

# Notes on morphology and morphophonemic analysis

## 1 Morphemes and phonemes

Morphemes are not the same as phonemes. A phoneme is the smallest unit that *distinguishes* meaning, a morpheme is the smallest unit that *has* a meaning.

|                   | <i>betting</i> | <i>bedding</i> |
|-------------------|----------------|----------------|
| <b>Allophones</b> | [bɛɾɪŋ]        | [bɛɾɪŋ]        |
| <b>Phonemes</b>   | /bɛtɪŋ/        | /bɛdɪŋ/        |
| <b>Morphemes</b>  | /bɛt/ + /ɪŋ/   | /bɛd/ + /ɪŋ/   |

## 2 Notation

Morphemes are composed of phonemes, hence they are often expressed between slashes /bɛd/. The **boundary** between morphemes is often represented with plus sign +, although a hyphen - is also used.

| <i>cat</i> | <i>betting</i> | <i>unidentifiability</i>          |
|------------|----------------|-----------------------------------|
| /kæt/      | /bɛt + ɪŋ/     | /ʌn + aɪdɛnt + ɪfai + ʌbəl + ɪti/ |
| /kæt/      | /bɛt-ɪŋ/       | /ʌn-aɪdɛnt-ɪfai-ʌbəl-ɪti/         |

Morpheme boundaries sometimes coincide with **word** boundaries, but not always. Word boundaries are often indicated with a hash sign #.

| <i>cat</i> | <i>betting</i> | <i>unidentifiability</i>            |
|------------|----------------|-------------------------------------|
| /#kæt#/    | /#bɛt + ɪŋ#/   | /#ʌn + aɪdɛnt + ɪfai + ʌbəl + ɪti#/ |

### 3 Types of morphemes

The basic distinction between morphemes is those which are **roots** versus everything else. The root is the central ‘core’ of the word. For a word like *cat*, the root is *cat*. For a word like *catty*, the root is also *cat*. For a word like *concatenation*, the root is *concatenate*.

Morphemes which attach to or otherwise modify the root are called **affixes**. Affixes can come before, after, or inbetween other morphemes, in which case they may be referred to as **prefixes**, **suffixes** and **infixes**, respectively. The *-ty* in *catty* and the *-tion* in *concatenation* are both suffixes; the *un-* in *unidentifiability* is a prefix. Infixes are affixes that are inserted into a root, such as nominalising *-(a)m* in Khmer:

| Verb/Adj            |          | N         |                    |
|---------------------|----------|-----------|--------------------|
| [kcəi]              | ‘unripe’ | [kamcəi]  | ‘something unripe’ |
| [k <sup>h</sup> ɛŋ] | ‘angry’  | [kamhɛŋ]  | ‘anger’            |
| [ʔaoj]              | ‘give’   | [ʔamnaoj] | ‘gift’             |

### 4 Morphological functions

Morphology can largely be classified as either **derivational** (word-building) or **inflectional** (grammatical). Derivational morphemes are things like *-able*, *-tion* or *-ity* in English, *-keit* in German, *-ov(ý)* in Czech, etc.

Inflectional morphology modifies a root to make it grammatically appropriate for the context. Tense, aspect, and mood of verbs, number and person agreement in nouns, and case in pronouns are examples of inflectional morphology.

Many languages, such as Mandarin, lack any kind of overt inflectional morphology; the verb 给 [gě] ‘give’ in 我给了他一本书 ‘I gave him a book’ is the same as in 他给了我一本书 ‘He gave me a book’. English has slightly more inflectional morphology, but it is still a relatively impoverished system compared to many other languages. Compare for instance the various English and Czech forms of the verb ‘to read’ (Czech forms given in standard orthography):

|                 | English         | Czech |                 | English     | Czech  |
|-----------------|-----------------|-------|-----------------|-------------|--------|
| <i>1sg Pres</i> | I read          | čtu   | <i>3pl Pres</i> | we read     | čteme  |
| <i>2sg Pres</i> | you read        | čteš  | <i>Inf</i>      | to read     | číst   |
| <i>3sg Pres</i> | he/she it reads | čte   | <i>3sg Imp</i>  | read!       | čti!   |
| <i>1pl Pres</i> | we read         | čteme | <i>2pl Imp</i>  | read!       | čtěte! |
| <i>2pl Pres</i> | you (all) read  | čtete | <i>1pl Imp</i>  | let's read! | čtěme! |

In Czech, the form of the verb changes in each case, overtly encoding information about person and number.

## 5 Morphological analysis

The first step in a morphophonological analysis is often to figure out what the morphemes are. The basic approach is to examine the data and look for sequences of phonemes which remain the same when two elements of meaning remain the same. Consider the following data from Swahili:

|             |                     |           |                     |
|-------------|---------------------|-----------|---------------------|
| atanipenda  | ‘he will like me’   | atanipiga | ‘he will beat me’   |
| atakupenda  | ‘he will like you’  | atakupiga | ‘he will beat you’  |
| atampenda   | ‘he will like him’  | atampiga  | ‘he will beat him’  |
| atatupenda  | ‘he will like us’   | ananipiga | ‘he is beating me’  |
| atawapenda  | ‘he will like them’ | anakupiga | ‘he is beating you’ |
| nitakupenda | ‘I will like you’   | anampiga  | ‘he is beating him’ |
| nitampenda  | ‘I will like him’   | amenipiga | ‘he has beaten me’  |
| utanipenda  | ‘you will like me’  | amekupiga | ‘he has beaten you’ |
| utampenda   | ‘you will like him’ | amempiga  | ‘he has beaten him’ |

Table 1: Swahili verbs.

The first column are all forms of the verb ‘like’, the second column all forms of the verb ‘beat’. Since the form *penda* is shared by all the words in the first column and *piga* by all the forms in the second, we can probably conclude that *penda* means ‘like’ and *piga* means ‘beat’.

The next step is to isolate morphological minimal pairs, which are much like phonological minimal pairs, except instead of finding pairs that differ in only one

phoneme, we are looking for forms that differ in terms of only one morpheme. For example, the first two forms in the data *atanipenda* ‘he will like me’ and *atakupenda* ‘he will like you’ form a minimal pair allowing us to identify the 3sg-object and 2sg-object morphemes:

a t a **n i** p e n d a  
a t a **k u** p e n d a

The remaining morphemes (for 3sg subject, 2sg subject, 1sg object, etc.) can be discovered in a similar fashion.

## 6 Morphophonemic analysis: Finnish

Morphophonemic analysis proceed from the assumption (or the desire to assume) that morphemes are stored in the lexicon in invariant, underlying forms (URs), which are then converted into surface forms (SRs) by the application of phonological rules (or constraints, in some theories). The analytic procedure is similar to that of phonemic analysis, except that instead of only consider the distribution of phones, morphophonemic analysis also considers the distribution of morphemes.

### Procedure for morphophonemic analysis

1. Make a provisional segmentation of words into morphemes
2. Find the allomorphs of each alternating morpheme
3. Within each allomorph, find the segment(s) that alternate
4. Set up the UR so that all allomorphs can be derived from a single UR by general phonological rule

We’ll illustrate these steps using some data on Finnish nouns and adjectives (Table 2). We’ll compare two forms of each verb, the nominative case (used to mark the subject) and the partitive (used to mark unknown amounts and irresultatives). The goal of this analysis is to account for a alternation that gives rise to different vowels in the pronunciation of two related forms.

| <i>Nom sg</i> | <i>Part sg</i> | <i>gloss</i> | <i>Nom sg</i> | <i>Part sg</i> | <i>gloss</i> |
|---------------|----------------|--------------|---------------|----------------|--------------|
| aamu          | aamua          | ‘morning’    | yoki          | yokea          | ‘river’      |
| hopea         | hopeaa         | ‘silver’     | kivi          | kiveæ          | ‘stone’      |
| katto         | kattoa         | ‘roof’       | muuri         | muuria         | ‘wall’       |
| kello         | kelloa         | ‘clock’      | naapuri       | naapuria       | ‘neighbor’   |
| kirya         | kiryaa         | ‘book’       | ovi           | ovea           | ‘door’       |
| kylmä         | kylmäæ         | ‘cold’       | tukki         | tukkia         | ‘log’        |
| koulu         | koulua         | ‘school’     | æiti          | æitia          | ‘mother’     |
| lintu         | lintua         | ‘bird’       | yærvi         | yærveæ         | ‘lake’       |
| hylly         | hyllyæ         | ‘shelf’      | koski         | koskea         | ‘waterfall’  |
| kømpelø       | kømpeløæ       | ‘clumsy’     | reki          | rekeæ          | ‘sledge’     |
| kaikki        | kaikkea        | ‘all’        | væki          | vækeæ          | ‘people’     |

Table 2: Finnish verbs.

A bit of background information. Although there aren’t any minimal pairs in the data above, there is no evidence on which to conclude that any of the vowels are allophones of any of the others. We will thus assume that Finnish has 8 phonemic vowels /i y o ø æ u a e/. The first step is to try and determine what the relevant morphemes are. In this case, we’re interested in the realisation of the nominative and partitive singular morphemes.

If we consider the first 6 pairs in the first column, we can see that the difference appears to be the presence versus absence of a suffix *-a* in the partitive. This suggests that that partitive affix is *-a* and the nominative ‘suffix’ is *-∅*, or in other words, the UR of the verb stem is the same as the nominative. This hypothesis is challenged when we examine pairs like [kylmä] ~ [kylmäæ] ‘shelf’ or [kømeplø] ~ [kømpeløæ] ‘clumsy’. However, it turns out that this alternation is predictable based on other properties of the stem, so it sometimes surfaces as *-a* and sometimes surfaces as *-æ*. For the purposes of this analysis, this factor is not relevant, so we will designate the **underlying representation** of the partitive as /-A/. Remember, this is an *abstract* representation – there is no surface allophone [A], but that doesn’t matter. We just want a way to represent that these two alternants are related by rule.

However, there is another alternation that emerges between the stems that can be seen in pairs like [kaikki] ~ [kaikkea] ‘all’, [ovi] ~ [ovea] ‘door’, and [kivi] ~ [kiveæ] ‘stone’: in some pairs, the final vowel of the stem alternates between *-i* and *-e*. That is to say: all of the pairs in the data set fall into one of two groups: one where the

nominative form ends in *-i* and the partitive form, minus the suffix (*-a* or *-æ*) ends in *-i* as well, and one where the nominative form ends in *-i* but the partitive form, minus the suffix, ends in *-e*. To account for this alternation, we want to posit a **single underlying form** for each verb stem and a **single phonological rule** that, when applied to each verb form, generates the correct output.

Just on the basis of this observation, we could start to formulate a hypothesis about the underlying representation of the stems. One possibility would be to suppose that the underlying forms of the stems are the same as the nominative forms, and there is a rule which changes *i* to *e* in some environment – say, when followed by a [-back] vowel. Let’s call this the *i-lowering* rule.

$$i \rightarrow e / \_\_ V [-back] \tag{1}$$

We can check whether this rule works by applying it to a number of putative underlying forms.

|                    |                  |                          |                      |
|--------------------|------------------|--------------------------|----------------------|
| UR                 | /#koulu#/<br>–   | /#muuri#/<br>–           | /#ovi#/<br>–         |
| <i>i</i> -lowering | –                | –                        | –                    |
| SR                 | [koulu]          | [muuri]                  | [ovi]                |
| UR                 | /#koulu+a#/<br>– | /#muuri+a#/<br>muure + a | /#ovi+a#/<br>ove + a |
| <i>i</i> -lowering | –                | muure + a                | ove + a              |
| SR                 | [koulua]         | *[muurea]                | [ovea]               |

Table 3: Derivations of surface representations using rule (1).

The first row shows three nominative forms and the SRs that result from the application of the *i*-lowering rule (1). Since the structural description of the rule is not met in any case, the rule does not apply, and the SR is the same as the UR. So far, so good. However, when we apply the rule to the partitive forms in the second row, we correctly generate forms like *ovea*, but incorrectly generate forms like *muurea* (which should be *muuria*, with no vowel change). So the *i*-lowering rule, as formulated in (1), is insufficient.

We could go on like this, positing rules and checking them against all the forms in the data set (or, perhaps, write a short computer program to perform this tedious task on our behalf). However, this is a somewhat haphazard approach that could

potentially take a great deal of time. In this case, we can quickly ‘zero in’ on the relevant rule by observing some general properties of the data set. Recall that we were told Finnish has eight vowels /i y o ø æ u a e/. Looking at the data, we find that all of these vowels occur at the ends of words *except* [e]. Since the alternation in question is between *e* and *i*, and *e* never appears word-finally, we can formulate a rule that reflects this observation. We’ll call this the *e-raising* rule.

$$e \rightarrow i / \_ \# \tag{2}$$

If we assume that the UR of the stem has the vowel that surfaces in the partitive forms, we can use rule (2) to correctly derive all the forms in the data.

|                  |             |             |           |
|------------------|-------------|-------------|-----------|
| UR               | /#koulu#/   | /#muuri#/   | /#ove#/   |
| <i>e-raising</i> | –           | –           | ovi       |
| SR               | [koulu]     | [muuri]     | [ovi]     |
| UR               | /#koulu+a#/ | /#muuri+a#/ | /#ove+a#/ |
| <i>e-raising</i> | –           | –           | –         |
| SR               | [koulua]    | [muuria]    | [ovea]    |

Table 4: Derivations of surface representations using rule (2).

On this formulation, the UR is generally reflective of the surface form; the rule is necessary just to correctly derive the nominative forms of verbs that have an *e*-final stem in the partitive. Like the Polish devoicing rule, this rule is **neutralising** since the phonemic distinction between /e/ and /i/ is neutralised in word-final position: an underlying /e/ becomes a surface [i].

Note also that this provides another example of why it can be dangerous or at least misleading to assume that the ‘most basic’ or ‘simplest’ surface form of a word based on morphological criteria (such as being the nominative case or the 1st person singular) is straightforwardly related to the underlying representation.

## 7 Morphophonemic analysis: Kerewe

Consider the following verbal data from Kerewe, a Niger-Congo (Bantoid) language spoken in Tanzania.

| <i>infinitive</i> | <i>1sg habitual</i> | <i>3sg habitual</i> | <i>imperative</i> |                          |
|-------------------|---------------------|---------------------|-------------------|--------------------------|
| kupaamba          | mpaamba             | apaamba             | paamba            | ‘adorn’                  |
| kupaanga          | mpaanga             | apaanga             | paanga            | ‘line up’                |
| kupima            | mpima               | apima               | pima              | ‘measure’                |
| kupuupa           | mpuupa              | apuupa              | puupa             | ‘be light’               |
| kupekeča          | mpekeča             | apekeča             | pekeča            | ‘make fire with a stick’ |
| kupiinda          | mpiinda             | apiinda             | piinda            | ‘be bent’                |
| kuhiiga           | mpiiga              | ahiiga              | hiiga             | ‘hunt’                   |
| kuheeka           | mpeeka              | aheeka              | heeka             | ‘carry’                  |
| kuhaanga          | mpaanga             | ahaanga             | haanga            | ‘create’                 |
| kuheeba           | mpeeba              | aheeba              | heeba             | ‘guide’                  |
| kuhiima           | mpiima              | ahiima              | hiima             | ‘gasp’                   |
| kuhuuha           | mpuuha              | ahuuha              | huuha             | ‘breathe into’           |

Table 5: Kerewe verbs.

Posit an underlying representation for each verb stem and formulate a single rule that correctly derives all of the surface forms.