
- One language seems to challenge (42) completely: Cantonese.

We will now follow Gordon (2001)'s points about each of these languages.

As a logical step in his argumentation, Gordon (2001) shows that independently established phonetic properties of the Type 4 cause their seemingly aberrant phonological behaviour. The argument is roughly the same in each case: it is not the obstruent itself which is bearing the tone, it is the fact that the vowel lengthens before such an obstruent but not at the end of a word which gives it this extra power to bear a contour tone.

Hausa (Newman, 1990; Wolff, 1993) has three types of syllable: CV, CVR, CVO and CVV. In an experiment, Gordon (2001) recorded a native speaker of the language pronouncing words including all those syllable patterns, with /a(ː)/ being the vowel in each case. In these words, he then measured the 'energy' of the first five formants. We will not go into the procedure in which he did this; here are the results (I copied this table as picture from Gordon (2001)'s manuscript, which makes it a little unsharp; standard deviations in parentheses):

	Syllable Type											
	CV		CVO			CVR			CVV			
	H	L	H	L	HL	H	L	HL	H	L	HL
Vowel	24.7	23.6	27.3	24.4	31.6	30.8	33.2	33.3	65.9	57.9	59.5
	(3.6)	(2.3)	(3.3)	(3.2)	(2.3)	(3.9)	(5.1)	(5.5)	(6.6)	(4.6)	(5.6)
Sonorous phase	24.7	23.6	27.3	24.4	31.6	65.4	65.4	57.6	65.9	57.9	59.5
	(3.6)	(2.3)	(3.3)	(3.2)	(2.3)	(2.2)	(6.3)	(7.1)	(6.6)	(4.6)	(5.6)

(43)

These data show that it is "not the coda obstruent in CVO which makes CVO a better licenser of contours than CV, but rather it is the *vowel*" (p. 27): vowels carrying (HL) contour tones in CVO syllables have a much greater energy

---

because the possibility of a contour is not dependent on the quality of an onset, as far as I can tell.

than vowels carrying level tones. This distinction in turn seemed to have been caused mainly by the relative length of the vowel. A similar distinction is not found in CVR syllables.

The obvious next question is why vowels have this special property in CVO, but not in CV. Gordon (2001) provides what he calls a 'functionally-driven' answer to this: Hausa has contrastive length in open syllables, but not in closed ones: CVVC syllables do not exist. Therefore lengthening can work in closed syllables without the danger of obscuring a phonemic contrast, whereas this is not true in open syllables. Like Zhang (2004), Gordon (2001) thus has to rely crucially on a notion of contrast: it has to be recorded somewhere (where?) that there is a contrast between CV and CVV, but not between CVC and CVVC, and the phonetic lengthening is sensitive to it.

This analysis seems supported by the facts of **Luganda** and **Musey**, the only two other languages distinguishing CVO and CV. In both languages, there is no length contrast in open syllables, but there is one in closed syllables.

For Luganda, a few details have to be made more precise. In the first place, phrase-final CVs in this language *can* carry a contour, presumably because they are lengthened (and phrase-final CVV is also lengthened, thus keeping the contrast). Notice that this means that in actual fact *all* syllable types in Luganda can carry the contrast phonologically, although they might not always be realised phonetically.

Furthermore, this language seems to distinguish between voiced and voiceless obstruents (Snoxall, 1967):

> Although CVO ending in a voiced obstruent and carrying a contour tone regularly realizes its contour tone, CVO ending in a voiceless obstruent and carrying a phonological falling tone, the only type of contour in Luganda, are reported to often carry only a "psychological low tone". Phonetically, the high part of the contour is realized on the CVO syllable while the low tone is characteristically phonetically realized through its lowering effect on the tone of the following syllable, though it should be pointed out that the vowel in CVO ending in a voiceless obstruent can realize the contour tone itself.

Gordon (2001)'s concluding discussion of the Hausa facts indeed is very much concerned with phonological issues. He argues that 'underlyingly' CVO syllables may be seen as equally long as CVV or CVR; hence their behaviour as phonologically tone-bearing. Since phonetically, on the other hand, they are not able to reflect their underlying tone, the vowel has to come at their rescue at this level.

The discussion of the difference between voiced and voiceless obstruents in Luganda seems to suggest the same thing: both 'underlyingly' can carry a tone, but only the ones which

It seems to me that this is also the only possible answer to a question which is not raised by Gordon (2001) as far as I can see: there are languages which neutralise length contrast in an open syllable, but not in a closed one (according to some analyses, German and Dutch should of this type); logically speaking there should be languages of this type having tonal contour as well (the Franconian dialects might be cases in point). Where then are there no languages in Gordon (2001)'s sample which have tonal contrast on CV but not in CVO? The phonological reasoning gives an answer; pure, direct phonetic grounding does not.

Cantonese seems to be a case more or less of this type, but it is more problematic: it allows contours on CV, CVR and CVVR, but not on CVO or CVVO. Again, Gordon (2001)'s (logical) response to this has been to carry out phonetic experiments to test the properties of the segments involved. A native speaker of Cantonese was asked to pronounce words with the rhimes /a, am, ap, aːm, aːp/. The energies were then measured in the same way as in previous experiments:

(44)

	CVO	CVVO	CV	CVR	CVVR
Vowel	20.5	41.2	69.9	25.7	51.2
Son. Phase	20.5	41.2	69.9	51.5	66.0

It turns out that there is far less energy on the syllable rhimes which disallow contours than in those which allow them. On closer inspection, it turns out that length is again the crucial factor: the sonorous part even of CV is longer than that of even CVVO (or CVVR for that matter).

Gordon (2001) argues that this length might be a function of a minimal requirement, demanding at least two moras in every prosodic word. It is not clear to me why CVVO should then not have approximately the same length as CV(ː) and hence be able to bear a tonal contrast as well.

Notice that this means that Cantonese is not a language of the type we argued to be missing above: the neutralisation at the end of the syllable does not go in the direction of CV, but if CVV.

It is also worth noting what explains what in Gordon (2001). In general, the line of reasoning seems to be that languages have a phonological mora structure and e.g. coda obstruents may or may not be allowed to be moraic. Phonetic implementation then may decide later whether or not the tonal contours can be realised on obstruents. Notice that this means that there still is some duplication in the analysis: phonology and phonetics seem to be making partly similar choices.

Zhang (2004)

Zhang (2004) also did a typological survey on a scale which is comparable to that of Gordon (2001): he studied 187 "genetically diverse tone languages with contour tones" (p. 166). These could then be divided in the following way:

- 22 have no restrictions on the distribution of countours (notice that this is very close to what Gordon (2001) found: 22/187=0.11 (Zhang); 12 / 105 =0.12 (Gordon (2001)).
- 159 have restrictions on contours that conform to the predictions of the direct approach (according to Zhang; to me it is not completely clear how this was tested, given the sensibility to language-specific phonetics of this approach)
- 5 languages have restrictions 'in both the expected and unexpected directions': Lealao Chinese, Margi, Zangcheng Chinese, Lao and Saek. These cases have been discussed in Zhang (2002), but not in the article under discussion. In stead of this, Zhang (2004) discusses a few cases which seem to be more problematic for the structure-only approach.

It is not very easy to identify those languages, since it involves knowing which of the relevant factors — vowel length, stress, final position, coda sonorancy — contribute how much to the phonetic profile of an individual potential tone-bearing unit, and this can be partially language-specific.

Here are a few examples of Zhang (2004)'s line of reasoning. In **Xhosa**, stress is on the penultimate syllable of the word, as is the only contour tone of the language (HL; we further find level H and L). Now notice the following result of Zhang (2004)'s measurements:


(45)
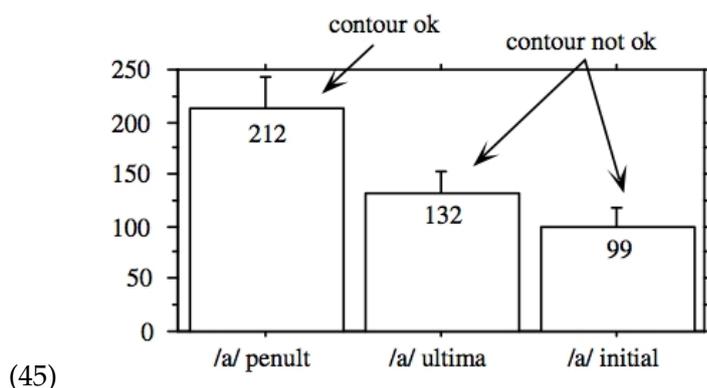
Both being in the final position and being in the stressed penultimate postion induces lengthening on the /a/; however, lengthening as a result of stress is much stronger than lengthening as a result of finality; it is this stronger effect which supports the tones.

It would actually be more interesting if a language could be found in which final lengthening was stronger than lengthening because of stress, and see whether Zhang (2004)'s predictions hold there as well. The problem with the Xhosa state of affairs is that it is rather straightforward to say that penultimate syllables are 'really' bimoraic, because of STRESSTOWEIGHT, whereas final syllables are 'only' phonetically lengthened, hence, not potential tone bearers.[4]

Similar measurements can be obtained for the interaction of vowel length and coda sonorancy. As a matter of fact, Zhang (2004) uses Gordon (2001)'s Cantonese data to argue for this point, as well as very similar data from Standard Thai:
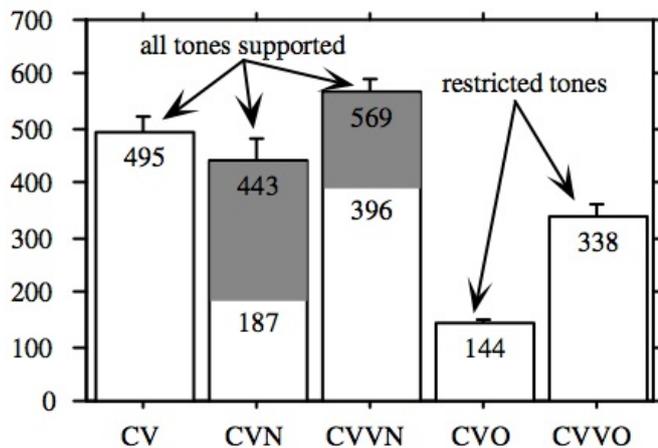
(46)

Like in Cantonese, the total proportion of length at the end of a syllable or before a sonorant (N) is much longer than before an obstruent; no explanation is provided as to why this is the case, but the fact is itself used as an 'explanation' for why tones are not possible in this case.

Navajo is "the opposite" of Thai and Cantonese, in the words of Zhang (2004): it has the full range of CV, CVR, CVO, CVV, CVVR and CVVO syllables allows tones on long vowels and not on short vowels, regardless of the quality of the coda. This corresponds again for a large part to a difference in relative length of the sonorant part (for some reason the Os are missing from the second and fifth column):

---

[4]Zhang (2004) cites Mende (Leben, 1973) as a case where "long vowels can carry LHL, LH, or HL in monosyllabic words, but only LH or HL in other positions. Short vowels can carry LH or HL in monosyllabic words, HL in the final position of disyllabic or polysyllabic words, but no contour in other positions." This is supposedly a problem for moraic theory, since it involves a four-way distinction, but it is not clear to me how this would be analysed in Zhang (2004)'s analysis, or whether moraic theory would necessarily only need to take mora's into account – and not e.g. the fact that words can have tonal template.

(47)
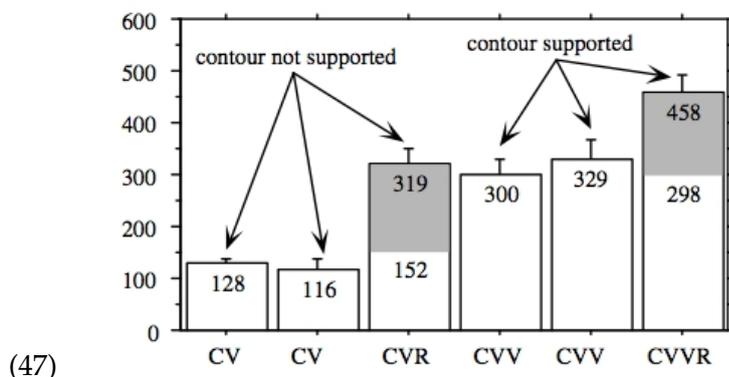
Notice however, that there is a problem with CVR syllables, which have a sonorous proportion which is at least as long as that of CVV. The solution is now to make a in the formula in (37a) language-dependent: some languages, like Navajo may choose to give greater weight to the duration of the vowel than others. Notice that this introduces an asymmetry in the system — why would certain languages not similarly choose to pay more attention to the sonorous part, which is after all the part that really has to carry the information of the contour, which is left unexplained.

In general, the level of explanation offered by Zhang (2004)'s approach does not seem very deep. We find phonetic length facts which correspond to a large degree with the occurrence or non-occurrence of tonal contours. But since nobidy will (presumably) deny that one needs phonetic length in order to express those contours in the first place, and since there is nothing in Zhang (2004)'s system which independently predicts the phonetic length, there seems some level of circularity involved. The explanation could be exactly the other way around: the phonetics module will always create enough space exactly on those phonological objects which need it, e.g. because of contour tones.

3.4   Tones on onsets

Let us briefly return to the issue of tones on onsets. As I have argued above, their absence pose a serious challenge to the phonetically inspired accounts: why should some sonorous part *preceding* the vowel not have the same influence of giving more space to potential contour tones (especially because intervocalic geminate sonorants can as well).

Next to the Kpelle case already cited above, Topintzi (2006) also discusses briefly another language which is of interest to us, Musey (Shryock, 1995). In this language we can distinguish between two types of consonants, 'Type A' and 'Type B':

(48)

Type A	sonorants and historically 'voiceless' obstruents	induce mid tone
Type B	historically 'voiced' obstruents	induce low tone

The tone inducing behaviour of these two types of tone can be observed on the basis of so-called tonal shift. Underlying tones on stem vowels can shift to the right if a suffix is added; the vowel will then itself show up with the tone which is 'induced' by the preceding consonant:

(49) Rightward displacement of lexical L tone in Musey
   a. Cliticisation of /na/
      i. Type A: sà → sanà → sānà 'person'
      ii. Type B: ɓù → ɓùnà 'goat'
   b. Subjunctive affix -m
      i. Type A: tò 'sweep' tōm̀ 'sweep it'
      ii. Type B: dò 'pick' dòm̀ 'pick it'

It thus looks, again, as if the onset consonants have some tone; importantly, however, this tone is predictable from the quality of the preceding consonant. This will bring us to the next topic in our course: the relation between voicing and tone.

4 Tone and voicing

4.1 Voicing and tone

It is well-known that voiced consonants have a phonetic lowering effect on (following) vowels, also in English (House & Fairbanks, 1953):

(50)
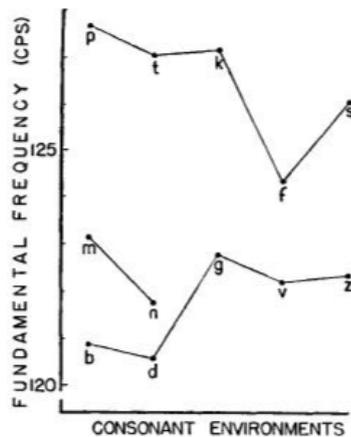
Theis relation between voicing and tone is also apparent from both the synchronic and the diachronic phonologies of many languages. Several authors have therefore suggested that we should use one and the same feature for voicing and tone (Halle & Stevens, 1971; Duanmu, 1990; Bradshaw, 1999; Kaye *et al.*, 1985; Ploch, 1999).

The classical case for the role of voicing in diachronic phonology is tonogenesis in Vietnamese (Haudricourt, 1954), where (putatively) a loss of voicing contrast on obstruents was accompanied with a gain in tonal contrast on the vowel:

(51)   a.   *pa > pa
       b.   *ba > pà

Synchronically, the effect is often known as that of *depressor consonants* (especially in the literature on African languages, but also elsewhere; Beach, 1924; Hombert, 1978).

For instance, in Suma (Bradshaw, 1999), imperfective tense is expressed by putting a high tone on the verbal stem, which is itself inherently toneless:

(52)   ɓúk   'applaud'
       éé   'leave behind'
       kírí   'look for'
       rɛ́m   'be able to'
       ndáŋgí   'boom'

However, if the stem starts with a voiced obstruent (not a prenasalised or implosive consonant), a Low tone is inserted before this High tone. This results in a rising tone if the stem is monosyllabic, and otherwise on a low tone on the first syllable and a high tone on the second:

(53)   bɔ̌m   'be blind'
       gǎy   'reprimand'
       vǎy   'bet'
       dɛ̀ɛ́   'swell'
       bùsí   'be bland'
       gɔ̀ɓí   'be bland'

Bradshaw (1999) adduces several reasons why the effect should be considered phonological in this case:

- In the first place, the difference between monosyllabic and polysyllabic words in (53) cannot be understood phonetically: if the effect were purely phonetic, we would expect e.g. the imperfective tense of 'be bland' to start with a rising tone ([bǔsí].

- In the second place, the effect is dependent on the morphology: we find this in verbs (which are inherently toneless) and not in nouns (which can be lexically specified for having an underlying tone):

  (54) a. bóná 'plant: Sporobolus pyramidales'
        b. dórò 'the trail of a small animal'
        c. vûy 'cord made from bark'
   d. vúlà 'acne, pustule'
        e. zîm 'taboo'

  Phonetics is supposedly blind to lexical categories or morphological structure, so that we do not expect this type of difference. (There presumably is some phonetic lowering in the nouns, which is smaller and not phonological.)

- A third argument comes from so-called monomoraic verbs in Suma (verbs which consist on one light syllable only). Since monomoraic syllables do not allow contours, such verbs show up only with the high tone. But as soon as another mora is added, for instance, in the nominalised forms of these verbs, the expected depressor tone surfaces again:

  (55)

Imperfective		Nominalized	
bɛ́	'refuse'	bɛ̀á	'refusing'
dɛ́	'do'	dɛ̀á	'doing'
ɗú	'collect'	ɗúá	'collecting'
fó	'agitate'	fóá	'agitation'

  Again, we would not expect a phonetic process to be sensitive to mora count, in Bradshaw (1999)'s view. (Notice that Zhang (2004) would presumably contest this point.)

Bradshaw (1999) analyses phenomena like these by assuming that voicing on obstruents and low tone on vowels are really one feature, which Bradshaw (1999) calls 'L/voice', but which we will simply call 'L' (in line with traditions in Dependency Phonology and Government Phonology). She assumes that this feature is 'multiplanar', i.e. it can be attached to two different planes: either to the Laryngeal node, in which case it will be realised as voicing, or to a mora, in which case it will be realised with a low tone:

$$\begin{array}{c} \mu \\ \text{Root} \quad\quad / \\ | \quad\quad\quad / \\ \text{Laryngeal} \quad / \\ \diagdown \diagup \\ (56) \quad\quad\quad L \end{array}$$
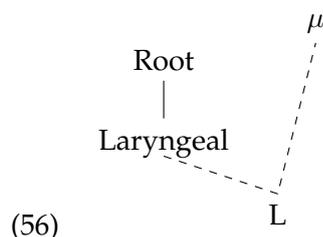
What happens in Suma is spreading: the feature L spreads from the voiced 'depressor' consonants to the following mora. However, it can only do so under the condition that this mora is not then forced to bear two tones — which may be a universal restriction on moras — but can hand over any other tones (such as floating inflectional tone) to other syllables. Lexically specified tones (such as in nouns) cannot be moved, due to faithfulness, and inflectional tones also cannot move if there is only one mora available. This explains exactly all the relevant contexts.

Another well-known effect of depressor consonants in African languages is that they trigger *downstep*: on the surface we hear two high tones, but the second high tone is significantly lower than the first one. Downstep is usually analysed in the literature as the effect of a floating low tone. Usually, downstep is transcribed with an exclamation mark '!' between the two high tones.

Bradshaw (1999) provides an example for downstep from the Bantu language **Makaa**, spoken in Cameroun. In this language, we find downstep in the nonpossessive associative construction (Heath, 1991), viz. after a stem-final voiced obstruent which is in an onset of the associative marker (AM):

(57) a. ò-káámbúg +AM+mə̀-bágə̀→ ò-káámbúg-!ó+mə̀-bágə̀
          'ants of the ashes'

We find this effect only if the associative marker has itself an underlying high tone. If it is unmarked for tone (as is the case for some of these interesting morphemes), a low tone shows up in these same conditions:

(58) a. káámbúg +AM+cúdú → káámbúg ù cúdú
          'ant of the animal'

It thus seems reasonable to assume that the voiced obstruent produces a L tone in both cases. Causing downstep thus is not very different from directly spreading a low tone.

A third process identified by Bradshaw (1999) concerns the spreading of L tone. For instance, in **Nupe** (George, 1970) the low tone of a nominal prefix spreads to the root vowel, *except* if the intervening consonant is a voiceless obstruent:

(59)   /è-tú/ → [ètú]   'parasite'
        /è-dú/ → [èdǔ]   'taxes'
        /è-lé/ → [èlě]   'parasite'

There are also a few processes where voicing seems to interact with High tone. One of these involves the blocking of docking of H tones. In Gbaya ɓokota (Bradshaw, 1999) there is a floating H tone which functions as the associative marker. When the second noun in the construction starts with a Low tone underlyingly, the associative High tone will dock onto it (60a); however, if the second word starts with a voiced obstruent and a low tone, this docking is blocked:

(60)   a.   wɔ̀+ mbòrō → wɔ̀ mbórō   'hunger for monkey'
              wɔ̀+ tòrō → wɔ̀+ tórō   'hunger for monkey'
              wɔ̀+ rì → wɔ̀ rí   'thirst for water'
              gbàà+ mbòŋgò → gbàā+ mbòŋgò   'corn seed'

        b.   wɔ̀+ gòrē → wɔ̀+ gòrē   'hunger for chicken'
              wɔ̀+ dùwà → wɔ̀+ dùwà   'hunger for goat'
              wɔ̀+ zàwà → wɔ̀ zàwà   'hunger for peanuts'
              gbàà+ zàwà → gbàā+ zàwà   'peanut seed'

This process can be understood if we assume that a constraint which wants the high tone of the associative marker to be realised is in competition with a constraint on the onset and the nucleus of a syllable sharing their L marking.

A final process we want to mention is 'voice insertion' in the Oceanic Austronesian language Yabem (Bradshaw, 1979; Ross, 1993; Bradshaw, 1999; Hansson, 2004). The last two syllables of the word (the last iambic foot) either both have a high tone or both have a low tone. Furthermore, all stops are either voiced or voiceless, ánd there is a correspondence between the tones and the voicing in the expected fashion: low tones go with voicing, high tones go with voicelessness. This is statically true for roots, and it causes alternations on prefixes:

(61)   -kátóŋ   'make a heap'   -gàbwà   'untie'
        tékwá   'his bone'   -gwàdè   'his cousin'
        típ   'all at once'   dìb   'thud'
        -kátóŋ   'make a heap'   -gàbwà   'untie'
        kó-péŋ   'shove (2 SG REAL)'   -gò-dèŋ   'move to (2 SG REAL)'
        tá-péŋ   'shove (1 PL. INCL)'   dà-dèŋ   'move to (1 PL. INCL)'

An important observation is that the key property is the tone: the voicing of the consonants can be predicted from the tones of the syllables; the opposite is not always true, since certain consonants are not contrastive for tone: the sonorants and the fricative *s*:

(62)   a. ká-sóm 'I speak'
        b. gà-sùŋ 'I shove'

Translating Bradshaw (1999)'s approach once again in terms of constraints, there thus is a requirement that all segments in the last foot of the word share L (or do not have this feature at all).

4.2   The representation of voice/tone

Bradshaw (1999)'s representation in (56) harbours two important assumptions:

- Low tone and [voice] are the same object
- They can be attached to different parts of the representation.

Together, these assumptions have to express the dual nature of Tone. On the one hand, it behaves really suprasegmentally, usually ignoring the mess that is segmental phonology (we could argue that the issue of sonority which we discussed before is itself already a little bit above this mess). At the same time, Low tone does not ignore one very small part of this mess, viz. voicing. High tone and Mid tone, on the other hand, are supposed to be purely tonal.

Unfortunately, no formal reason is given why this would be the case; this remains an arbitrary stipulation, motivated only by the empirical facts under discussion. It has to be noted in this light that there are a few more principled proposals in this light (e.g. Kaye *et al.* (1985); Ploch (1999)), which equate H with aspiration or [spread glottis]; M does not have a separate status in a theory involving register (section 1.3). However, Bradshaw (1999) argues that this predicts languages in which e.g. past tense would be realized as a Low tone, except if the verbal stem would start with a voiceless obstruent, in which case it would be realised with a falling tone (the inverse case of Suma in (52)).

In light of this discussion about feature geometry, we could wonder exactly which feature we are dealing with in this case, but unfortunately, the evidence about this seems inconclusive at this point: most of the languages with depressor consonant behaviour have only two different tones.

It is interesting to see what an alternative analysis would be. There are several options, but here we will consider one in terms of "grounded" phonology, viz. one in which we keep a representational distinction, but our phonological (Peng, 1992) or constraints (Hansson, 2004) are directly motivated by the phonetics.

Bradshaw (1999) briefly discusses the approach by Peng (1992), who assumes the following syllable structure for a closed syllable:

(63)

$$
\begin{array}{c}
\sigma \\
\diagup\diagdown \\
\mu \quad \mu \\
\diagup\diagdown \quad | \\
c \;\; v \quad c
\end{array}
$$
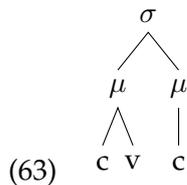
Furthermore, Peng (1992) uses the notion *path* from Grounded Phonology (Archangeli & Pulleyblank, 1994):

(64) There is a *path* between α and β iff:

  a. α and β belong to a linked set Σ of nodes/features, and
  b. in the set Σ, there is no more than one instance of each node/feature.

According to this definition, the onset consonant is in the same path as the vowel, whereas the coda consonant is not: in order to get from coda to vowel, one needs to go through two moras, which is disallowed by (64b).

Peng (1992) proposes that the laryngeal nodes of obstruents and the tone on the mora in a path should be subject to certain 'grounded' constraints. Bradshaw (1999) criticizes this assumption:

> This is where stipulation enters into this model. [. . .] What stops us from stipulating that [voice] and H must cooccur or that [voice] and L must not cooccur? It is only the added stipulation that path conditions must be phonetically motivated.

Everything which is said here is strictly speaking true. There is no phonological reason why we should stick to phonetically grounded paths. Furthermore, as we will see later in this course, not every phonetically grounded relation seems to lead to a phonological constraint. This is the big problem of 'grounded' phonology: which aspects of phonetics are available for grounding and which are not?

On the other hand the suggestion that Bradshaw (1999)'s model fares much better in this respect, is not justified:

- In the first place, we have just seen that the fact that L and [voice] are one feature, occurring on different tiers is a stipulation as well; it does not follow from any inherent property of these features, since [spread glottis] or H (or for that matter any other type of feature) do not display a similar behaviour.
- In the second place, Bradshaw (1999) herself is not able to derive the effects of the path condition (that onset consonants are relevant whereas coda consonants are not) in any way.

  Notice, by the way, that there is something unconfortable about the path condition also for grounded phonology: the fact that onset consonants are relevant but not coda consonants may itself be grounded,

since voicing tends to have a more easily discernible effect on following vowels than on preceding ones.

If we could combine the two approaches of Bradshaw (1999) and Peng (1992) we might actually get at interesting results. Assuming with both of these authors that moras are tone-bearing units, we might want to say that a consonant under a L mora is interpreted as voiced. This would keep the explanation for the connection between the onset and the tone of the syllable.

Hansson (2004) presents a constraint-based version of an idea similar to Peng (1992)'s in order to account for the Yabem facts in (61) on p. 37, and he briefly compares this to Bradshaw (1999); his main point is that his own analysis relies on "constraint interaction rather than representational assumptions." This remark is a little bit strange, because in his actual analysis, Hansson (2004) *does* invoke all sorts of representational assumptions, such as the idea that [+voice] and L are *distinct* features. (This is an example of a tendency with authors to believe that there own representational assumptions, especially when they are shared by more people, are *not* representational assumptions at all.)

Further, he notes that Bradshaw (1999)'s account is "very brief and sketchy" and he gives a fuller analysis, which we will discuss only briefly and sketchily: we are only interested in the interaction between voicing and tone. These follow from the interaction of the following three constraints:

(65) a. L ⊃ [voice] ("Low tone implies voicing"): For each syllable σ, *if* σ carries Low tone,, *then* every segment within σ must be [+voiced].

   b. *VOIOBS: A segment must not be simultaneously [-sonorant] and [+voiced].

   c. FAITH[voice]: No changes in the specification of the feature [voice].

They function in our tableaux in the following way (we give tableaux with potential inputs with the 'wrong' voicing specification for the consonants:

(66) a.

/tìp/	L ⊃ [voice]	*VOIOBS	FAITH[voice]
tìp	*!*		
dìp	*!	*	*
☞dìb		**	**

   b.

/-dáŋ/	L ⊃ [voice]	*VOIOBS	FAITH[voice]
dáŋ		*!	
☞táŋ			*

The fricative /s/ is always voiceless, also in syllables with a low tone. Hansson (2004) argues that the reason for this is that there is a high-ranking constraint forbidding voiced fricatives.

Hansson (2004) further argues that the constraint L ⊃ [voice] is not inherently directional; it does not tell us which of the two features (L or [voice]) will have to change. In many cases (like most of those we have seen above) it looks like it is the tone which is adapting to the voicing of the consonants; in Yabem it happens to be the other way around.

Discussing the differences between his own approach and the one by Bradshaw (1999) once more near the end of his paper, Hansson (2004) observes:

> The constraint-based analysis presented in this paper can be contrasted with the representation-based analysis outlined by Bradshaw (1999). The latter captures the interaction of tone and voicing by stipulating complete identity of the two in structural - representational terms: a single featural element [L/voice] carried by consonants and moras alike. In the above analysis, by contrast, the interaction is captured directly as such as an implicational relation embodied in a ranked and violable constrain. This obviates the need to define any representational connection between tone and voicing in featural or feature-geometric terms. What makes this possible is the shift from a process-based to a "target"-based perspective on sound patterns, inherent in constraint-based frameworks like OT. The conditioning of obstruent voicing by tone is viewed not as an operation spreading some autosegment from a vowel (or its mora) to a neighboring consonant, but simply as the satisfaction of a static target configuration imposed on output syllables (essentially: "both L and [+voi], otherwise H").

I can once again only see this as rhetorics, plus maybe some level of misunderstanding, viz. based on the assumption that separate features L and [voice] are in any sense more 'real' or more 'direct' than a feature which captures both manifestations at the same time, and that we should do away with such an assumption as soon as we can. However, authors such as Halle & Stevens (1971); Bao (1990); Duanmu (1990); Bradshaw (1999) have at least tried to also give a uniform phonetic characterisation of the single feature.

Hansson (2004) following observations seem to be more to the point:

> This highlights a fundamental characteristic of the constraint-based perspective on phonological feature interaction. There are in principle no inherent formal limitations against the interaction of features residing on distinct and unrelated representational tiers. Feature interaction is instead dependent on being explicitly encoded in some well-formedness constraint (in this case L⊃[voi]) in the form of an implicational relation. The burden of explaining interaction — what features interact and why —- is shifted to the domain where it properly belongs: the grounding of individ-

ual constraints (synchronically and/or diachronically) in factors
of speech production and perception. This is arguably superior to
attempts at artificially reifying the phonetic motivation for inter-
action through a representational-geometric architecture which,
no matter how finely articulated, is at best a very crude reflection
of the potential richness of phonetic realities (see Ohala, 1995).

The valid point of duplication is raised here: phonology does not have to
explain everything which is explained already by the phonetics. However, it
should also be recognized that in this particular case the proposal was not to
enrich phonological theory, but to reduce it. Also, it is not a "fundamental
characteristic" of the constraint-based perspective that we relegate explan-
tions to the phonetics; it is not the case that factors of speech production and
perception are the domains "where [the explanation] properly belongs; and
the 'crudeness' of phonological representations compared to the 'potential
richness' of phonetic realities can also be counted as a merit, since it restricts
the number of possible options. If Bradshaw (1999) is right, for instance, that
high tone does not interact with voicelessness in the same way that low tone
interacts with voicing, this is surely a problem for any grounding theory.

Bradshaw (1999) points out that voiceless obstruents from a purely pho-
netic point of view might be expected to have an even larger eect than voiced
obstruents, among other things because the rising tone which occurs after
voiced obstruents is more marked, both from a perceptual and from an ar-
ticulatory point of view. Rising tones are therefore also more marked typo-
logically, so we would expect the (falling tone inducing) effect of voiceless
consonants to be at least as big as the (rising tone inducing) effect of voiced
consonants.

Another problem for the purely phonetic view is that sonorants are quite
similar in their phonetic influence on following vowels, as the following
charts (from Hombert, 1978) show.

(67)
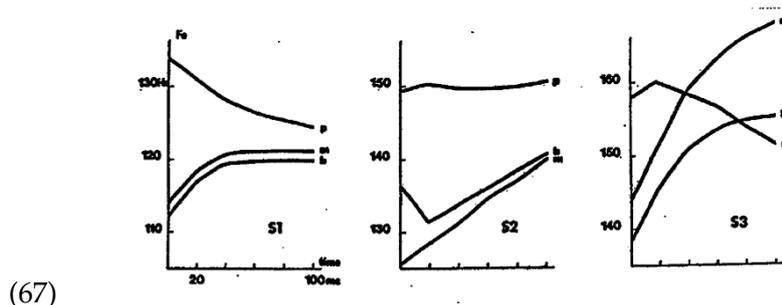
Yet sonorants do not participate in the phonological alternations, presumably
because they are not phonologically voiced.

This is not to say that there are no problems for a representational account
such as the one offered by Bradshaw (1999), but I believe that they are of

a different nature. For instance, tone and voicing do not always interact: there are languages in which tones can spread freely without considering intervening obstruents, and it is not immediately clear how we can represent this in a framework such as the one presented by Bradshaw (1999).

4.3 Why voicing-tone relations are relatively rare

Moreton (2006) observes that consonant-tone-relations (CTP) are relatively rare, compared to tone-tone interacctions (TTP) — phonological processes relating the tone of one syllable to that in the next.

In a non-systematic but still fairly broad[5] typological overview of (living) tone languages which had a voicing, aspiration or fortis-lenis contrast in obstruents — so that both TTP and CTP where logically possible, Moreton (2006) looked at static phonotactic patterns and morphophonemic alternations which involved either CTP and TTP (allophonic patterns were excluded, since they might be argued to be phonetic). The results were that TTP were much more frequent than CTP (we give a few representative examples; see Moreton (2006) for the complete survey):

(68)   a.   Tone-to-tone processes: 19 Ethnologue families, 5 continents

       i.   **Africa**: In *Tsonga (Niger-Congo)*, when an H-toned prefix is added to a word with only L tones, all tones but the last become H.

      ii.   **Eurasia**: In *Lhasa Tibetan (Sino-Tibetan)*, H-tone spreads in compounds, neutralising the tone contrast on the second member.

     iii.   **Central and NorthAmerica**: In *Dakhel (Na-Dene)*, disyllabic nouns can have LH, HL or HH, but not *LL.

     iv.   **South America**: In *Barasana (Tukanoan)*, an HL root or suffix suppresses H tone on a following suffix.

      v.   **Oceania**: In *Skou (Sko)*, the second element of compounds determines the tone of the whole compound, unless it is L.

     b.   Tone-voice and tone-aspiration interactions: 8 Ethnologue families, 4 continents

       i.   **Africa**: In *Lamang (Afro-Asiatic)*, syllables beginning with voiced obstruents have L tones; other syllables contrast L and H.

---

[5]"Cases of TTP and CTP were located by searchin (1) the collection of language-description books held by the University of North Carolina at Chapel Hill and written in Western European languages, (2) print and on-line journals focused on language description, such as *Oceanic Linguistics*, (3) general works on tonal phonology such as Bradshaw (1999), and (4) the World Wide Web, using to search on the string consisting of *tones* plus the name of each language family listed in Ethnologue (Gordon, 2005)."

ii. **Eurasia**: In *Mulao (Tai-Kadai)*, aspirated initial stops occur only with lower tones, while unaspirated ones occur with all tones (notice that this is a rather problematic type of interaction; we will not go into this).

iii. **North America**: In *Kiowa (Kiowa-Tanoan)*, medial voiced stops become voiceless after falling tone.

iv. **Oceania**: In *Skou (Sko)*, the H/L contrast is neutralized to phonetic mid tone in syllables with voiced-obstruent onsets; there is a voicing contrast before falling tone.

TTP thus seems much more frequent than CTP. Moreton (2006) notes that this finding is surprising in light of the fact that phonologization can create TTP only in tone languages, whereas CTP can also arise in non-tonal languages undergoing tonogenesis, as well as the fact that CTP relate phonetically adjacent elements in the same syllable, while the elements participating in TTP are usually at a larger distance from a phonetic point of view.

We are now confronted with a situation in which two things are possible in the phonology, but one seems typologically more marked than the other. The more marked situation arises less often. We could respond to this situation in various ways, but the most common seems to be to assume:

1. That the fact that CTP and TTP are both possible (a fact o "hard typology") should be accounted for by phonological theory;

2. That the fact that CTP is more marked than TTP (a fact of "soft typology") should be relegated to e.g. phonetics.

However, Moreton (2006) argues that this cannot be right in this case; this particular case of soft typology has to be relegated to the phonological system in conjunction with a theory of language acquisition.

If our soft typological fact was the (indirect) result of phonetics, it should mean that the phonetic effects of TTP are bigger than those of CTPs: tones should have a more substantial effect on the F_0 of neighbouring tones than consonants do. Moreton (2006) did a literature survey on this issue, computing the effects that neighbouring tones and precedign consonants had. The results of these measurements are displayed in the following table:
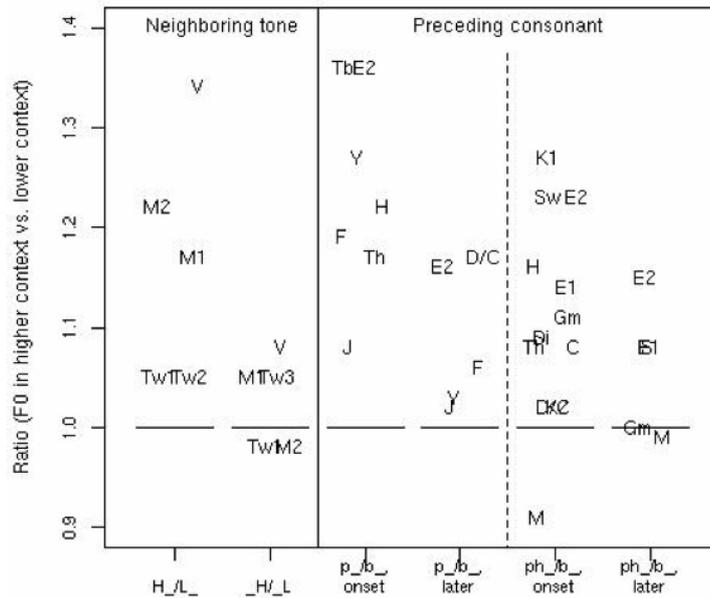
(69)

This table should be read as follows. Every plotting designates one of the studies taken into account by Moreton (2006). On the vertical axis, 1.0 means that there is no effect, higher numbers mean that there is a positive effect, and lower numbers that there is a negative effect. On the horizontal axis, H and L represent high and low toned contexts, p/b represents a voicing difference and ph/b an aspiration difference.

These studies thus provide no evidence whatsoever that the phonetic interaction between two tones is greater than that between a tone and the laryngeal features of obstruents. If there is any difference at all, it runs in the opposite direction.

We thus need to turn to a different kind of explanation. Moreton (2006) tries to find this in a theory of 'phonologization', i.e. of the way in which a learner may introduce phonetic effects into the phonological grammar. The idea is that the learner will have a preference for certain types of constraints; this preference can be stated in formal terms alone, so that even if both phonetic events are equally likely, TTP will be more easily phonologized than CTP.

In order to show how this works, Moreton (2006) presents a small model world, in which there is a Speaker who has the following lexicon of 16 words.

(70)   mápáp, mápàp, màpáp, màpàp,
       mábáp, mábàp, màbáp, màbàp,
       mápáb, mápàb, màpáb, màpàb,
       mábáb, mábàb, màbáb, màbàb

Furthermore, this Speaker pronounces all these words with equal frequency and fully faithfully, except that phonetically-biased coarticulatory distortions may occur. The other individual in this world is the Learner, who can undo most of the phonetic distortions, but sometimes will make a mistake and interpret a phonetic distortion as the result of phonology.

The relevant constraints are the following:

(71)   a. MAX-H: 'Don't delete an H tone'
       b. LH: 'No LH Tone Sequences'
       c. MAX-VOICE: Don't delete a [voice] feature.
       d. *BH: No voiced obstruent should be followed by a high tone.

Obviously, these constraints are a bit crude and not very insightful; we will return to this below. The constraints give us the following factorial typologies:

(72)   a. Tone-tone:
           i.   Faithful realisations: MAX-H≫*LH
           ii. Rightward L-Tone spreading: *LH≫MAX-H
       b. Obstruent-tone:
           i.   Faithful realisations: MAX-VOICE, MAX-H, ≫*bH
           ii. Post-/b/ Lowering: *bH, MAX-VOICE, ≫MAX-H
           iii. Pre-H devoicing: *bH, MAX-H, ≫MAX-VOICE

A crucial idea is that the two markedness constraints *bH and *LH are 'off-stage' in the initial state of the Learner, and they must somehow be added to the set of ranked constraints. This can mean that the constraint is in the set Con, but the grammar of an individual language does not contain all of the constraints of Con at least during acquisition, or it means that constraints are made 'on the fly' during acquisition.

According to some algorithm, whenever new data comes in, the Learner will have to decide whether or not to add a new constraint to her grammar. She will do this only if the space of possible grammars to search in will not be too big: adding a constraint which can be ranked in many different places will make the search space bigger than adding a constraint which does not interact with too many other constraints.

This reasoning paves the way for modularity, which is a property of grammars. Suppose that a grammar has m constraints which only talk about tone, and n constraints which only talk about segments. The grammar is then purely modular; the rankings of the tonal constraints with respect to each other matter, as do the rankings of the purely segmental constraints. Technically, this means that of the $(m + n)!$ mathematically possible grammars, we

only have to consider $m!n!$ relevant grammars. (Let us say that $m = 3$ and $n = 2$; this is a difference then between $5! = 120$ vs. $3! \times 2! = 6 \times 2 = 12$.)

If we add a new tonal constraint to our grammar, the number of possible grammars goes up to at most $(m + 1)!n!$ (in our example, it goes up from 12 to $4! \times 2! = 24 \times 2 = 48$. But if we add a tone-segment constraint, which unites the two dimensions, all of a sudden it is no longer possible to separate the two modules, and we go up to a possible number of rankings which may approach $6! = 720$. The Learner will thus be much more cautious before introducing such a intermodular constraint than before introducing a new modular constraint.

Moreton (2006) also built a computer programme which included this insight, plus our miniature world with the Learner trying to set up a grammar from what she hears from the Speaker. Indeed, under such a simulation, the cross-modular markedness constraint *bH was much more resilient to introduction than the purely modular markedness constraint *LH.

Notice that the whole line of reasoning depends only in part on the actual structure of the constraints involved. We could translate these constraints for instance in terms of Bradshaw (1999)'s analysis and still get the same result:

(73) a. *LH: L linked to μ_1 may not be immediately followed by H (linked to μ_2)

  b. *bH: : L linked to a consonant may not be immediately followed by H (linked to μ)

Both constraints have an OCP-like flavour. The modularity in this case would consist of the same feature being sensitive to attachment at two different dimensions vs. being sensitive to only one of these two dimensions. The relation between vocalic height and tone might be so 'multi-dimensional' (different features, different attachment) that it it would become very hard to be phonologized at all.

5 Tone and vowel height

5.1 The phonetics of tone and vowel height

Phonetically, tone corresponds to fundamental frequency (F_0) and vowel height corresponds to the first formant (F_1). Phonetic studies have shown that there is a clear correlation between these two: in a survey of 31 languages, Whalen & Levitt (1995) showed that in each of them the high vowels /i/ and /u/ have a higher pitch than the low vowel /a/, and many other studies have confirmed this for other languages as well. The following represents an average F_0 for Italian vowels (Esposito, 2002):
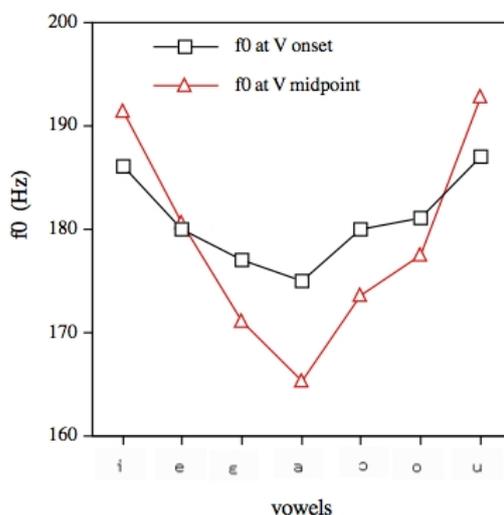
(74)

There is no general agreement as to what exactly explains this phonetic effect. One possible explanation is that in order to produce the high vowels, the jaw is lifted and this stretches the vocal chords, which produces the effect of a higher tone. E.g. Myers & Tsay (2003) claim:

> The articulation of the larynx is such that the raising of pitch height can be enhanced by moving the entire larynx upward with extrinsic muscles. This rotates the thyroid cartilage relative to the cricothyroid, thereby lengthening the vocal folds and raising pitch. This lifting of the larynx in turn raises the hyoid bone somewhat, to which the larynx is anchored at the top. The tongue root is also partly anchored to the hyoid bone, so raising this bone makes it easier to raise the tongue body as well. Thus raising pitch can reduce the energy required to produce a rise in vowel height.

Other correlations have been found as well, e.g. with rounding (Iivonen, 1987). But more importantly, High toned vowels tend to be longer than vowels with mid or low tone (Gandour, 1977).

The striking thing is that there are no uncontested examples of phonological reflexes of this phonetic effect. A possible explanation runs along the following lines (Hombert, 1977): Although intrinsic perturbations caused by consonantal influences and vowel height have similar absolute values, they are perceived differently. In particular, perturbations caused by prevocalic consonants are much more salient, perceptually, than perturbations caused by vowel height. There are two possible reasons for this difference in perceptual saliency:

1. A voiced consonant causes a relatively rising F_0 contour at the onset

of the following vowel. Contrary to this, the intrinsic F_0 associated with different vowel qualities is manifested by over-all higher versus lower F_0 levels. Dynamic changes in F_0 are much easier to detect than differences in level between two F_0 signals.

2. The F_0 perturbations caused by consonants are perceptually more salient, because they can be detected independently from the conditioning segments. On the other hand, in the case of F_0 perturbations caused by vowels, it is not possible to dissociate the F_0 differences from the conditioning segments, since they are both present simultaneously.

5.2 Height and Tone in Fuzhou

The most well-known case of a putative height-tone interaction may be that of the Northern Min language Fuzhou (Wang, 1967; Maddieson, 1976; Yip, 1980; Wright, 1983; Jiang-King, 1999; Myers & Tsay, 2003). In this dialect we find vowel alternations which seem dependent on tonal register:

(75)

LM, LML	æ	ai	au	ay	ei	ou	œy	ieu	uoi	'loose finals'
H, HL, M	ɛ	ei	ou	oy	i	u	y	iu	ui	'tight finals'

We basically find higher variant of the vowels when the tone is set in the higher register. We now find two approaches to this: one links the height shift directly to the register (e.g. Yip, 1980; Myers & Tsay, 2003); another tradition holds that this relation is rather indirect and mediated e.g. by length or mora structure (Wright, 1983; Jiang-King, 1999).

The classical instance of a 'direct' approach is Yip (1980), who provides the following rule:

(76)   V   →   -low   / ___
        [αlow]   -αhi   [+upper]

Jiang-King (1999) provides several arguments against such an approach.

- In the first place, it is not only height which seems affected: roundness and backness are also involved. Under the assumption of a direct height-height correlation these additional changes would be left unexplained.
- In particular also the simplification of some diphthongs (e.g. ei→i) is left unexplained.
- The hypothesis yields an 'implausible' vowel inventory: we have to assume that monophthongal high vowels {i, y, u} are not underlying, but only derived. Note, however, that this is only a problem for a direct approach implemented with a rule such as (76); we could try to set up

the alternations differently (with a constraint-based analyses) such that this problem would not arise.

- Cross-linguistic study shows that in some other Northern Min languages such as Fuqing or Fuan, Low tones may under certain circumstances also trigger tight finals.

Furthermore, Jiang-King (1999) observes that syllable structure position is also relevant for the relevant alternations. Consider, for instance the behaviour of /i/ and /u/:

(77)

	tight		loose	
a.	ki^{HM}	'reach'	$kεi^{MLM}$	'lucky'
	ku^{HM}	'alone'	kou^{MLM}	'old; reason'
b.	$tsieŋ^{H}$	'stick'	$tsiεŋ^{MLM}$	'fight'
	puo^{HM}	'vigorous'	$puɔ^{MLM}$	'to peel; to shell'

The high vowels change if they are in the syllabic nucleus (77a), but not if they are in the off-glide (77b). This cannot be captured by a 'direct' approach.

All in all, Jiang-King (1999) lists the following properties for tight vs. loose finals in Fuzhou:

(78)

	Tight finals	**Loose finals**
Segments:	monophthongs	diphthongs
	tense vowels	lax vowels
	round harmony	no harmony
	no VVC	VVC
Tone:	H, ML, HM	MHM, MLM

Jiang-King (1999) proposes an *indirect* approach in which the relation between tone and the segmental content is mediated by the moraic syllable structure: certain tones prefer certain syllable structures, and so do certain segmental feature configurations. In particular she argues that 'tight' syllables have one mora, whereas 'loose' syllables have two:

(79)

tight final	loose final
σ	σ
$\mid$	$\wedge$
$\underline{\mu}$	$\underline{\mu}\ \mu$

Note that we underlined the first mora in both representations. This is a way of distinguishing between a *head* and a *non-head* mora: the first, which is underlined, will always dominate a vowel and more in general be the most prominent of the two moras.

One reason for adopting this representational refinement in the study of Fuzhou is that it allows us to explain why there are at most *three* tones in a tonal contour.

One alternative way of getting the same result would be to assume that Fuzhou syllables can be trimoraic, but this runs against the assumption that syllables are universally at most binary branching. (Some prosodic constituents have been argued to not obey this restriction, but those then have unbounded branching; languages with at most ternary branching have not been attested.)

Yet another alternative would be to assume that there is a feature geometric node for tonal contours, as in (23). However, the problem then is that such a node might be attached to *both* moras, so that we could get a tonal contour of 4:

(80)

$$
\begin{array}{c}
\sigma \\
\diagup\ \diagdown \\
\underline{\mu}\quad\ \mu \\
|\qquad| \\
\mathrm{o}\qquad\mathrm{o} \\
\diagup\diagdown\ \diagup\diagdown \\
T_1 T_2 T_3 T_4
\end{array}
$$
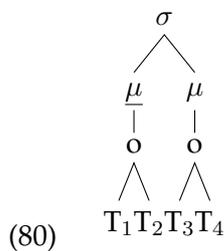
Jiang-King (1999) therefore chooses to follow Hyman (1988) and assume that the asymmetry between head moras and non-head moras is reflected in their ability to host tones: the former can host two tones, the latter only one. She proposes that this is regulated by the following OT constraint:

(81) HDBIN (head binarity): A mora M must bear two tones *iff* M is a syllable head.

If this constraint interacts with faithfulness constraints on tone (82) and with the Well-Formedness Condition (7), we get tight syllables for some tonal inputs and low syllables for others.

(82) a. PARSE-TONE: Underlying tones need to be parsed (no deletion of tones)

   b. LEX-TONE: Only lexically sponsored tones are allowed (no insertion)

(83) a.

H		PARSE-TONE	LEX-TONE	HDBIN	WFC
a. ☞	H | μ			*	
b.	H L ∨∕ μ		*!		*
c.	H μ	*!		*	

b.

HM		PARSE-TONE	LEX-TONE	HDBIN	WFC
a. ☞	H M ∨∕ μ				*
b.	H M ∨⌐ μ μ				**!
c.	H M | | μ μ			*!	
d.	H M | μ μ	*!		*	

c.

MLM		PARSE-TONE	LEX-TONE	HDBIN	WFC
a. ☞	MLM ∨∕ | μ μ				*
b.	MLM | ∖ μ μ			*!*	*
c.	MLM | | μ μ	*!		*	*
d.	MLM ∨∕ μ			*!	**

It thus follows from this system that an input with three tones will lead to a bimoraic unit; tonal distribution will determine syllable weight. But notice that several aspects of the analysis have not been worked out:

- It is not clear how we can actually prevent a system with four tones, given the high ranking of faithfulness in this system (in particular, the high ranking of PARSE-TONE), we expect actually any underlying tone to surface. Notice, however, that this is due to a mistake in Jiang-King (1999)'s ranking argument; nothing goes wrong in the tableaux if we shift PARSE-TONE to a position below HDBIN, but this shifting will now penalize any underlying tones which are attached to non-head positions. (We will also need to make precise that ternary branching tone structures are not allowed on any mora, possibly universally.)

- Another problem is that it is not exactly clear what happens with a putative input — which we should take into consideration given Richness of the Base. Toneless syllables do not seem allowed in Fuzhou, so somehow we should make sure that these get a default tone. The most likely candidate for this seems H; but this means that also LEX-TONE is dominated, for instance by part of the WFC, viz. the part that says that every mora needs a tone.

- We also do otherwise not dispose of an explanation for the tonal inventory of Fuzhou. For instance, there is no formal explanation of why there cannot be a level Low tone. However, Jiang-King (1999) does discuss a similar effect in the related language Fuqing.
  In this language, tight syllables can have one of the tones HM, H or M, whereas loose syllables have one of the tone HL, ML, ML. Jiang-King (1999) accounts for this by assuming that L tone can only be linked to the second (non-head) mora of the syllable; whenever there is a Low tone, we thus need a bimoraic unit.
  Formally, Jiang-King (1999) assumes (i) that [+Upper] and [-Raised] are the marked features, (ii) the unmarked feature values [-Upper] and [+Raised] are absent from the representation, and (iii) there is a 'Tonal Sonority Hierarchy' [+Upper] > [-Raised]. This gives us the following inherent ranking of constraints (such an inherent ranking presumably finds its origin in the phonetics, although it is not clear how):

  (84) *NUC$^\mu$/[-RAISED]≫*NUC$^\mu$/[+UPPER]

  A constraint *NUC$^\mu$/[-RAISED], in effect militating against Low tones on head moras, will make sure that underlying Low tones will always end up on the second mora; but again this means that both faithfulness constraints need to be dominated by this constraint.
 Notice, by the way, that Fuqing allows at most binary tones, which seem to argue for a ranking of WFC≫HDBIN (but Jiang-King (1999) does not seem to notice this; as a matter of fact, the constraint rankings she gives seem entirely wrong).

- Finally, we also will need to make sure that underlying moraic structure is not important in Fuzhou (or Fuqing). For languages with contrastive

> vowel length, it is usually assumed that we have faithfulness to under-
> lying moras. Such constraints should be low-ranked in the Northern
> Min languages.

In spite of these technical problems, we seem to have thus established that
tones on 'tight syllables' trigger a monomoraic structure, whereas tones on
'loose syllables' trigger a bimoraic structure. It now remains to be seen how
these two structures in turn correspond to the vocalic behaviour of these syl-
lables.

Jiang-King (1999) proposes that the two syllable types look as follows:

(85)
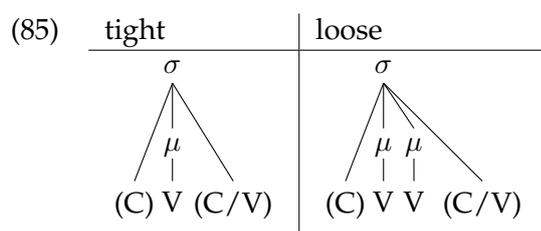
Various of the observed differences in (78) now follow more or less automat-
ically. For instance, there is a contrast between monophthongal high vowels
in tight syllables vs. diphthongs in loose syllables (77a). Now let us assume
that in both cases we have an underlying high vowel. In tight syllables, this
shows up on the single mora; the question is why we find diphtongisation
on loose syllables.

Jiang-King (1999) observes that in the diphthongal structures both parts
of the diphtong share all features, except for [+high]. The idea is that there
is again a sonority related constraint which disprefers high vowels in head
moras:

(86) $^*\textsc{Nuc}^\mu/[+\textsc{high}] \gg {}^*\textsc{Nuc}^\mu/[+\textsc{low}]$

If this constraint is outranked by Parse-[+high] (plus the constraints which
get the number of moras from the underlying tonal contours and nothing
else), we get as an effect that underlying high vowels will be realised on the
head mora in tight syllables. However, in loose, syllables, we have more
freedom: we can satisfy $^*\textsc{Nuc}^\mu/[+\textsc{high}]$ without violating Parse-[+high],
viz. by the following output candidate (for pɛiŋ^{MLM} 'combine'):

(87)   -b+h   -b = [-back]; +h = [+high]; o = vocalic nodes
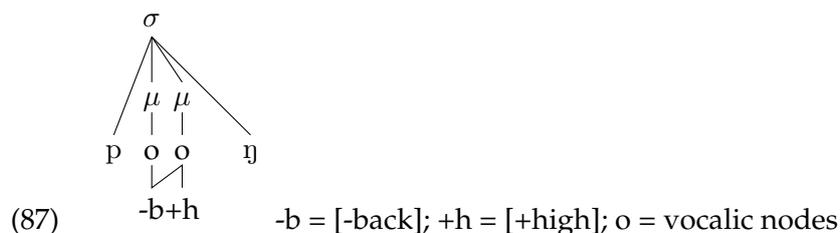
Another property that follows is the rounding harmony, which we find in tight finals but not in loose finals, at least if we assume that segments which are linked to the same mora are closer together — and therefore more likely to undergo assimilation — than segments which are further apart.

An even more interesting type of harmony is that within tight syllables we cannot find diphthongs of the shape [ɑi], i.e. a low vowel, followed by a high vowel. We will find [ɐi] instead, and we could assume that this is due to the following constraint:

(88) *HI/LO/μ: a mora cannot be filled by both [+Hi] and [+Lo] F-elements.

Notice that there is some phonetic grounding for a constraint such as this (it prevents articulatory effort) while at the same time it crucially refers to an abstract phonological category, the mora.

The tense/lax distinction requires some more elaboration. We find pairs such as the following:

(89)

tight		loose	
tsieŋ^{HL}	'felt'	tsiɛŋ^{MLM}	'fight'
ko^{H}	'song'	kɔ^{MLM}	'individual'

This suggests that lax vowels occur in bimoraic syllables, and tense vowels in monomoraic syllables. Jiang-King (1999) proposes the following constraint for this which is obviously doing the job:

(90) LAXING: If α is parsed onto two moras, then α is [Lax] (where $\alpha \neq$ [high])

The constraint may seem a little bit *ad hoc* and even running counter to some typological generalisations. For instance, it has been proposed for many Germanic languages that tense vowels are long and lax vowels are short. However, this has also been contested, and e.g. van Oostendorp (2000) proposes that in these Germanic languages the relevant constraint is the following (approximately, we abstract away here from some of the details of the syllable structure analysis given in that work):

(91) CONNECT: A vowel α is in a heavy syllable *iff* α has the feature [lax]

Notice that this formulation seems more precise because of the *iff* clause, which sseems intended, but not expressed, by Jiang-King (1999).

5.3 A direct account

All in all, it seems that Jiang-King (1999) has proposed a plausible account of duration (abstract duration in terms of mora structure) plays an intermediary

role between vowel height and tone. However, it is fair to say that this view has also been contested.

Myers & Tsay (2003) argue in favour of what they call a 'formal functional view' of tone. Their objections against indirect approaches are the following:

- "These alternative analyses neglect the facts that not all of the alternations change prosodic structure (e.g [æ]-[ɛ], [ɑi]-[ɐi]." Notice that this is simply not true, since Jiang-King (1999) does discuss these changes in terms of changes in prosodic structure.
- "Moreover, while interactions between tone and vowel height are relatively rare, they are not entirely unheard of; in addition to Fuzhou, they have also been reported in Cantonese (Hashimoto, 1972), Hausa (Pilsczikowa-Chodak, 1972; Newman, 1975), Lahu (Matisoff, 1973) and even Mandarin (Chao, 1948; Tsay & Sarwusch, 1994)." Notice that this is a strange list of references; e.g. Newman (1975) contests the claim that there is interaction between tone and vowel height in Hausa, Tsay & Sarwusch (1994) is co-authored by one of the authors of this article, and a few other references are textbooks or grammars for non-linguists. Furthermore, the fact is not mentioned that Jiang-King (1999) does talk (briefly) about Cantonese, Mandarin and Hausa. The list is therefore far from impressive.

In any case, Myers & Tsay (2003) propose a direct approach. In their view, "since the route from pitch raising to vowel raising is rather indirect, the physiological effect is rather small (Honda, 1983; Honda *et al.*, 1994)". Because it is so small (from an articulatory point of view) it will not be very often phonologized.

Myers & Tsay (2003) propose functionally motivated "coordination constraints", following Boersma (1998):

(92) *COORD(gesture$_1$, gesture$_2$): Gesture$_1$ and gesture$_2$ are not coordinated.

The idea is that we have a universal ranking of such constraints, in which e.g. *COORD(H,æ)≫*COORD(H,ɛ)≫*COORD(H,i). Furthermore, in order to account for the fact that the effect of vowel height is much smaller than the effect of voicing, they stipulate that (universally) coordination constraints on voicing and tone are ranked higher than those on height and tone.

6   Franconian tones: phonology vs. phonetics

6.1   Introduction

Most Limburg dialects of Dutch, as well as neighbouring dialects of German — the dialect area is usually referred to as 'Franconian' — display a lexi-

cal contrast between two types of tone, a falling tone (traditionally called *stoottoon*) and a level high tone (*sleeptoon*). The following examples are from Maasbracht:

(93) *falling tone*   *level high tone*
      mı́m̀ 'minus'   mı́ń 'vile'
      dǽm̀ 'fir'   dǽń 'then'
      klám̀ 'trap'   kláḿ 'hardly'
      bı̂ı 'bee'   bı́ı́ 'with'
      zı̂ı 'side'   zı́ı́ 'she'
      pı̂ıp 'to squeak'   pı́ı́p 'pipe'

The tonal contours can also be used to express (inflectional) affixation, e.g. for singular/plural pairs:

(94) *falling tone*   *level high tone*
      bέm̀ 'legs'   bέń 'leg'

The following two pictures represent the F0 values for these two tones (for a speaker from the Roermond dialect, very close to Maasbracht[6]):

(95) falling tone   level high tone

The 'falling' tone is characterised by a clear downward movement; the 'level high' tone also moves slightly downward, but then goes up again towards the end. There are several ways to translate this into the phonology, but many analysts have converged on the following (see Gussenhoven, 2004, for an authoritative overview):

(96) falling tone   level high tone
       HL   H(L)H
       μ μ μ μ
      m ɪ n   m ɪ   n

---

[6]The data were analysed with the *Praat* programme; http://www.praat.org/. The data are almost identical to those presented in Gussenhoven (2000).

Limburg tones have several properties which can be readily explained in a phonetically-based framework of phonology. Take for instance the following observations:

1. Tonal contours are only expressed on a long vowel or a V+sonorant ('VR') rhyme.
2. Every word has at most tone contour, which is expressed on the syllable carrying main stress.

These fit with observations we have already seen in previous classes and seen from this angle, the motivation behind some of the choices made in Limburg dialects may therefore be phonetic. Still, the way it works out here (as well as in other tone languages) has to be phonological. The reason is that the ban on tones outside of the main stress and syllables with a long vowel or sonorant rhyme is absolute. This applies also to loanwords, which will always have one of the two tones on the stressed syllable, given appropriate conditions. This means that analyses in which there is no role for a (phonological) system are difficult to maintain. Zhang (2004) formalizes the analysis in terms of a family of OT constraints:

(97)   CONTOUR(x_i)-C$_{contour}$(y_i)
       no contour tone x_i is allowed on syllable with the C$_{contour}$ value of
       syllable y_i or smaller

The constraints in this family are inherently ranked. "These rankings reflect the speaker's knowledge that a structure that is phonetically more demanding should be banned before a structure that is less so; and that a syllable should be able to host a tone with a lower complexity before it can host a tone with higher complexity." This places Zhang (2004) in the school of thought which was exemplified in our course by Hayes.

   Of greater interest is the fact that the Limburg dialects also seem to run counter to at least some of the proposed universals. In some cases the problem may be only apparent. Take, for instance, the following (Zhang (2004)'s example (13a)):

(98)   If a language has contour tones, then it also has a level tone.

Under the analysis presented above, Limburg dialects only have contour tones HL and HLH. However, there is actually even more than one way out of this problem. For instance, we may analayse HLH as actually being a level high tone, although this would be a rather abstract analysis in many cases — one which we might want to avoid if we follow the lines of Zhang (2004). A more viable option might be to consider that other factors are interfering here. For instance, the fact that the first mora of the stressed syllable might be attributed to independent functionally motivated constraints, such as the

fact that stress tends to be interpreted as high toned in any case. This would leave the choice of lexical specification restricted to the second mora, which could be only high or low. (98) (which is by the way called a 'strong implicational tendency', not a universal by Zhang (2004)) would thus be violated for a reason in Limburg.

Yet, the question always is how many ways out we allow ourselves. Notice that Zhang (2004)'s arguments have all been taken from the realm of articulatory ease, but the argument about the high tone on stress has to come from some other area (maybe processing). It is the question how much knowledge about how many areas the child may bring to bear on language acquisition (and to what extent this gives us a parsimonious theory).

In the following sections, we will go into more detail into two more specific problems, having to do with the interaction with consonants and vowels respectively. In both cases, the generalisations we can make seem to run counter to what we would expect on the basis of phonetic facts only. We will see how these seem to rule out certain accounts.

6.2 Interaction with consonants

Lexical tone interacts with laryngeal features on consonants in a number of ways; they all have in common that a falling tone prefers not to be followed by a voiceless obstruent. Here is an example; if a stressed vowel is followed by a sonorant plus a voiced obstruent, we have a potential contrast:

(99) Roermond contrast in voiced environment

Falling tone	Level tone
bændə 'gang'	dʊndər 'thunder'
hærdər 'shepherd'	mærɣəl 'marl'

If the last consonant of this sequence is voiceless, however, only the level tone is allowed:

(100) Roermond gap in voiceless environment

Falling tone	Level tone
(missing)	pɪmpəl 'booze'
(missing)	hærsəs 'brains'

This gap suggests the following generalisation:

☞ *Consonant-tone interaction in Roermond*
A voiceless consonant disprefers a low tone on a consonant at its left.

Now interactions between laryngeal features and tone are well-known from other languages. In Suma (Bradshaw, 1999), for instance, imperfective verbs start with a High tone (*éé* 'leave behind', *kírí* 'look for'), except when they

begin with a voiced obstruent, in which case the first tone is rising (i.e. Low-High: *bŭsi* 'be bland'). Diachronically, the effect is known as well from other tone languages, where voicing contrasts on onset consonants turn into tone contrasts on following vowels. A second type of interaction is that the spreading of a high tone is blocked by an intervening voiced obstruent and the spreading of a low tone is blocked by an intervening voiceless obstruent (Hyman & Schuh, 1974). Botswana Kalang'a (Bradshaw, 1999) is a language where the former process applies. In this language high tones normally spread to toneless (i.e. phonetically low-toned) syllables in a following word (/tʃipó + tʃipó/ ← [tʃipótʃípó] 'your gift'), except if a voiced obstruent intervenes (/zwipó + zwipó/ ← [zwipózwipó] 'your (pl.) gifts').

In all known cases, it is the following vowel that is affected by the consonant, and any theory has to account for this. But all of these theories are then facing a serious problem with the Limburg facts. This is specifically true for the phonetic optimisation approach: if it is optimising to link an laryngeal feature in the onset to a following vowel, the Limburg facts seem 'crazy' and in need of a separate constraint.

A diachronic (evolutionary) approach for part of the Limburg facts is possible, however. Most scholars assume that the rise of tone is connected in one way or another to a process of schwa apocope by which e.g. *ouge* 'eye' turned into *oug*. Although there is no consensus on what this relation is exactly, this gives us the opportunity to give a typologically more plausible analysis to the Limburg phenomenon: we may assume that the Low tone originated on the schwa following the voiced obstruent:

(101) $\begin{array}{ccc} \text{ɔ uɣə} & & \text{ɔ uɣə} \\ |\quad | & \Rightarrow & |\quad \diagdown \\ \text{H\quad L} & & \text{H\quad L} \end{array}$

The origin of this phenomenon could thus have been a phonetically natural one (whatever causes the misperception of voicing on obstruents as low tone on the following vowel), but it needs to be noted that the system has been subsequently heavily phonologized in ways that stretch far beyond these phonetic origins.

One example of this is from the Moresnet dialect (Jongen, 1972). Different from most other dialects, short vowels followed by an obstruent also get tone. This tone is falling if the vowel is followed by voiced obstruent, but level if followed by a voiceless obstruent. Thus, in a word like *bedde* 'bed-PL' the first vowel has a falling tone, whereas in words like *teppich* 'carpet' and *kes* 'casket' it is level.

Interestingly, words that are devoiced by Final Devoicing have a falling tone, according to Jongen. Thus, *bet* 'bed-SING' has a falling tone. This seems to suggest that the tone of the short vowel is determined at the underlying level. That is, since *bet* has a voiced consonant underlyingly, the vowel preceding it has a falling tone, even at the surface level.

It is hard to see how a synchronic or diachronic account which is based directly on phonetics is going to make sense of this, because the effect is clearly very opaque: something happens only to devoiced segments, not to segments which are underlyingly voiceless, or voiced on the surface. But 'devoiced', referring to the phonological derivation, is obviously not a phonetic category.

The only approach which has some hope of shedding light on this, is the one by Boersma: the tonal change on the vowel acts as a recoverability marker for the underlying voicing. The hearer gets a cue from the 'wrong' tone about the underlying status of the consonant. Yet, if tone on a vowel is usually a cue for the voicing of the *preceding* segment, it is still not clear how this would work exactly.

Notice that a more abstract approach is available; there have been various places in the literature where it is suggested that from a phonological point of view, Low tone and [+voice] are the same animal, which we will represent here as L. Under such an approach, the tonal shift is a type of faithfulness:

$$
\text{(102)} \quad \begin{array}{c} \text{bɛd} \\ | \\ \text{L} \end{array} \quad \Rightarrow \quad \begin{array}{c} \text{bɛd} \\ | \\ \text{L} \end{array}
$$

Phonetically low tone and voicing are not obviously the same thing. We thus need this level of abstractness to maintain this analysis.

6.3   Interaction with vowels

The interaction between tone and vocalic height is very puzzling from a typological point of view. In the first place, the fact that there is any phonological interaction between these two dimensions at all sets Limburg apart from all other known tone languages. There is some phonetic interaction, but this is never phonologized — in itself an observation that is worth considering.

In Limburg, the connections between vowel quality and tone are many, but the basic observations are that we find diphthongization under a falling tone — e.g. long /eː/ or /iː/ turning into [ɛi] — and monophthongization under a level high tone — e.g. /ɛi/ turning into [eː]. Other cases involve lowering of diphthongs and mid vowels under a falling tone — e.g. /ɛi/ changing to [ɑi], and /eː/ to [ɛː] — and raising of mid vowels under a level high tone — changing e.g. /ɛː/ to [eː]. A further problem with these developments is that they go in the wrong direction: if there is a phonetic effect, it is that high tones want to go with high vowels; but here the level high tones seem to prefer low vowels and vice versa.

By way of an example, the following gives an overview of vocalic changes which happened to the Maastricht dialect (de Vaan, 2004).

(103)

West-Germanic vowel	Falling tone	Level tone
iːC, yːC, uːC	ɛi, øy, ou	iː, yː, uː
iːr, yːr, uːr	eː, øː, oː	iː, yː, uː

In order to explain effects such as these (we will concentrate on the lowering), which run in the reverse direction of what we want Gussenhoven & Driessen (2004) propose the following phonetic explanation. Falling tones have to be perceptually short. Now normally high vowels are actually shorter than mid vowels (and it is more difficult to keep to the length of high vowels from an articulatory point), so this would again normally run in the wrong direction, but for this reason Gussenhoven & Driessen (2004) assign a special role to the listener. Since she knows that high vowels are usually shorter, she will automatically always add something to their length; and in an experiment they proved that if we present an [iː] and an [eː] of exactly the same length, say 220 ms, to a set of speakers, they will perceive the former as longer than the latter. (If we present an elephant and a mouse of the same absolute height, the elephant will be perceived as very small and/or the mouse as very big.) This is why a vowel which is usually short can be used as a cue for length.

Various things can be said about this approach; for instance, it runs the danger of paving the way for very unrestricted analyses. If we find an effect A, we can attribute this to articulation; but if we find ¬A, we can attribute this to the fact that the listener subtracts ease of articulation. We can thus explain one effect and the contrary effect at the same time, which is not a very desirable state of affairs.

Again, a more abstract approach presents itself. Let us assume that the basic idea of Gussenhoven & Driessen (2004) is correct, viz. that falling tones in Maastricht want to be represented as short. This seems particularly attractive in this dialect, where length is the most important cue for the tonal distinction (Gussenhoven & Aarts, 1999). But let us also assume that this is a *phonological* generalisation.

Now we can observe that Maastricht contrasts short [i] to long [iː] (there are minimal pairs), but it has only one tense mid vowel [e]. This means, then, that [iː] is really, phonologically, long, whereas [eː] is phonologically only tense. Seen from that perspective, the change from [iː] to [e] is thus a phonological shortening:

(104)

$$
\begin{array}{ccc}
\sigma & & \sigma \\
\diagup | \diagdown & & \diagup | \diagdown \\
x \; x \; x \; x & & x \; x \; x \\
| \; \diagdown\!\diagup \; | & & | \; | \; | \\
v \; i \quad r & \rightarrow & v \; e \; r
\end{array}
$$
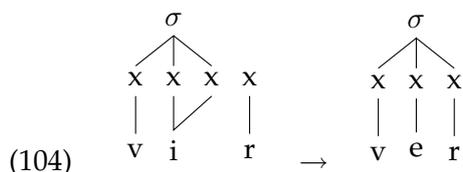
Again, it is hard to see how this could be captured directly in phonetic terms (since [e] is not necessarily shorter than [iː]).

6.4 Tone ⊃ Diphthong and Diphthong ⊃ Tone

To conclude, we will look at a few slightly more complicated examples, comparing two dialects (Maasbracht and Sittard, spoken at 30 km distance). Sittard presents us with a case of 'Tone ⊃ Diphthong':

(105)   Tone ⊃ Diphthong: if we have a falling tone, long mid vowels diphthongize.

In this dialect long mid vowels diphthongize, but only if they carry a falling tone (not a dragging tone). Below, we compare forms of the Sittard dialect with equivalent forms of another Limburg dialect, Maasbracht. The Sittard forms are taken from Dols (1953); the Maasbracht forms are taken from Hermans (1994).

(106)   (examples with a falling tone)

Maasbracht	Sittard	
a. front, non-round mid long vowel		
keːze	keize	'to choose'
keːzel	keizel	'gravel'
bedreege	bedreige	'to cheat'
b. front, round mid vowel		
vrøːtele	vrœyte	'to grub'
bedrøːftj	bedrœyf	'sad'
vrøːg	vœyg	'early'
c. back round mid vowel		
aːlmoːs	aːlmous	'alms'
voːt	vout	'foot'
boːk	bouk	'book'

The inverse is not true: there is no diphthongization under a dragging tone, hence Tone ⊃ Diphthong does not apply:

(107)   (examples with a dragging tone)

Maasbracht	Sittard	
a. front, non-round mid long vowel		
beːt	beːt	'bite'
smeːt	ʃmeːt	'smith'
streːk	ʃtreːk	'region'
b. front, round mid vowel		
døːr	døːr	'door'
vøːr	vøːr	'before'
røːk	røːk	'smell'
c. back round mid vowel		
toːn	toːn	proper name
woːs	woːs	'sausage'
doːn	doːn	'to do'

From the perspective of phonetics these facts are totally unexpected; they directly go against the universal phonetic correlation between tone and vowel height.

Maasbracht shows evidence for the reverse correlation:

(108)   Diphthong ⊃ Tone: centering diphthongs only occur under a falling tone.

(109)   (examples with a falling tone)

Maasbracht	Sittard	
a. front, non-round mid long vowel		
biəs	beːs	'beast'
iəlɛnj	eːlɛnj	'misery'
fiərtɪx	feːrtɪx	'forty'
b. front, round mid vowel		
dyər	døːr	'proper name'
dryəx	drøːx	'dry'
yərmœnj	øːrmœnt	place name
c. back round mid vowel		
buən	boːn	'bean'
vijuəl	fejoːl	'violin'
fuəj	foːj	'tip'

'Normal' mid vowels also occur under a falling tone, hence Diphthong Tone. For examples cf. (106).

From the perspective of a direct phonetic approach, again these facts are unexpected, so that we should try now a (brief) analysis of these facts; like Jiang-King (1999), we will assume that syllable structure is intermediating. We make one important simplifying assumption:

☞ Vowels that are realized as a diphthong are underlyingly tense, but short.

This assumption is simplifying, because Richness of the Base requires us to assume also inputs which have underlying inputs, maybe even with the 'wrong' tone. The theoretical machinery adopted here cannot completely deal with these wrong inputs yet.

(110)   underlyingly short   underlyingly long
         (cf. (106))   (cf. (109))

$$
\begin{array}{cccc}
\text{H} & & \text{H} & \text{L}\\
| & & | & |\\
\mu & & \mu & \mu\\
| & & | & |\\
\text{k}\ \text{e}\ \text{z}\ \text{ə} & & \text{e}\ \text{e}\quad\text{lɛnj}
\end{array}
$$

We further claim that short vowels in open stressed syllables are lengthened due to Prokosch' Law (Prokosch, 1938):

(111)   PROKOSCH: A stressed syllable is bimoraic.

The idea is that the new mora inserted by Prokosch' Law is filled by default elements; low tone at the tonal level (112) and high vowels at the segmental level (113).

(112)   HEIGHTASYMMETRY (after de Lacy (1999))
         a. A High tone may only be linked to a head position.
         b. A Low tone may only be linked to a dependent position.

(113)   MARGIN (Prince & Smolensky, 1993)
         a. A low vowel may only be linked to a head position.
         b. A high vowel may only be linked to a dependent position.

In Maasbracht there is no diphthongization of the Sittard type, and this must be due to some well-formedness constraint which we will assume to be of the following type:

(114)   ANTIDIPHTHONG A vocalic feature must be linked to the head position.

(115)   Grammar fragments
         a. Sittard: MARGIN≫ANTIDIPHTHONG
         b. Maasbracht: ANTIDIPHTHONG≫MARGIN

Maasbracht, however, has a different type of diphthongisation, the centering diphthongs. Here we assume that schwa lacks a place node (Anderson, 1982; van Oostendorp, 2000), so that these dipthongs do not violate AntiDiphthong (!) Furthermore, we assume that HeightAsymmetry also applies within segments, and that segments have a head, e.g. the place node:

(116) HeightAsymmetry (within segments)
       A High tone may only be linked to a head of a segment.

The idea is thus that the most prominent tone cannot be realised on a segment which is least prominent. That is why H cannot occur on schwa.

There is a further fact of Maasbracht tonology we might mention: like in Fuzhou, there is a connection with tenseness on the vowels as well, although it runs in the opposite direction.

(93) has shown that there can be contours on CVV and CVR syllables in Maasbracht Dutch. However, CVR syllables with a (high) tense head only carry falling tones (high tense vowels are the only short tense vowels):

(117)   (all examples have a falling tone)
        min   proper name
        pil   'to bungle'
        ʒyl   proper name
        byl   'bag'
        pun   'money' (colloquial)
        dul   'purpose'

A constraint that directly relates tenseness and tonal quality would have to be very complicated and non explanatory. It can however be assumed that the consonants after tense vowels are in a different syllable position from those immediately following lax vowels: they behave like they are more external. Independent evidence for this comes from the stress behaviour: syllables with a tense high vowel followed by a tautosyllabic sonorant are always stressed, whereas lax vowels in the same environment are always skipped by the stress rule.

(118)
Tense vowel		Lax vowel	
tal[ýt]	'talus'	réb[œs]	'rebus'
gamb[ít]	'gambit'	bát[ɪk]	'batik'

6.5 Conclusion

We thus see that the Limburg tones interact with consonants and with vowels in ways that are very hard to capture in terms of phonetics directly. If phonetic generalisations have played a role at all, it is most likely that they

have done so in a diachronic sense, more or less in the way this is depicted in Evolutionary Phonology. But after these generalisations have entered the grammatical phonological system, they have been subjected to rules of this cognitive game. That seems to be the most important lesson to take home after this course.

Bibliography

Anderson, Stephen (1982). 'The analysis of French schwa: or, how to get something for nothing'. *Language*, 58, 3: 534–573.

Archangeli, Diane & Douglas Pulleyblank (1994). *Grounded Phonology*. Cambridge, Mass.: The MIT Press.

Bao, Zhi Ming (1990). *On the Nature of Tone*. Ph.D. thesis, MIT, Cambridge, Mass.

Baudouin de Courtenay, Jan (1895). *Versuch einer Theorie phonetischer Alternationen: Ein Capitel aus der Psychophonetik*. Strassburg/Crakow: Trübner.

Beach, D.M. (1924). 'The science of tonetics and its application to Bantu languages'. *Bantu studies*, 2: 75–106.

Bermúdez-Otero, Ricardo (2005). 'Diachronic Phonology'. In: de Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press.

——— (forthcoming). *Stratal Optimality Theory*. Oxford: Oxford University Press.

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

Bradshaw, Joel (1979). 'Obstruent harmony and tonogenesis in Jabêm'. *Lingua*, 49, 189–205.

Bradshaw, Mary (1999). *A Crosslinguistic Study of Consonant-Tone Interaction*. Ph.D. thesis, Ohio State University.

Bybee, Joan (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Carstairs-McCarthy, Andrew (1999). *The origins of complex language: an inquiry into the evolutionary beginnings of sentences, syllables and truth*. Oxford: Oxford University Press.

Chao, Y.-R. (1948). *Mandarin Primer*. Harvard University Press.

Chen, Matthew (2000). *Tone Sandhi*. Cambridge, UK: Cambridge University Press.

Chomsky, Noam (2004). 'Beyond explanatory adequacy'. In: Belletti, Adriana (ed.), *Structures and Beyond: The Cartography of Syntactic Structures, Vol. 3*. Oxford: Oxford University Press.

Clements, Nick (1978). 'Tone and syntax in Ewe'. In: Napoli, Diane (ed.), *Elements of Tone, Stress, and Intonation*, pp. 21–99. Washington: Georgetown University Press.

Clements, Nick & Samuel Jay Keyser (1983). *CV Phonology: A Generative Theory of the Syllable*. Cambridge, Mass.: MIT Press.

Dols, Willy (1953). *Sittardse diftongering. Een hoofdstuk uit de historische grammatica*. Sittard: Albert's Drukkerijen.

Duanmu, San (1990). *A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages*. Ph.D. thesis, MIT.

——— (1994). 'Against Contour Tone Units'. *Linguistic Inquiry*, 25, 4: 555–608.

Esposito, Anna (2002). 'On Vowel Height and Consonantal Voicing Effects: Data from Italian'. *Phonetica*, 59, 2: 197–231. URL `http://vislab.cs.vt.edu/Publications/2002/PDFfiles/Esp02.pdf`.

Féry, Caroline (2003). 'Liaison and syllable structure in French'. URL `http://www.sfb632.uni-potsdam.de/homes/fery/pdf/liaison%20actual.pdf`. Manuscript, Potsdam University.

Flemming, Edward (1995). *Auditory Representations in Phonology*. Ph.D. thesis, University of California at Los Angeles.

Gandour, J. (1977). 'On the interaction between tone and vowel length: Evidence from Thai dialects.' *Phonetica*, 34, 54–65.

George, I. (1970). 'Nupe tonology'. *Studies in African Linguistics*, 1: 100–122.

Goldsmith, John (1990). *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.

Gordon, Matthew (2001). 'A Typology of Contour Tone Restrictions'. *Studies in Language*, 25: 405–444.

Gordon, Raymond (2005). *Ethnologue. Languages of the World*. Dallas: SIL International, 15 edn. URL `http://www.ethnologue.com/`.

Gussenhoven, Carlos (2000). 'The lexical tone contrast of Roermond Dutch in Optimality Theory'. In: Horne, Merle (ed.), *Intonation: Theory and Experiment*, pp. 129–167. Dordrecht: Kluwer. URL `http://www.let.ru.nl/gep/carlos/roermond.pdf`.

——— (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos & Flor Aarts (1999). 'The Dialect of Maastricht'. *Journal of the International Phonetic Association*, pp. 155–160.

Gussenhoven, Carlos & Wilske Driessen (2004). 'Explaining two correlations between vowel quality and tone: The duration connection'. In: *Prosody 2004*. Nara, Japan: Prosody.

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

Hale, Mark & Charles Reiss (2000). 'Phonology as cognition'. In: Burton-Roberts, N., Philip Carr, & Gerard Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Foundations*, pp. 161–184. Oxford: Oxford University Press.

Halle, Morris & Kenneth Stevens (1971). 'A note on laryngeal features'. In: *Quarterly progress report 101*, pp. 198–213. Cambridge, Mass: Research Laboratory of Electronics, MIT.

Hansson, Gunnar Ólafur (2004). 'Tone and Voicing Agreement in Yabem'. In: B. Schmeiser, A. Kelleher, V. Chand (ed.), *WCCFL 23 Proceedings*. Cascadilla Press.

Hashimoto, O.-K. Yue (1972). *Phonology of Cantonese*. Cambridge, UK: Cambridge University Press.

Haudricourt, André-George (1954). 'De l'origine des tons en vietnamien'. *Journal Asiatique*, , 242: 69–82.

Hayes, Bruce (1999). 'Phonetically-driven phonology: the role of Optimality Theory and grounding'. In: Darnell, Michael, Edith Moravscik, Michael Noonan, Frederick Newmeyer, & Kathleen Wheatly (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*, pp. 243–285. Amsterdam: John Benjamins. URL `http://www.linguistics.ucla.edu/people/hayes/phonet.htm`.

Heath, D. (1991). 'Tone in the Makaa associative construction'. In: *Tone in five langauges of Cameroon*. Arlington: SIL & University of Texas.

Hermans, Ben (1994). *The Composite Nature of Accent" With Case Studies from Limburgian and Serbo-Croation*. Ph.D. thesis, Free University of Amsterdam.

Hoffmann, Carl (1963). *A grammar of the Margi language*. London: Oxford University Press.

Hombert, Jean-Marie (1977). 'Development of tones from vowel height'. *Journal of Phonetics*, 5: 9–16.

——— (1978). 'Consonant types, vowel quality, and tone'. In: Fromkin, Victoria (ed.), *Tone. A Linguistic Survey*, pp. 77–111. New York: Academic Press.

Honda, K. (1983). 'Relationship between pitch control and vowel articulation'. In: Bless, D.M. & J.H. Abbs (eds.), *Vocal fold physiology*, pp. 286–299. San Diego, CA: College-Hill Press.

Honda, K., H. Hirai, & J. Dang (1994). 'A physiological model of speech production and the implication of tongue-larynx interaction'. In: *Proceedings of the Third International Conference on Spoken Language Processing*, pp. 175–178.

House, A. & G. Fairbanks (1953). 'The influence of consonant environment upon secondary acoustical characteristics of vowels'. *JASA*, 25: 105–113.

Hyman, Larry (1976). 'Phonologization'. In: Devine, A.M. & Laurence D. Stephens (eds.), *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*, pp. 407–418. Saratoga: Anma Libri.

——— (1985). *A Theory of Phonological Weight*. Dordrecht: Foris.

——— (1988). 'Syllable Structure Constraints on Tonal Contours'. *Linguistique Africaine*, 1: 49–60.

Hyman, Larry & R. Schuh (1974). 'Universals of tone rules: Evidence from West Africa.' *Linguistic Inquiry*, 5: 81–115.

Iivonen, A.K. (1987). 'Monophthonge des gehobenen Wienerdeutsch'. *Folia linguistica (Acta Societatis Linguisticae Europaeae)*, 21, 1–3: 293–336.

Jiang-King, Ping (1999). *Tone-vowel interaction in Optimality Theory.* Berlin: Lincom Europa.

Jongen, René (1972). *Phonologie der Moresneter Mundart; Eine Beschreibung der Phonologie der Moresneter Mundart. Eine Beschreibung der segmentalen und prosodischen Wortformdiakrise*. Assen: Van Gorcum.

Kaye, Jonathan, Jean Lowenstamm, & Jean-Roger Vergnaud (1985). 'The internal structure of phonological elements. A theory of charm and government'. *Phonology Yearbook*, 2: 305–328.

Kayne, Richard S. (1994). *The Antisymmetry of Syntax*. Cambridge: The MIT Press. Linguistic Inquiry Monograph Twenty-Five.

Kenstowicz, Michael (1994). *Phonology in Generative Grammar*. Oxford, UK: Blackwell.

Kirchner, Robert (1997). 'Contrastiveness and faithfulness.' *Phonology*, 14: 83–111.

de Lacy, Paul (1999). 'Tone and Prominence'. Manuscript, Rutgers.

Lanham, L.W. (1958). 'The tonemes of Xhosa'. *African Studies*, 17: 65–81.

Leben, William (1973). *Suprasegmental phonology*. Ph.D. thesis, MIT, Cambridge, MA.

Maddieson, Ian (1976). 'The intrinsic pitch of vowels and tone in Foochow'. *UCLA Working Papers in Linguistics*, 33: 191–202.

Matisoff, James (1973). *The grammar of Lahu*. University of California Press.

Moreton, Elliott (2006). 'Underphonologisation and modularity bias'. Manuscript; ROA 830.

Morev, L. N. (1983). *Shanskii jazyk*. Izdatelstvo Nauka.

Myers, James & Jane Tsay (2003). 'A Formal Functional Model of Tone'. *Language and Linguistics*, 4, 1: 105–138.

Newman, Paul (1975). 'The non-correlation of tone and vowel height in Hausa'. *Studies in African Linguistics*, 6: 207–213.

——— (1990). 'Hausa and the Chadic languages'. In: Comrie, Bernard (ed.), *The world's major languages*, pp. 705–724. New York: Oxford University Press.

Odden, David (1995). 'Tone: African languages'. In: Goldsmith, John (ed.), *Handbook of phonological theory*, pp. 444–475. Oxford, UK: Blackwell.

Ohala, John J. (1995). 'Phonetic explanations for sound patterns. Implications for grammars of competence'. In: Elenius, K. & P. Branderud (eds.), *Proceedings of the 13th International Congress of the Phonetic Sciences*, vol. 2, pp. 52–59.

van Oostendorp, Marc (2000). *Phonological Projection*. Berlin/New York: Mouton de Gruyter.

Peng, Long (1992). *A Unified Theory of Tone-Voice*. Ph.D. thesis, University of Arizona.

Piggott, Glyne (1992). 'Variability in feature dependency. The case of nasality'. *Natural Language and Linguistic Theory*, 10: 33–77.

Pilsczikowa-Chodak, Nina (1972). 'Tone-vowel height correlation and tone assignment in the patterns of verb and noun plurals in Hausa'. *Studies in African Linguistics*, 3: 399–421.

Ploch, Stefan (1999). *Nasals on my mind. The phonetic and the cognitive approach to the phonology of nasality.* Ph.D. thesis, School of Oriental and African Studies.

Plomp, R. (1967). 'Pitch of complex tones'. *JASA*, 41: 1526–1533.

Prince, Alan & Paul Smolensky (1993). 'Optimality Theory: Constraint Interaction in Generative Grammar'. Manuscript, Rutgers University and University of Colorado at Boulder. Available at ROA and published 2004 at Blackwell.

Prokosch, Eduard (1938). *A Comparative Germanic Grammar*. Baltimore: Linguistic Society of America.

Ritsma, Roelof J. (1967). 'Frequencies dominant in the perception of complex sounds'. *JASA*, 42, 191–198.

Ross, Malcolm (1993). 'Tonogenesis in the North Huan Gulf chain'. In: Edmondsen, J. & K. Gregerson (eds.), *Tonality in Austronesian Languages*, pp. 133–150. Honolulu: University of Hawaii Press.

Scheer, Tobias (2004). *A Lateral Theory of Phonology. Volume I: What Is CVCV, and Why Should It Be?* Berlin: Mouton de Gruyter.

Shryock, Aaron (1995). 'Investigating Laryngeal Contrasts. An Acoustic Study of the Consonants of Musey'. Working Papers in Phonetics 89, UCLA.

Snoxall, R.A. (1967). *Luganda-English Dictionary*. London: Oxford University Press.

Topintzi, Nina (2006). *Moraic Onsets*. Ph.D. thesis, University College London.

Tsay, Jane & James R. Sarwusch (1994). 'The interaction of F0 and vowel quality in toneal and non-tonal languages'. *Journal of the Acoustical Society of America*, 95, 5: 2922.

de Vaan, Michiel (2004). 'Wgm. *ī en *ū vóór *r* in Zuid-Limburg'. *Taal en Tongval*, 54, 2: 171–182.

Wang, William S.-Y. (1967). 'Phonological features of tone'. *International Journal of American Linguistics*, 33: 93–105.

Watkins, Laurel (1984). *A Grammar of Kiowa*. Lincoln: University of Nebraska Press.

Welmers, William (1962). 'The phonology of Kpelle'. *Journal of African Languages*, 1: 69–73.

Whalen, Douglas H. & Andrea G. Levitt (1995). 'The universality of intrinsic F0 of vowels'. *Journal of Phonetics*, 23: 349–366.

Williams, Edwin (1976). 'Underlying Tone in Margi and Igbo'. *Linguistic Inquiry*, 7: 463—484.

Wolff, Ekkehard (1993). *Referenzgrammatik des Hausa. Zur Begleitung des Fremdsprachenunterrichts und zur Einführung in das Selbststudium.* Munster: LIT.

Wright, Martha Susan (1983). *A Metrical Approach to Tone Sandhi in Chinese Dialects.* Ph.D. thesis, University of Massachusetts.

Yip, Moira (1980). *The Tonal Phonology of Chinese.* Ph.D. thesis, MIT, Cambridge, Mass.

——— (2002). *Tone.* Cambridge: Cambridge University Press.

Zhang, Jie (2002). *The effects of duration and sonority on contour tone distribution. Typological survey and formal analysis.* Routledge.

——— (2004). 'The role of contrast-specific and language-specific phonetics in contour tone distribution'. In: Hayes, Bruce, Robert Kirchner, & Donca Steriade (eds.), *Phonetically-Based Phonology.* Cambridge: Cambridge University Press.