Gemination and Degemination in English Prefixation: Lexical Strata, Semantics, and Phonetic Evidence

Sonia Ben Hedia
Heinrich-Heine-Universität Düsseldorf

The 22nd Annual University of Texas at Arlington Student Conference in Linguistics & TESOL, 19th and 20th of February 2015
Morphological structure and phonetic detail

**Claim:**

- Phonetic detail entails information about morphological structure

**Morphological Segmentability Hypothesis (Hay 2003)**

- Strength of morphological boundary depends on ‘decomposability’ and is mirrored in phonetic detail

Morphological structure and phonetic detail

Decomposability:

<table>
<thead>
<tr>
<th>Highly decomposable</th>
<th>Less decomposable</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Semantically transparent</td>
<td>• Semantically opaque</td>
</tr>
<tr>
<td>• Phonologically transparent</td>
<td>• Phonologically opaque</td>
</tr>
<tr>
<td>• Easy to segment</td>
<td>• Difficult to segment</td>
</tr>
</tbody>
</table>

- *happiness* vs. *business*
- *discernment* vs. *government*
Morphological structure and phonetic detail

• Hay (2007): Gradient decomposability of affixed forms is reflected in phonetic detail
  • Decomposability measured in relative frequency (whole word frequency : base frequency)
  • The prefix *un* is pronounced longer when part of more decomposable word

→ Other phonetic phenomena could also give us an insight about morphological structure
A test case: Gemination

• Due to affixational processes: sequence of two identical consonants at morphological boundaries, e.g.

  un#natural
  im#mature

• What happens on the phonetic level?
  • Longer duration than a singleton ( = ‘gemination’)?
  • Same duration as a singleton ( = ‘degemination’)?
Assumptions about gemination in English

Gimson’s Pronunciation of English (2014):

“In general such prefixes result in a doubled consonant when the prefix-final and the stem-initial consonants are identical, e.g. unnecessary is pronounced with a double length [n:]. (This rule does not apply to in- and its variants, so for example illogical is pronounced with only a single /l/).” (p. 248)

Cohen-Goldberg (2013: 1055f):

“Similarly in English, although geminates are banned from monomorphemic words (*spaghe[tt]i) and words containing less productive affixes (e.g.in-: i[n]umerable), they are allowed in words containing more productive affixes and compounds (e.g. un-: u[nn]ecessary; boo[kk]eeper).”
Assumptions about gemination in English

- Gemination depends on the affix involved:
  - *un*- geminates
  - *in*- does not geminate

- Assumption is in line with theory of Lexical Phonology (cf. Kiparsky 1982, Mohanan 1986)
  - Affixes belong to different Lexical Strata
  - Level 1 (like *in*-) affixes display a weak morphological boundary and a great degree of integration with the base
  - Level 2 affixes (like *un*-) display a strong morphological boundary and a lesser degree of integration with the base

Prediction for gemination

<table>
<thead>
<tr>
<th>Strata/Level</th>
<th>Affix</th>
<th>Prediction for gemination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td><em>in-</em></td>
<td><em>in</em> + numerous $\rightarrow$ <em>i/n/umerous</em> $\rightarrow$ <em>i[n]umerous</em></td>
</tr>
<tr>
<td>Level 2</td>
<td><em>un-</em></td>
<td><em>un</em> + natural $\rightarrow$ <em>u/nn/atual</em> $\rightarrow$ <em>u[n:]atural</em></td>
</tr>
</tbody>
</table>
Previous phonetic research

• Only one study empirically investigated gemination in English (Oh and Redford 2012)
• They found some variation in the gemination of *in*-prefixed words:

  *immigration*al: No gemination
  *immemorial*: gemination

→ There seems to be variation within one affix!

→ So, what is the pattern of the variation in gemination?
  Is it really the affix which determines gemination/degemination?

The research question

Is the /n/ in *unnatural* longer than the /n/ in *unable*?

If yes, *un-* geminates!

Is the /m/ in *immature* longer than the /m/ in *impossible*?

If yes, *in-* geminates!

→ Implications for theories of morpho-phonology, such as Lexical Phonology
Methodology

• *in*- and *un*- prefixed words with a double or a single consonant at the morphological boundary

• Switchboard Corpus (Godfrey & Holliman 1997)

• Telephone conversations, North American English

• For the prefix *in* the allomorph /ɪm/ was investigated

• Manual segmentation and acoustic measurements in Praat (Boersma & Weenink 2014)

• Coding of additional variables, e.g. frequencies, duration of the preceding segment, word duration
Overview of the data

<table>
<thead>
<tr>
<th></th>
<th>Double Consonant</th>
<th>Single Consonant</th>
<th>Total per affix</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>In-</em></td>
<td>94</td>
<td>65</td>
<td>159</td>
</tr>
<tr>
<td><em>Un-</em></td>
<td>22</td>
<td>133</td>
<td>155</td>
</tr>
</tbody>
</table>
Overview of the *un*-data
Overview of the *un*-data
Overview of the *un*-data
Overview of the *in*-data
Overview of the *in*-data

Data distribution: *in*-
Overview of the *in*-data
Methodology: Statistical Analysis

• Linear regression:
  • look at the effect of one variable
  • while holding all others constant

• Predict duration of the nasal on the basis of the number of consonants
Methodology: Analysis Model 1

Dependent Variable: Absolute Duration of the Nasal in seconds

Variable of interest: Number of Nasals

Covariates: Word Form Frequency, Relative Frequency, Preceding Segment Duration, Number of Segments in the Word
Methodology: Analysis Model 2 in

**Dependent Variable:** Absolute Duration of the Nasal in seconds

**Variables of interest:**
1. Number of Nasals
2. Semantic Transparency
3. Type of Affix

**Covariates:** Word Form Frequency, Relative Frequency, Preceding Segment Duration, Number of Segments in the Word
## Results: Model 2 *in-*

### Semantic Transparency and Type of Affix:

<table>
<thead>
<tr>
<th>Semantic Transparency</th>
<th>In-Locative</th>
<th>In-Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transparent</td>
<td><em>immigrate</em></td>
<td><em>immature</em></td>
</tr>
<tr>
<td>Opaque</td>
<td><em>imply</em></td>
<td><em>impunity</em></td>
</tr>
</tbody>
</table>
Results: Model 1 *un-*

- Doubles are longer than singles $\Rightarrow$ Gemination
Results: Model 2 *in-*

- Doubles are longer than singles $\Rightarrow$ Gemination
Results: Model 2 *in*- 

Interaction of Semantic Transparency with Type of Affix: 

- Opaque: no effect (e.g. *imply* vs. *impunity*)
- Transparent: /m/ is shorter in locative *in* (e.g. *immigrate*) than in negative *in* (e.g. *immature*)
Summary: Results

There is a difference in duration between double consonants and singletons for *un*-affixed and for *in*-affixed words.

/\n/ in *unnatural* is longer than the /\n/ in *unable*!

AND

/\m/ in *immature* is longer than the /\m/ in *impossible*!

Both, *in*- and *un*- prefixed words can geminate!
Summary: Results

• Effect of semantic transparency in interaction with type affix for \textit{in}-prefixed words

• These effects are robust in natural speech and also hold if we control for phonetic effects such as word duration
Implications

Both, *in*- and *un*- prefixed words can geminate!

• Challenges simple categorical effects of Lexical Strata as suggested by Lexical Phonology
• Demands for models of morpho-phonological interaction which take other factors into account

Effect of semantic transparency in interaction with type of affix for *in*-prefixed words

• Morphological structure is mirrored by phonetic detail
• Challenges models of speech production that state that post-lexical phonology has no access to morphological information (cf. Levelt 1999)

Future Research

• Replication of results with more data and different affixes

• Conduct production experiment and test the assumptions of different models of the phonology-morphology interaction (e.g. Morphological Segmentability Hypothesis, Hay 2003)

Thank you very much for your attention!
References


Appendix
What is gemination?

• Gemination is used for related phenomena on different levels: Phonetic, Phonological, Orthographic

  – Phonetic: long consonants which can be distinguished from short consonants by their length
  – Phonological: Distinction of meaning in languages like Italian, Finnish, Arabic, e.g. *papa* vs. *pappa*
  – Orthographically: Doubling of consonants
What is gemination?

<table>
<thead>
<tr>
<th>Level</th>
<th>Italian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonetics</td>
<td>[pap:a]</td>
<td>[ʌnʌɛtrəl]? [ʌnɑɛtrəl]? [ʌnætrəl]?</td>
</tr>
<tr>
<td>Orthography</td>
<td>pappa</td>
<td>unnatural</td>
</tr>
</tbody>
</table>
# Data Distribution

<table>
<thead>
<tr>
<th>Segmental Sequence</th>
<th>Example</th>
<th>Number of Types</th>
<th>Number of Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Un+n</td>
<td>Unnecessary</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>Un+C</td>
<td>Unfit</td>
<td>51</td>
<td>66</td>
</tr>
<tr>
<td>Un+V</td>
<td>Unable</td>
<td>43</td>
<td>67</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>99</strong></td>
<td><strong>155</strong></td>
</tr>
<tr>
<td>Im+m</td>
<td>Immemorial</td>
<td>22</td>
<td>94</td>
</tr>
<tr>
<td>Im+p/b</td>
<td>impossible</td>
<td>64</td>
<td>65</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>86</strong></td>
<td><strong>159</strong></td>
</tr>
</tbody>
</table>
Results: Model 1 \textit{Un-}

Effects of Covariates

- Longer words $\rightarrow$ longer consonant duration
- More segments in a word $\rightarrow$ shorter consonant duration
Results: Model 2 \(lm\)-

Effects of Covariates

- Longer words $\rightarrow$ longer consonant duration
- More segments in a word $\rightarrow$ shorter consonant duration
Model 1: un

Call:
\texttt{lm(formula = log(AbsDurCon) \sim TransitionType + WordDur + NoSegWord, data = unComplex1)}

Residuals:

<table>
<thead>
<tr>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.89180</td>
<td>-0.21009</td>
<td>0.00323</td>
<td>0.21242</td>
<td>0.83250</td>
</tr>
</tbody>
</table>

Coefficients:

|                      | Estimate | Std. Error | t value | Pr(>|t|)  |
|----------------------|----------|------------|---------|-----------|
| (Intercept)           | -2.34771 | 0.13833    | -16.972 | < 2e-16   ***|
| TransitionTypesingle-C| -0.34597 | 0.08050    | -4.298  | 3.09e-05  ***|
| TransitionTypesingle-V| 0.72417  | 0.08119    | -8.919  | 1.51e-15  ***|
| WordDur              | 1.10768  | 0.17350    | 6.384   | 2.04e-09  ***|
| NoSegWord            | -0.09859 | 0.01554    | -6.342  | 2.52e-09  ***|

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3187 on 150 degrees of freedom
Multiple R-squared:  0.5626,  Adjusted R-squared:  0.5509
F-statistic: 48.24 on 4 and 150 DF,  p-value: < 2.2e-16
Model 2: im

Call:
lm(formula = log(AbsDurCon) ~ NoCons + WordDur + NoSegWord + MorphBound * Affix, data = imComplex2)

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals</td>
<td>-0.64335</td>
<td>-0.15523</td>
<td>-0.01938</td>
<td>0.15791</td>
<td>0.75189</td>
</tr>
</tbody>
</table>

Coefficients:

|                      | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | -2.84293 | 0.10155    | -27.995 | < 2e-16  ***|
| NoConsdouble         | 0.26600  | 0.06236    | 4.266   | 3.49e-05 ***|
| WordDur              | 0.50065  | 0.14642    | 3.419   | 0.000806 ***|
| NoSegWord            | -0.03508 | 0.01358    | -2.584  | 0.010720 * |
| MorphBoundtransparent| -0.18686 | 0.08473    | -2.205  | 0.028932 * |
| AffixinNeg           | -0.00515 | 0.09343    | -0.055  | 0.956118 |
| MorphBoundtransparent:AffixinNeg | 0.36325 | 0.11229    | 3.235   | 0.001493 **|

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2435 on 152 degrees of freedom
Multiple R-squared:  0.4394,  Adjusted R-squared:  0.4173
F-statistic: 19.86 on 6 and 152 DF,  p-value: < 2.2e-16