Morphological complexity in word recognition: Foundational issues

Robert Fiorentino

Neurolinguistics & Language Processing Lab
Department of Linguistics
University of Kansas
Plan for the four talks

Talk 1: Morphological complexity in word recognition: Foundational issues

Talk 2: Decomposing morphologically complex words: Across contexts and word types I

Talk 3: Decomposing morphologically complex words: Across contexts and word types II

Talk 4: Compositional aspects of complex word processing
Morphological complexity in word recognition: Foundational issues

Today’s plan

Talk 1: Morphological complexity in word recognition: Foundational issues

Basic questions and introduction to processing approaches to morphological complexity
Identifying the pieces, processes, and neural bases of language
Identifying the pieces, processes, and neural bases of language

Basic units

Mary likes hockey
Identifying the pieces, processes, and neural bases of language

*Internal structure*

NP

Mary

VP

V

likes

NP

hockey
Identifying the pieces, processes, and neural bases of language

*Internal structure*

```
S
  VP
    V likes
    NP hockey
NP Mary
```
Identifying the pieces, processes, and neural bases of complex word recognition

*Basic units??*

teacup

tea
cup
Identifying the pieces, processes, and neural bases of **complex word recognition**

*Internal structure?*

- **Teacup**
  - **Tea**
  - **Cup**
Morphological complexity in word recognition: Foundational issues

Desiderata: Learn/say something about language with respect to:

- Representations
- Mechanisms
- Neuroanatomy
Morphological complexity in word recognition: Foundational issues

Desiderata: Learn/say something about language with respect to:

- Constituency/basic units
- Structure building
- Composition of meanings
- Organization of language in the brain
Morphological complexity in word recognition: Foundational issues

Desiderata: Learn/say something about cognition with respect to:

- **Representations** (e.g. nature of abstraction/internal representation in mental computation)
- **Mechanisms** (e.g., binding; predictive processing)
- **Neuroanatomy** (e.g., moving closer toward finding answers to the why questions regarding brain structure and function)
Morphological complexity in word recognition: Foundational issues

Do morphemes matter in the processing of putatively complex words?

While many linguistic approaches to the lexicon recognize the morpheme as a basic-level representational unit, whether and to what extent the mind/brain makes recourse to a morpheme-based route to word recognition has remained controversial.
Morphological complexity in word recognition: Foundational issues

*Idiosyncratic* knowledge: consequences for representation and computation?

Irregular inflection e.g. “taught”

Limited productivity e.g. “humidity”

Semantic opacity e.g. “honeymoon”

Lexicalization e.g. “teacup” vs. “teadesk”
Morphological complexity in word recognition: Foundational issues

Three kinds of solution

- It’s *all* idiosyncratic (save similarity-based groupings) – atomic whole-word and not morpheme is basic-level representation (atomist/non-morphological models)

Morphemes are a “highly problematic theoretical construct” that should be done away with (Hay & Baayen 2005).
Morphological complexity in word recognition: Foundational issues

Three kinds of solution

- It’s all idiosyncratic (save similarity-based groupings) – atomic whole-word and not morpheme is basic-level representation (atomist/non-morphological models)

“…orthographic (and phonological) cues are directly mapped onto meanings, without the mediation of morphemes as a representational level” (Kuperman, 2013).
Morphological complexity in word recognition: Foundational issues

Three kinds of solution

- It’s *all* idiosyncratic (save similarity-based groupings) – atomic whole-word and not morpheme is basic-level representation (*atomist/non-morphological models*)

Various proposals in this line, such as that putative constituent relationships are learned associations among whole-word forms (e.g., Bybee, 1995)
Morphological complexity in word recognition: Foundational issues

Three kinds of solution

- It’s all idiosyncratic (save similarity-based groupings) – atomic whole-word and not morpheme is basic-level representation (atomist/non-morphological models)

...or by virtue of overlap in form (e.g., orthographic/phonological) and/or semantic features without recourse to abstract morphological representations (e.g., Seidenberg & Gonnerman, 2000)
Morphological complexity in word recognition: Foundational issues

Three kinds of solution

- Morpheme-based organization for the formally and semantically, etc. “regular” stuff; the rest is atomic (dual-route/supralexical models)

e.g. Pinker’s *Words and Rules* account

Regular words: “products of a rule”

Irregular words: “memorized and retrieved as individual words”
Morphological complexity in word recognition: Foundational issues

Three kinds of solution

- Morpheme-based organization for the formally and semantically, etc. “regular” stuff; the rest is atomic (dual-route/supralexical models)

“In the supralexical architecture, morphemic representations are located between whole word form and higher-level semantic representations. Thus, morphemic representations can only become active after they have been contacted by whole-word form representations.” (Diependaele, Sandra, & Grainger 2005)
Morphological complexity in word recognition: Foundational issues

Three kinds of solution

- Morpheme-based organization for complex words; idiosyncratic knowledge is about combinations (full decomposition/early decomposition models)

“We argue for a model of morphological complexity that has all complex words assembled by the grammar…” (Stockall & Marantz, 2006)
Morphological complexity in word recognition: Foundational issues

**Three kinds of solution**

- It’s all idiosyncratic (save similarity-based groupings) – atomic whole-word and not morpheme is basic-level representation (**atomist/non-morphological models**)

- Morpheme-based organization for the formally, semantically, etc. “regular” stuff; the rest is atomic (**dual-route/supralexical models**)

- Morpheme-based organization for complex words; idiosyncratic knowledge is about combinations (**full decomposition/early decomposition models**
Morphological complexity in word recognition: Foundational issues

**What is the right solution?** Trying to answer this question is going to lead us to engage a range of issues including:

- Processing across word types (inflection, derivation, and compounding)
- The impact of factors like frequency of occurrence, semantic transparency, productivity, regularity/allomorphy, lexical status, etc.
- Alternative methods: how they work and what they tell us
Psycholinguistic and Neurolinguistic Methods for Investigating Word Recognition in Isolation

Psycholinguistics:

• Priming (e.g., masked priming, overt priming, cross-modal priming, long-lag priming, transposed-letter priming)
• Lexical decision
• Eye movements

Neurolinguistics:

• Neuroimaging (e.g., EEG, MEG, fMRI, PET)
• Aphasiology
Morphological complexity in word recognition: Foundational issues

Reasoning about the basic nature of putatively complex word representations also (arguably) needs to be situated within the larger computational system of language.

This in turn raises difficult (but arguably useful to wrestle with) questions like:

What is the endgame of word recognition (and should this affect how we might expect word recognition to work)?

What kind of representations eventually make contact with other parts of the language system?
Morphological complexity in word recognition: Foundational issues

My approach (for the talks): engage these issues/discuss these methods using some of my own work as an ‘anchor’ each day, but introducing other relevant studies and the discussion of other methods as well

As most of my research has been on compounding, that word formation type will consistently come up, and we will start there today
Morphological complexity in word recognition: Foundational issues

What are we looking for when we look for “morphological processing” or “decomposition” in the mind/brain?

In my work, I’ve suggested we think of/look for three broad types of computation that a morpheme-based route to putatively complex word recognition entails:

- **Segmentation**
- **Activation**
- **Composition**
Morphological complexity in word recognition: Foundational issues

Unpacking complex words/Unpacking “decomposition”

- **Segmentation** mechanism: parsing sublexical string for morphemic candidates
- **Activation/access** to long-term memory representations of morphemes in potentially complex words
- **Composition** mechanism(s): construction of complex word representations
Morphological complexity in word recognition: Foundational issues

**Segmentation - Activation - Composition** of complex word representations

Tracking each of these kinds of computation in order to better understanding when and how morpheme-based processing unfolds will require a cross-method approach

Let’s start with the basic questions of “Is there a game in town? Is there any reason to think that complex word processing involves these operations on morphemes at all?”
Morphological complexity in word recognition: Foundational issues

**Segmentation - Activation - Composition** of complex word representations

Begin by examining evidence regarding the activation of morphemic constituents of compounds
Test case: compounding

Why compounds?

- English compounds don’t wear their complexity on their sleeve (e.g., no salient, high frequency affix that can be stripped)
- English compounds show variations semantic transparency

*Are compounds processed via a morpheme-based route in the general case?*
Competing representational claims:

Are compounds (CW) represented as atoms, like Single Words (SW) or are they decomposed?

```
<table>
<thead>
<tr>
<th>CW</th>
<th>SW</th>
<th>vs.</th>
<th>CW</th>
<th>SW</th>
</tr>
</thead>
<tbody>
<tr>
<td>flagship</td>
<td>crescent</td>
<td></td>
<td>flagship</td>
<td>crescent</td>
</tr>
</tbody>
</table>
```

Competing processing predictions:

(When, if ever) are compounds recognized in terms of constituents online?
Use what you know. We know something about the properties affecting lexical access for “single words”:

**Frequency of occurrence:** words appearing more frequently are processed more quickly than words that appear less frequently (e.g. *cat* vs. *hog*)
Design: pairwise matching on whole-word properties

- The compounds (whole word) and single words have equal whole-word log frequency, letter-length, and syllabicity.

<table>
<thead>
<tr>
<th>In whole-word properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound = Single word</td>
</tr>
<tr>
<td>flagship = crescent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Compound</th>
<th>Single word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Freq.</td>
<td>.68</td>
<td>.69</td>
</tr>
<tr>
<td># Letters</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td># Syllables</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Design: advantage for constituents over whole words

- **However**, each sub-part of the compound has **higher** frequency, **shorter** length, and **fewer** syllables than the whole words.

In internal properties

<table>
<thead>
<tr>
<th>Compound</th>
<th>Single word</th>
</tr>
</thead>
<tbody>
<tr>
<td>flagship</td>
<td>crescent</td>
</tr>
</tbody>
</table>

| Log Freq. | 1.49 | 1.95 | .69 |
| # Letters | 4    | 4    | 8   |
| # Syllables | 1    | 1    | 2   |

Stimuli

60 Compounds and 60 pairwise matched Single Words

120 nonwords (including 16 word-nonword foils)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean Log. Freq.</th>
<th>Mean # Letters</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound Word (CW)</td>
<td>.455</td>
<td>7.82</td>
<td>flagship</td>
</tr>
<tr>
<td>CW 1st/2nd constituents</td>
<td>1.96/1.96</td>
<td>3.82/4.0</td>
<td>flag/ship</td>
</tr>
<tr>
<td>Single Word (SW)</td>
<td>.455</td>
<td>7.78</td>
<td>crescent</td>
</tr>
<tr>
<td>Nonword (NW)</td>
<td></td>
<td>7.81</td>
<td>nishpern</td>
</tr>
<tr>
<td>Word-Nonword Foil (WNW)</td>
<td></td>
<td>7.94</td>
<td>crowskep</td>
</tr>
</tbody>
</table>
Task: Lexical decision (word/nonword judgment)

Participants: N=12 Native English speakers
Direct comparison of whole-word vs. morpheme level

For compounds, which frequency, length, and number of syllables matters?

- Whole word? \textit{flagship}
- Sub-parts? \textit{flag} and \textit{ship}

It depends on the representational/processing model.
Predictions

If all compounds treated as **atoms**, then *no advantage from the parts*.

- **flagship** is just like **crescent**

If compounds are processed in decomposed form, there should be an advantage for short, high-frequency constituents

- **flagship** should be **faster** than **crescent**

If decomposition holds *across-the-board*, it should hold for both common and rare compound words

- significant effects across frequency bins
Time-course of compound vs. monomorphemic word processing

I. The whole-word perspective

II. The late decomposition perspective

III. The initial decomposition perspective
Time-course of compound vs. monomorphemic word processing

I. The *whole-word* perspective
Time-course of compound vs. monomorphemic word processing

I. The *whole-word* perspective:

- Similar processing profiles for compounds and single words; no access advantage due to constituents.

```
Whole-word form access
```

```
Initial activation at whole word level
```

```
No role for morphological constituents
```

```
Endpoint measure (RT)
```


Time-course of compound vs. monomorphemic word processing

II. The *late decomposition (supralexical)* perspective
Time-course of compound vs. monomorphemic word processing

II. The *late decomposition (supralexical)* perspective:
- Initial access to the lexicon is for whole-word chunks.
- Later, morphemes activated (for transparent words).

Initial activation at whole word level

Subsequent morpheme activation *at least some of the time*

- **Whole-word form access**
- **Decomposition**
- Endpoint measure (RT)
Time-course of compound vs. monomorphemic word processing

III. The *initial decomposition* perspective:

- Compounds are processed immediately in terms of their constituents; they have an advantage in access properties.
- Thus, if the constituents are activated early, we should detect differences in processing due to constituents.
Compounds responded to more quickly than single words, contra to atomist account.
This distinction evident across high, medium, and low-frequency subsets of matched compounds and single words
Adding a brain-level measurement

Drawing inferences regarding initial lexical access from response data alone is complicated by the very nature of the measurement.

What’s going on during all this time?
Adding a brain-level measurement

Drawing inferences regarding initial lexical access from response data alone is complicated by the very nature of the measurement.

**CW**

Fixation point: + 0ms teacup ~700ms Button push for lexical decision

Shorter initial access time?

**SW**

Fixation point: + 0ms station ~750ms Button push for lexical decision

Longer initial access time?
Adding a brain-level measurement

Drawing inferences regarding initial lexical access from response data alone is complicated by the very nature of the measurement.

<table>
<thead>
<tr>
<th>CW</th>
<th>Same initial access time</th>
<th>Shorter post-access time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixation point:</td>
<td>Button push for lexical decision</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>~700ms</td>
</tr>
<tr>
<td></td>
<td>teacup</td>
<td>Button push for lexical decision</td>
</tr>
<tr>
<td></td>
<td>0ms</td>
<td>~750ms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SW</th>
<th>Same initial access time</th>
<th>Longer post-access time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixation point:</td>
<td>Button push for lexical decision</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>~700ms</td>
</tr>
<tr>
<td></td>
<td>station</td>
<td>Button push for lexical decision</td>
</tr>
<tr>
<td></td>
<td>0ms</td>
<td>~750ms</td>
</tr>
</tbody>
</table>
Adding a brain-level measurement

Recording brain activity time-locked to the onset of the word via MEG allows access to several components in the signal *before the response*, that may map onto the subcomputations of lexical processing.
What is MEG?

- MEG (magnetoencephalography) is a non-invasive method of recording brain activity with **millisecond temporal resolution**
- Measures magnetic fields (100 million times smaller than those detected by compass) from underlying neural currents

Photograph courtesy of Kanazawa Institute of Technology, Japan
Components in the MEG Signal

M170: Visual Word Form processing; some evidence of morpho-orthographic processing (e.g., Solomyak & Marantz, 2010)

M250: Possibly reflects orthography/phonology conversion; role poorly understood

M350: Index of lexical activation (frequency, priming, etc.)

A. Magnetic Field Contours at 170ms, 250ms, and 350ms Following Visual Word Onset

B. Averaged Waveforms from All Left Hemisphere Sensors Following Visual Word Onset

Fixation:

0ms Word presentation: teacup

Button push for lexical decision ~700ms
Effects of frequency on M350 (Embick et al, 2001)

Fig. 2. Waveform illustrating M150, M250, and M350 components.

Fig. 4. Latency of M350 by frequency category (Mean and S.E.).

Table 1
Word-frequency categories

<table>
<thead>
<tr>
<th>Category</th>
<th>n/Million</th>
<th>Log frequency</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>700</td>
<td>2.8</td>
<td>Number</td>
</tr>
<tr>
<td>2</td>
<td>140</td>
<td>2.1</td>
<td>Ask</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>1.4</td>
<td>Wheel</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.7</td>
<td>Candle</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>Clam</td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
<td>-0.7</td>
<td>Snarl</td>
</tr>
</tbody>
</table>
Results: Earlier Peak Latency of M350 component for CW vs. SW
Fiorentino & Poeppel (2007): Conclusions

• Results support a decompositional view of compound processing

• Results do not support putative constraints on decomposition due to frequency, lexicalization, or length (cf. Bertram & Hyönen, 2003)

• Consistent with initial automatic decomposition in the general case, not initially constrained by experience-based constraints
Fiorentino & Poeppel (2007): Conclusions

The response time results have been replicated a number of times/converge with those of other studies/labs (e.g., Ji et al., 2011 among others)

As we will return to in later talks, these effects also hold:

- For semantically opaque compounds (Ji et al., 2011; Fiorentino, 2014)
- When the single words to which we compare the CW have pseudoembedded morphemes (e.g., platform)
- In experiments including large numbers of novel compounds (e.g., Fiorentino et al., 2014)
Moreover, the MEG findings are consistent with a host of studies implicating the activation of lexical content from long-term memory around 200-400 ms in left temporal lobe (roughly speaking)

e.g., MEG studies: Embick et al., 2001 (frequency), Sekiguchi et al., 2000 (long-lag repetition priming), Pylkkänen et al. 2006 & Beretta et al., 2006 (polysemy/homonymy distinctions), Solomyak & Marantz, 2009; for other morphological processing studies, see e.g., Stockall & Marantz (2006); Fruchter et al. (2013)
What do these findings suggest about segmentation?

Much previous evidence for rapid decomposition comes mainly from *affixation* (e.g., prefixed or suffixed words) ➔ Leaves open the possibility that rapid automatic decomposition may rely on *affix stripping*

- See, e.g., Taft & Forster (1975); Duñabeitia et al. (2007), among others

The data we have seen so far suggests that the presence of a salient affix is not necessary for decomposition
Morpheme-based route to complex word recognition

Segmentation - Activation - Composition

Not limited to affix stripping

M350; response times (with caution)

Constraints on morpheme combination? Assembly of meanings from parts?

Activation data allow inferences regarding segmentation

..and motivate questions regarding the (existence and) nature of post-decompositional processes
Fiorentino & Poeppel (2007): Conclusions

Response Time findings: can be thought of as a version of the “base frequency” effect (evidence of constituent frequency influencing processing when whole-word, “surface” frequency kept constant)

In this study, the constituent vs. whole-word comparison was done by comparing compounds to matched monomorphemic words
Fiorentino & Poeppel (2007): Conclusions

Other studies do manipulate constituent frequency and report response time effects:

Andrews (1986): effects for both constituents in lexical decision

…although Andrews ultimately concludes such effects may be post-decompositional since the effects for derived words, also tested in that study, depended on the presence of compounds in the experiment.
Several other studies report constituent frequency effects, although there is some inconsistency across studies with respect to which constituent (initial or final) shows such effects (e.g., Juhasz et al., 2003 among others) and whether first- and second-constituents influence processing on the same timecourse (e.g., as measured by ERP: Vergara-Martínez et al., 2009 or eye-movements, e.g., Pollatsek et al., 2000)
Base frequency effects: Some background

- Found across word type (e.g., derivation: Taft, 1979, inflection: Lehtonen et al., 2007)
- Effects may not be the same across whole-word frequency (e.g. Baayen et al., 2007)
- Effects of experimental context
- How to interpret base frequency effects on RT? (e.g., Taft, 2004)
Base frequency effects: Some background

- Found across word type (e.g., derivation: Taft, 1979, inflection: Lehtonen et al., 2007)

- Effects may not be the same across whole-word frequency (e.g. Baayen et al., 2007)

- Effects of experimental context

- How to interpret base frequency effects on RT? (e.g., Taft, 2004)
Base frequency effects: Some background

Some standard assumptions:

- Across-the-board, obligatory decomposition would seem to predict base-frequency effect (faster response times for words with high base frequency than low base frequency, when whole-word frequency is kept constant)

- Failure to find a base-frequency effect would seem to provide evidence for whole-word storage for the given word type
Base frequency effects: Some background

Consider the following example:

**seeming**  **mending** matched on surface frequency

**seem** has higher base frequency than **mend**

Finding an RT advantage for say **seeming** over **mending** would provide evidence for decomposition
Base frequency effects: Some background

Consider the following example:

`seeming`  `mending`  matched on surface frequency

`seem` has higher base frequency than `mend`

Does `failing` to find such an effect always indicate `non-decomposition`?
Base frequency effects: Some background

Consider the following example:

seeming  mending  matched on surface frequency

seem has higher base frequency than mend

Taft (2004) points out that benefits of base frequency can be obscured by late-emerging effects of morpheme re-combination (e.g. with their origin in the difficulty of combining each of these verbs with –ing)
Summary

- Today, I have highlighted some of the main ‘big-picture’ questions regarding the role of morphemes in word recognition, outlining major alternative conceptions

- Introduced a sample psycholinguistic/neurolinguistic study (Fiorentino & Poeppel, 2007) suggesting that morphemes do matter in complex word recognition, at least for the compounds tested in that study

Next: move on to some harder cases!
Thank you!

fiorentino@ku.edu

Neurolinguistics & Language Processing Lab:
www.neuroling.ku.edu