1 The metatheory of distributional analysis in Harris

- Recall the fruitful assumption Alice Kober made in her work on Linear B.

Kober assumed that the utterances of a natural language comprise sequences of basic elements which combine in broad patterns that are, for the most part, regular.

Thus she expected the distribution of basic elements to be stateable simply and succinctly, in terms of very general categories of those elements.

To put it another way, she aimed to isolate elements that distribute as freely as possible.

Based on this expectation, Kober speculated that the irregular distribution of ‘third graphs’ resulted from a distortion caused by the writing system, not from irregularity in the language itself. This speculation proved to be correct.

- Kober’s approach is common scientific methodology.

The basic goal of (most) scientific theory is to find a small set of elemental object types whose interactions are determined by a small set of simple rules.

We (generally) assume that a more economical set of rules is a diagnostic of a more correct set of proposed element types.
• Structuralist linguists like Zellig Harris were very explicit in their deployment of this methodology.

On phonemes:

The phonemes ... have been so selected that as many as possible of their combinations should occur in one utterance or another. (pg. 156)

On morphemes

[W]e determine our elements in such a way that it will be possible to make simple and compact statements about their distribution. (pg. 179)

• Another important determinant of Harris’s methodology was the environment of Empiricism in American science at the time, and in particular Behaviorism.

It was widely felt that science should not deal in objects that cannot be ‘directly’ measured and quantified.

This was taken to mean that psychologist should speak not of mental states, but of correlations between stimuli and behavioral responses.

And it was taken to mean that linguists should speak not of meanings, but perhaps of correlations between utterances and “features of social situations” (see Harris 12.41).

2 H’s understanding of ‘morpheme analysis’

• The usual understanding of morpheme is roughly this:

Morpheme A minimal pairing of sound and meaning.

Suppose that S is a sequence of phonemes, M is a meaning, and S denotes M. Then the pair (S,M) is a morpheme iff S cannot be divided into parts that denote parts of M’s meaning.
Examples

Consider the word *catastrophe*.
We could divide its pronunciation into pieces that are themselves words: at least *cat* and *strophe*. Yet the meanings of these words are clearly no part of the meaning of *catastrophe*. A catastrophe is a disastrous event, not a feline segment of music.

Therefore the pair ⟨*catastrophe*, “disastrous event”) is a minimal pairing of sound and meaning: a morpheme.

Now consider the word *reiterations*.
Suppose we divide it into the pieces: *re-*, *iterat*, *ion*, *s*. These pieces can be assigned meanings which jointly constitute the meaning of *reiterations*. Something like this: *re-* ‘again’; *iterat-* ‘to list’; *ion* ‘event of’; *s* ‘[plural]’; *reiterations* ‘plural events of listing again’.

Therefore *reiterations* is not monomorphemic. It is composed of (at least) the morphemes: *re-*, *iterat-*, *-ion* and *-s*.

- Harris aims to develop a procedure for discovering the basic elements of a language that is as *mechanical* as possible.

  The procedure should rely as much as possible on objectively measurable quantities, and not at all on assumed knowledge of the language.

- Consequently, Harris talks not about pairing phoneme sequences with meanings, but more generally about finding restrictions on sequences of phonemes (pg. 157).

- He observes that phonology provides some of these restrictions, but not all. There are restrictions on the distribution of phonemes in environments larger than the syllable.¹

  This is true of phonemes singly …

  *Example*
  
  English phonology permits the phoneme /l/ to appear in the environments /sta.p — aying/: *Stop lying.*

¹Phonology establishes (at least) the phonemes of the language, and the possible syllables of the language.
But /l/ cannot appear in the environment /stap —aying mi/:  
*Stop lying me* is not a sentence of English.

This restriction is evidently not entailed by any phonological rules.

... and in sequences:

<table>
<thead>
<tr>
<th>Environment</th>
<th>Phoneme sequence</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Her eyes were</em></td>
<td><strong>welling</strong></td>
<td>Possible</td>
</tr>
<tr>
<td></td>
<td><strong>incredible</strong></td>
<td>Possible</td>
</tr>
<tr>
<td></td>
<td><strong>ngiwell</strong></td>
<td>Ruled out by the phonology</td>
</tr>
<tr>
<td></td>
<td><strong>inkwell</strong></td>
<td>Ruled out <em>here</em> by ...</td>
</tr>
<tr>
<td><em>Were —s available</em></td>
<td><strong>inkwell</strong></td>
<td>Possible</td>
</tr>
<tr>
<td></td>
<td><strong>ngiwell</strong></td>
<td>Ruled out by the phonology</td>
</tr>
<tr>
<td></td>
<td><strong>welling</strong></td>
<td>Ruled out <em>here</em> by ...</td>
</tr>
<tr>
<td></td>
<td><strong>incredible</strong></td>
<td>Ruled out <em>here</em> by ...</td>
</tr>
</tbody>
</table>

- Evidently there must be a body of rules, distinct from phonology, which governs the distribution of phoneme *sequences* in arbitrary environments.

- The objects governed by these rules, Harris will call *morphemic segments*. The smallest morphemic segments are single morphemes.

### 3 Harris’s procedure for discovering morphemic segments

In the following subsections, we will discuss the diagnostics by which Harris judges if a phoneme sequence is (or pronounces) a morpheme segment, in a certain environment.

Again, the fundamental principle is this:

[W]e determine our elements in such a way that it will be possible to make simple and compact statements about their distribution. (pg. 179)

So the guiding question is: What mechanical procedures will lead to maximally simple statements of distribution?
3.1 Independence

- One condition Harris apparently considers necessary for morphemehood is independence (see pg. 157).

**Independence** $X$ is independent in $E$ if there is some $Y$ distinct from $X$ (and possibly $\epsilon$) that can replace $X$ in $E$.

**Example**
What independent phoneme sequences can we find in the utterance:

*They reconstructed the arguments*

(a) **He** reconstructed the arguments. $\Rightarrow$ **they**
(b) They **reconstruct** the arguments. $\Rightarrow$ **ed**
(c) They reconstructed the argument **re**. $\Rightarrow$ **s**
(d) They **de**constructed the arguments. $\Rightarrow$ **re**
(e) **Jay** reconstructed the arguments. $\Rightarrow$ **th**
(f) They reconstructed the ar **boreums**. $\Rightarrow$ **gument**
(g) They reconstru **_ed** the arguments. $\Rightarrow$ **ct**

\[ \ldots \text{etc.} \]

**Exercise**
What independent sequences can you find in the utterance:

*There’s a bug in my soup.\]

- Obviously the independent sequences deduced in (e–g) of the Example above are not all morphemes.

- What further conditions must an independent phoneme sequence meet before it can be considered a morpheme?

3.2 Shared distribution

We set up as morphemic segments only those tentatively independent phonemic sequences which have distributional similarities with other tentatively independent morphemic sequences. (pg. 179)

**The condition of shared distribution** An independent sequence is a morphemic segment if it shares its distribution with a class of other independent sequences.
Examples
1. /kt/ in *They reconstr-ued the arguments*.
   /kt/ appears in a large set of environments:

   \[
   \{ \text{He phu— it, Doctor Le—er, nice a—, } \ldots \} 
   \]

   But probably no other sequence shares this set of environments, or any large subsets of it.
   Hence /kt/ probably isn’t a morphemic segment in *They recon-
   structed the arguments*.

2. /gyument/ in *They reconstructed the ar—s*.
   (Same point as in 1.)

3. /they/ in — *reconstructed the arguments*
   /they/ appears in a very large set of environments (see below),
   shared exactly by we, and substantially by a great many other
   phoneme sequences, such as those spelled: you, the girls, the strange
   people we met yesterday, ....

   \[
   \{ \text{— arrived, What did — say?, } \ldots \} 
   \]

4. /ri/ in *They —constructed the arguments*
   (Discuss)

Problems
1. Is the distributional pattern of a sequence the set of environments in which
it is morphemic?
   If so, then there is a danger of infinite regress.

2. Is the distributional pattern the set of environments in which it is inde-
pendent (though not necessarily morphemic)?
   If so, the patterns will be very ‘noisy’; probably no two sequences will
   share their distribution.
3. How similar must the distribution of $X$ be to that of other sequences for $X$ to be considered morphemic?

4. With how many other sequences must $X$ share a similar distribution for $X$ to be considered morphemic?

5. How do we distinguish homonyms?

6. Is it adequate, in every language, to identify the distribution of candidate morpheme purely in terms of phoneme sequences?

### 3.3 Coherence of environment

Given a tentatively independent sequence of phonemes $A$ in a particular total environment, we seek some distributional feature which correlates with the distribution of this phoneme sequence; i.e. we ask what other utterance position, or the neighborhood of what other tentatively independent phoneme sequence characterizes all the sequences $M, N$ which occur with (before, etc.) our given sequence $A$. If we find such, we define our given phoneme sequence $A$, in the environments in which we have considered it, as a morphemic segment. (pg. 161)

**The condition of environmental coherence** An independent sequence $X$ is a morphemic segment in $E$ if the set of environments in which $X$ occurs (including $E$) is coherent.

**Environmental coherence** A set of environments is coherent if its members share some significant aspect of their distributions, thus allowing compact description of the set.

**Examples**

1. How would we decide that the ‘adverbalizing’ /li/ is a morpheme in *quickly*?

   Presumably not because /li/ shares its distribution with many other phoneme sequences. It’s distribution is evidently uncommon:

   $$\{ \text{ he quick—arrived, shocking—sudden, \ldots } \}$$
Yet the set of environments in which /li/ is coherent. /li/ always follows an adjective, and the resulting “adj+/li/” combination occurs where adverbs can occur.

It remains, then, to define the categories of adjective and adverb distributionally!

2. How would we decide that /n/ is not a morphemic segment in “neat”?

There are many phoneme sequences that can replace /n/ in /niyt/: /b/, /p/, /m/ /f/, /w/, /t/, /sk/. Many of these can also replace /n/ elsewhere: —it, —ill, —at, .... So we might get the wrong impression that /n/ is a morpheme in /niyt/, by virtue of the condition of Shared Distribution.

However it is quite clear that the set of environments in which /n/ occurs is not coherent. The sequences it, ill, at, etc., share no significant aspect of their distributions. They are not broadly substitutable for each other.

Problems

1. What is the relevant set of environments? The set where the sequence in questions is a morpheme segment, or where it is independent?
   If the first, then there is a danger of infinite regress. If the second, then the data will be noisy.

2. To what extent must the members of a set share their distributions for that set to be considered coherent?

3. For the purposes of deciding whether a set of environments is coherent, is it always adequate to identify those environments just as sequences of phonemes?
3.4 ‘Leftover’ morphemes

We want to be able describe a stretch of speech exhaustively as a sequence of morphemes. (pg. 177)

The exhaustivity condition For an utterance $XE$, if $E$ is a sequence of morphemes, then necessarily $X$ must be a sequence of morphemes.

- Given the exhaustivity condition, it is not strictly necessary that a phoneme sequence, to be considered morphemic, must satisfy the conditions of Independence or Shared Distribution.

Example:

Is *cran* a morpheme *cranberry*?

The sequence *cran* is independent in *cranberry*, witness: *blueberry*, *gooseberry*, etc.

Yet *cran* occurs nowhere else. So we cannot conclude that it is a morpheme based on any pattern in its distribution.

Yet Harris argues (pg. 177) that *cran* is nevertheless morphemic, by virtue the Exhaustivity Condition. This follows from his conclusion that *berry* is a morpheme.

Question: How can Harris conclude that *berry* is morphemic in *cranberry*?

Question: Why might somebody want to insist that *berry* is morphemic in *cranberry*?

3.5 Meaning cannot be used in discovering morpheme segments

- Harris apparently assumes that the meaning of a morphemic segment can—in principle—be defined by correlating its use with features of the social situations in which it is used.
• He then concludes that:
  
The correlation of morphemic segments with features of social situations cannot be used in establishing [morphemic] segments [...] because there is at present not way of determining meaning differences as exactly as one can measure sound differences. (pg. 173)

• And he admits that, as a result, his procedure may yield morphemic segments different from those produced by a meaning-based analysis.

Discuss Why would the meaning-based and non-meaning based procedures ever yield the same results? Why would they ever differ?

4 Discontinuous morphemes

• Sometimes we find independent phoneme sequences with clearly describable distributions that are discontinuous. That is, there is at least one ‘break’ in the phoneme sequence, where distinct elements intervene.

We will call these discontinuous morphemes.

Examples

1. Arabic tri-consonantal roots

<table>
<thead>
<tr>
<th>kataba</th>
<th>‘he wrote’</th>
<th>kadaba</th>
<th>‘he lied’</th>
</tr>
</thead>
<tbody>
<tr>
<td>katabtu</td>
<td>‘I wrote’</td>
<td>kadabtu</td>
<td>‘I lied’</td>
</tr>
<tr>
<td>ka:taba</td>
<td>‘he corresponded’</td>
<td>kadabtu</td>
<td>‘I corresponded’</td>
</tr>
<tr>
<td>ka:tabtu</td>
<td>‘I corresponded’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this pattern, we note at least that:
(a) $k...t...b$ is independent in *kataba, katabtu, ka:taba* and *ka:tabtu*,
(b) $k...d...b$ is independent in *kadaba* and *kadabtu*.

Based on the meanings given, we might also conclude that $k - t - b$ corresponds somehow to the meaning ‘write’, and $k - d - b$ to the meaning ‘lie’.
2. Latin agreement

<table>
<thead>
<tr>
<th>Latin</th>
<th>English</th>
<th>Latin</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>filius bonus</td>
<td>‘good son’</td>
<td>taurus malus</td>
<td>‘bad bull’</td>
</tr>
<tr>
<td>filius bon</td>
<td></td>
<td>taurus mal</td>
<td></td>
</tr>
<tr>
<td>fili bonus</td>
<td></td>
<td>taur malus</td>
<td></td>
</tr>
<tr>
<td>filia bona</td>
<td>‘good daughter’</td>
<td>libra mala</td>
<td>‘bad scale’</td>
</tr>
<tr>
<td>filia bon</td>
<td></td>
<td>libra mal</td>
<td></td>
</tr>
<tr>
<td>fili bona</td>
<td></td>
<td>libr mala</td>
<td></td>
</tr>
</tbody>
</table>

In this pattern we note at least that:
(a) *us*. . . *us* is independent in *filius bonus* and *taurus malus*.
(b) Neither of the two *us’s* is independent of the other. (That is, *us* is independent neither in the environment *fili—bonus* nor in the environment *filius bon*—.
(c) *a*. . . *a* is independent in *filia bona* and *libra mala*.
(d) Neither of the two *a’s* is independent of the other.
Based on similar examples, we would conclude that these discontinuous sequences have very coherent distributions, and are hence good morphemes.

**Total dependency** When *X* cannot be replaced without replacing *Y*, then there is a **total dependency** between *X* and *Y*.

**Partial dependency** When replacement of *X* restricts the range of possible replacements for *Y*, then there is a **(partial) dependency** between *X* and *Y*.

**Discontinuous dependency** Whenever *X* is (at least partially) dependent on *Y*, and *X* and *Y* are not adjacent, we will speak of a **discontinuous dependency** between *X* and *Y*.

**Exercises:** Isolate any (discontinuous) dependencies in the following sentences:

1. He runs daily.

2. Both boys are laughing.
5 Discussion:
Is surface phonological distribution enough?

- If morphemes are defined by distribution, we cannot speak of one morpheme having two allomorphs with distinct distributions. Conversely, every distinct pattern of distribution should define a separate morpheme.

Problems

- Isn’t it the same have in both You have eaten and Have you eaten? ?
- Stressed and unstressed pronouns distribute differently in many languages (e.g., the Germanic languages). But are they utterly distinct morphemes?
- Sometimes the distribution of a morpheme (class) is difficult or impossible to define just by describing sequences of phonemes. That is, it may be difficult to identify certain syntactic or morphological environments in purely phonological terms. In such cases, Harris’ method will not (easily) discover the relevant morpheme (class). Consider the example from German below.

*German finite main clause verbs*

(1) Die Helga hat gestern dem Hans auf die Nase geküsst

the(NM) Helga has yesterday the(DT) Hans on the nose kissed

‘Helga kissed Hans on the nose yesterday.’

(2) Gestern hat die Helga dem Hans auf die Nase geküsst

yesterday has the(NM) Helga the(DT) Hans on the nose kissed

‘Yesterday, Helga kissed Hans on the nose.’

(3) Auf die Nase hat die Helga dem Hans gestern geküsst

on the nose has the(NM) Helga the(DT) Hans yesterday kissed

‘On the nose, Helga kissed Hans yesterday.

(4) Dem Hans hat die Helga gestern auf die Nase geküsst

the(DT) Hans has the(NM) Helga yesterday on the nose kissed

‘Hans, Helga kissed on the nose yesterday morning.’

(5) Geküsst hat sie ihm nicht

kissed has she him not

‘Kiss him, she didn’t.’