Autosegmental Phonology

Marc van Oostendorp

September 13, 2005

Background

- In this class, we present an overview of some of the basic principles of autosegmental phonology. This theory proposes that the primitive elements of phonology (features) are not grouped together in unordered bundles (segments), but that they lead their own, independent lives.
- Phonological structure thus can be seen as a 'score' of individual instruments – roughly corresponding to articulatory organs – which play together along the same beat.
- Tone behaviour is a classical example of autosegmental theory in action.

1 The lowest level of linguistic organisation

Any person who did not study linguistics, but at the same time is able to use an alphabetic writing system, tends to divide words into individual vowels and consonants — together we call these *segments*.

If we think about this for a moment, this particular choice does not make a lot of sense. From a psycholinguistic point of view, the evidence for the existence of syllables is much stronger than the evidence for segments. It proves to be considerably easier to count the number of syllables in *encyclopedia* than the number of segments in *strings*. People who are not able to read or write will find the latter task even next to impossible, although they can perform the former one without a lot of problems.

From a phonological point of view, furthermore, segments are certainly not the ultimate primitives of our theory: we know that there are much smaller elements — features, which together somehow make up the words of natural language. It is an important question, what this organisation looks like. A very influential view of this is *autosegmental phonology*. According to

this theory, we can see the organisation of speech sounds in the human mind more or less like a musical score: every feature has its own part, which is to some extent independent of all other parts. Their only relation is that they are all attached to one central line, the *skeleton*, which keeps track of the time. The elements of the skeleton — which resemble the notion of a segment in certain ways — are usually depicted as x's (we will return to this in much more detail in a few weeks from now).

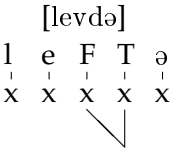
An important indication for the correctness of this view is assimilation. For instance, the Dutch plural past tense suffix has two allomorphs: -də and -tə. The former is chosen after a stem ending in a voiced segment, the latter if the stem ends in a voiceless obstruent:

(1) a. *lee[*vd*]en* 'lived'
b. *ma[*ft*]en* 'sleeped'

The question is: why would it be preferable to have a voiced segment next to another voiced segment? If all segments have their own independent dimension of voicing, there is no particular reason to expect this preference to hold (rather than, say, a preference for voiced obstruents to be chosen after coronal segments).

From an autosegmental point of view, on the other hand, the requirement makes sense. The structure of (1a) would be as in (2a), whereas (1b) would look like (2b):

(2) a.



b.

We have simplified the picture to a large extent; letters like l, e, f, d, ə, etc., are stand-ins for complex configurations of features, each on their own autosegmental tiers. Another aspect of this picture I want to draw your attention to, are the lines between segments and x-slots. These lines are called *association lines*. We can now generalise over these structures in the following way:

(3) Two consonantal x-slots need to share their association to [voice].

The only two configurations which are allowed by this constraint are the two which are actually attested in Dutch. There thus is *one* feature which is shared by *two* segments.

2 Tone

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:

(4)

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):

(5)

	Subject 'to'	Subject 'ma'
rɔr	tò rɔ̀r ìréɛ́	má rɔ́r ìréɛ́
	tò mò rɔ̀r ìréɛ́	má mó rɔ̀r ìréɛ́
	tò mà rɔ́r ìréɛ́	má má rɔ́r ìréɛ́
tom	tò tòm íréɛ́	má tóm íréɛ́
	tò mò tòm íréɛ́	má mó tòm íréɛ́
	tò mà tóm íréɛ́	má má tóm íréɛ́

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:

(6)

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:

(7) 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:

(8) 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*:

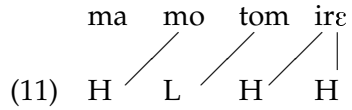
(9)

ma	mo	tom	irɛ
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:

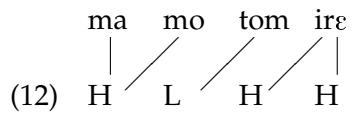
(10) WELLFORMEDNESS CONDITION (WFC): Every tone in the output representation should be linked to exactly one vowel, and vice versa.

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:



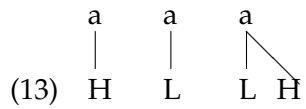
Every tone is now linked as much to the right as possible, without creating illicit 'contour' tones. Notice, however, that there is still one problem: the very first vowel (the one of the subject marker) does not bear a tone at all. There is no way we can solve this problem, given the requirements of 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:



3 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):



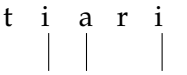
For Margi, the advice that autosegmental phonology gives us, turns out to be good advice. In the first place, this representation helps us to understand what is going on with tones. First, look at the following facts concerning the definite suffix *-ári*. The left-hand column represents the underlying shape of the stems to which this suffix is added (ˇa represents the rising tone):

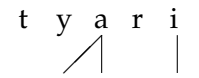
(14) a. sál sál-árì ‘man’
kùm kùm-árì ‘meat’
b. ʔímí ʔímy-árí ‘water’
kú kw-árì ‘goat’
c. tì ty-ǎrí ‘morning’
hù hw-ǎrì ‘grave’

(14a) 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.

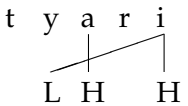
(14b) shows that if the stem ends in a high vowel with a high tone, this turns into a glide. Since glides, like all consonants, cannot carry their own tone, it looks as if the high tone disappears.

(14c) 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:

(15) a. Input:


b. Output:


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 *tyárǐ*):

(16)


The reason why we do not find this, is because there is a very hard constraint on autosegmental representations of the following kind:

(17) 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 (16), the realisation of the low tone will always precede that of the high tones.

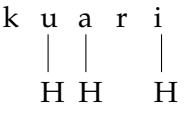
If we think about our representations in this way, it stands to reason that association of an element X to an element Y means that the realisation of X overlaps with that of Y in time. Thus the pronunciation of the low tone in (15a) will happen during the pronunciation of the /i/.

But given all of this, (16) 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).

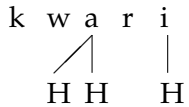
We can thus conclude that grammars can entertain all kinds of representations, including those which are not completely well-formed (because they display contour tones, or floating tones, or toneless vowels); but they will never entertain possibilities which do not make any sense at all.

Another remark to be made with respect to (15), 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:

(18) a. Input:



b. Output:



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:

(19)

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

Bibliography

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

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

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

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

Exercise 1

In an innovative variety of the dialect of Shanghai we find an interesting tonal pattern. Consider first the following underlying representations for several morphemes (tonal specifications are added in parentheses; M denotes a mid tone, a third tonal level in some languages):

- *çi* 'fresh' (HL); *wa* 'yellow' (LH); *du* 'big' (LH) ŋ 'dish' (LH); *ço* 'small' (MH)

Now consider the tones of following phrases:

- ço+ŋ → ço (M) ŋ (H) 'small fish'
- çi+ŋ → çi (H) ŋ (L) 'fresh fish'
- wa+ŋ → çi (L) ŋ (H) 'yellow fish'
- ço+wa+ŋ → ço (M) wa (H) ŋ (L) 'small yellow fish'
- çi+wa+ŋ → çi (H) wa (L) ŋ (L) 'fresh fish'

Describe what is going on here in autosegmental terms, and give the tonal pattern for 'big yellow fish'.