# 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

Hay, J. (2003). *Causes and consequences of word structure*.

# Morphological structure and phonetic detail

**Decomposability:**

| Highly decomposable | Less decomposable |
| --- | --- |
| • Semantically transparent<br>• Phonologically transparent<br>• Easy to segment | • Semantically opaque<br>• Phonologically opaque<br>• Difficult to segment |

*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

Kiparsky, P. (1982). Lexical morphology and phonology. In *Linguistics in the morning calm.*
Mohanan, K. P. (1986). *The theory of lexical phonology*.

# Prediction for gemination

| Strata/ Level | Affix | Prediction for gemination |
|---|---|---|
| Level 1 | *in-* | in +numerous → i/n/umerous → i[n]umerous |
| Level 2 | *un-* | un + natural → u/nn/atural → u[n:]atural |

# 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:

*immigrational*: 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?

Oh, G. E., & Redford, M. A. (2012). The production and phonetic representation of fake geminates in English.

# 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

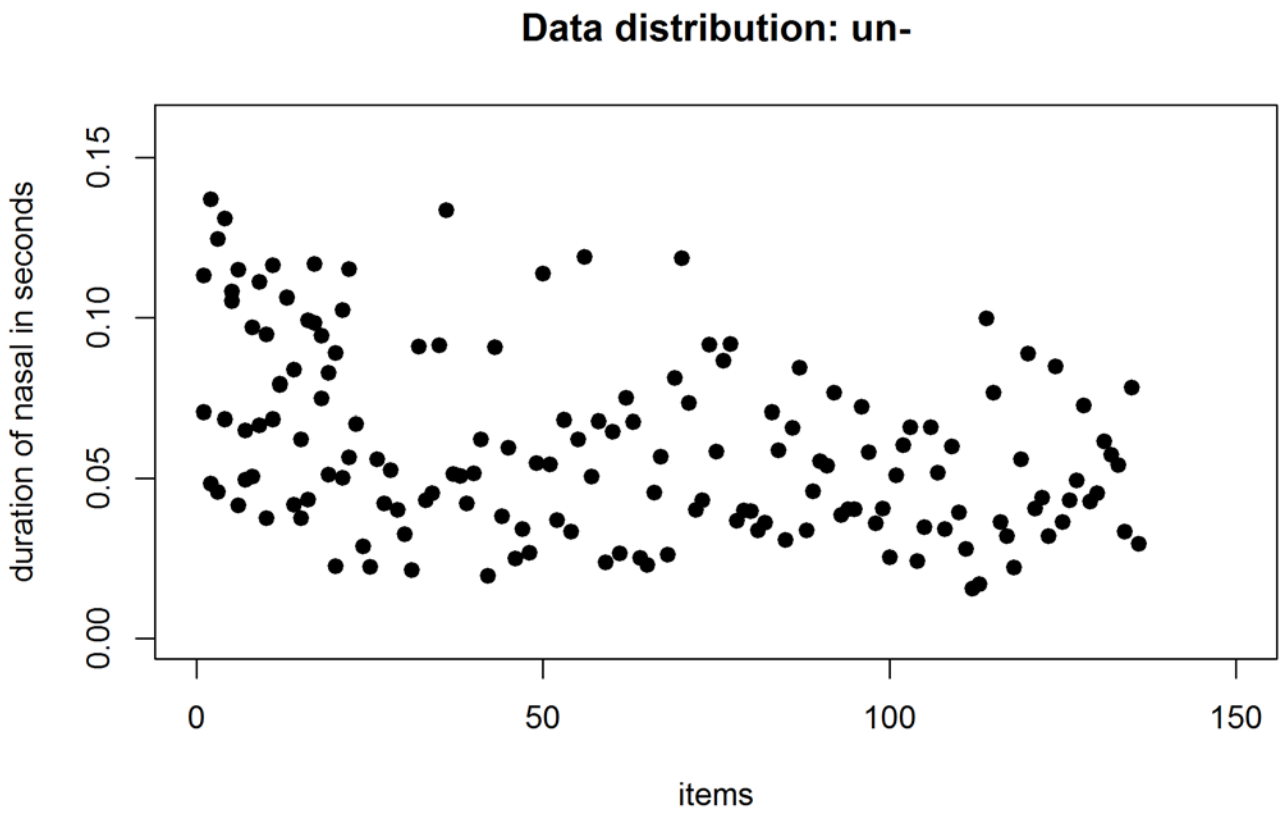
# Methodology

- *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

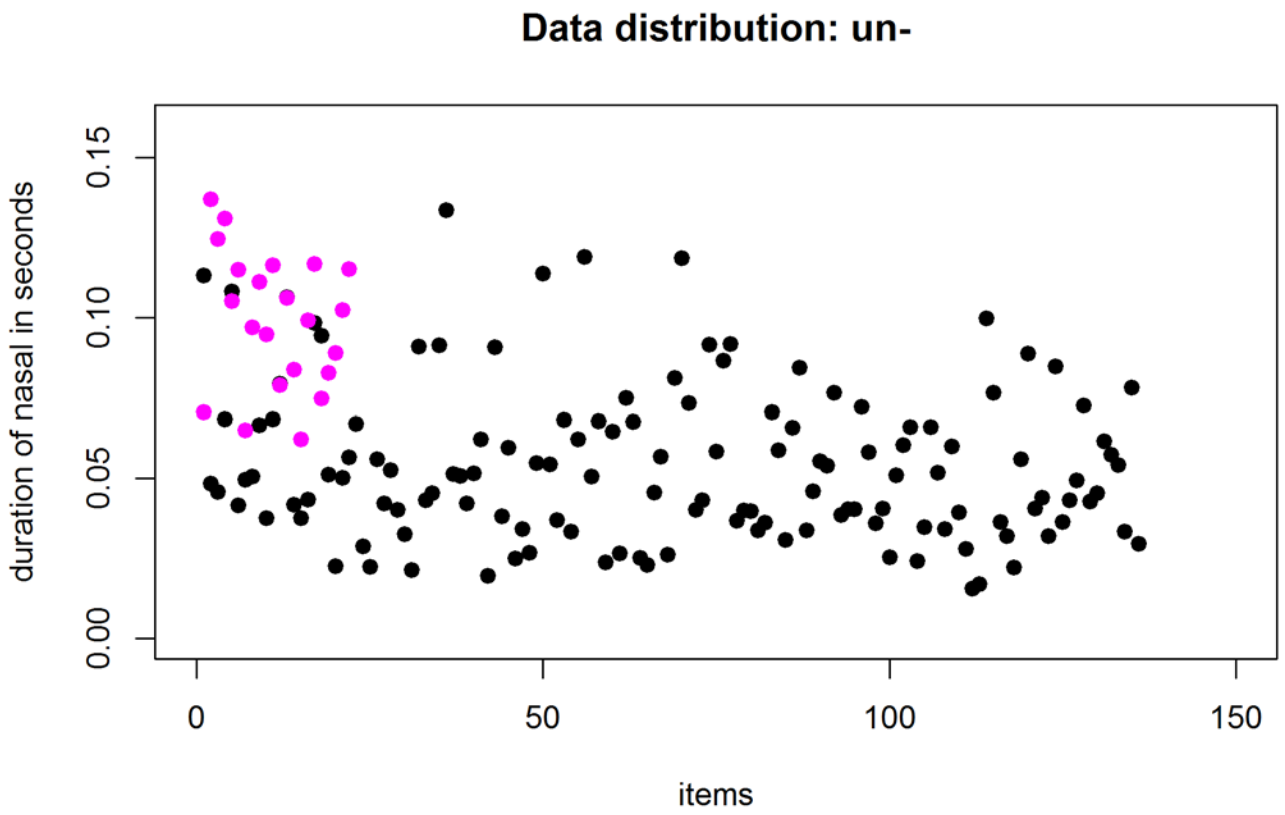
# Overview of the data

| | Double Consonant | Single Consonant | Total per affix |
|---|---|---|---|
| *In-* | 94 | 65 | 159 |
| *Un-* | 22 | 133 | 155 |

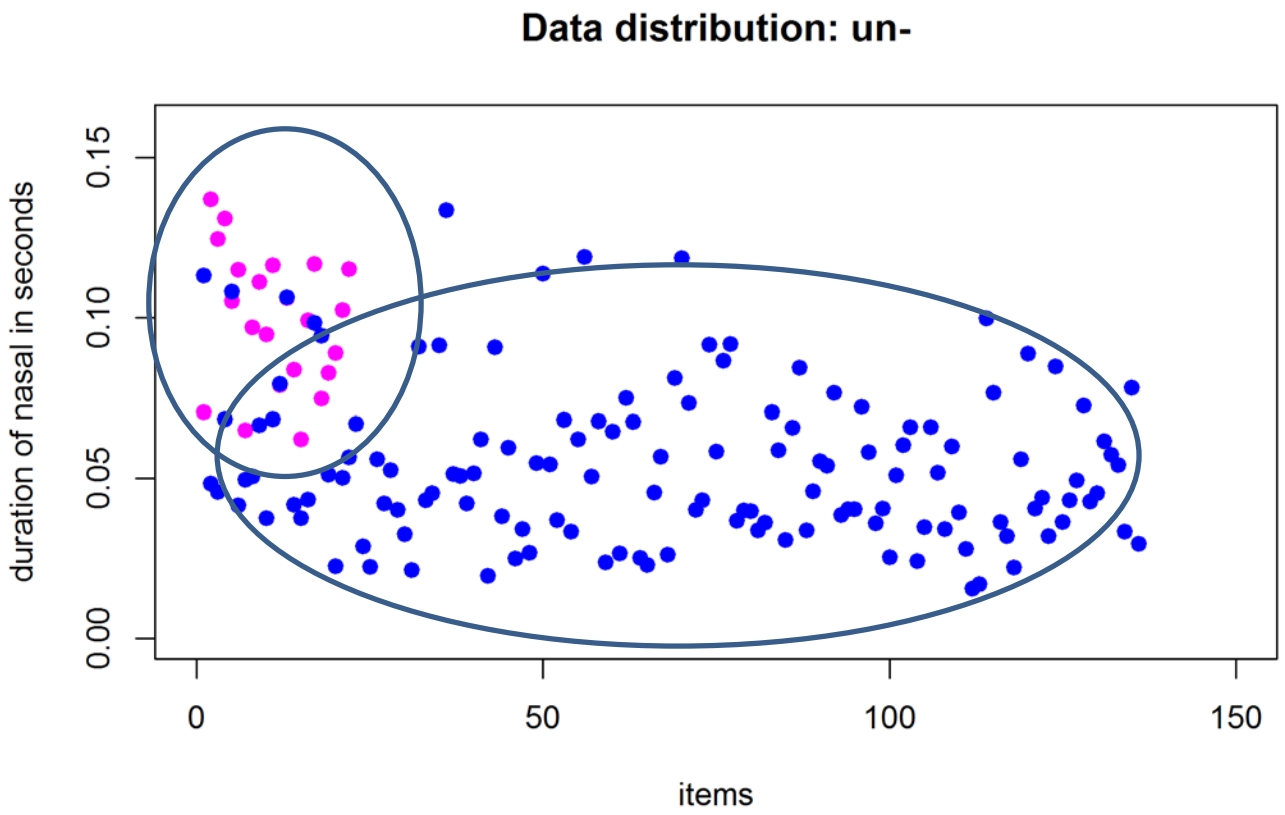
# Overview of the *un*-data



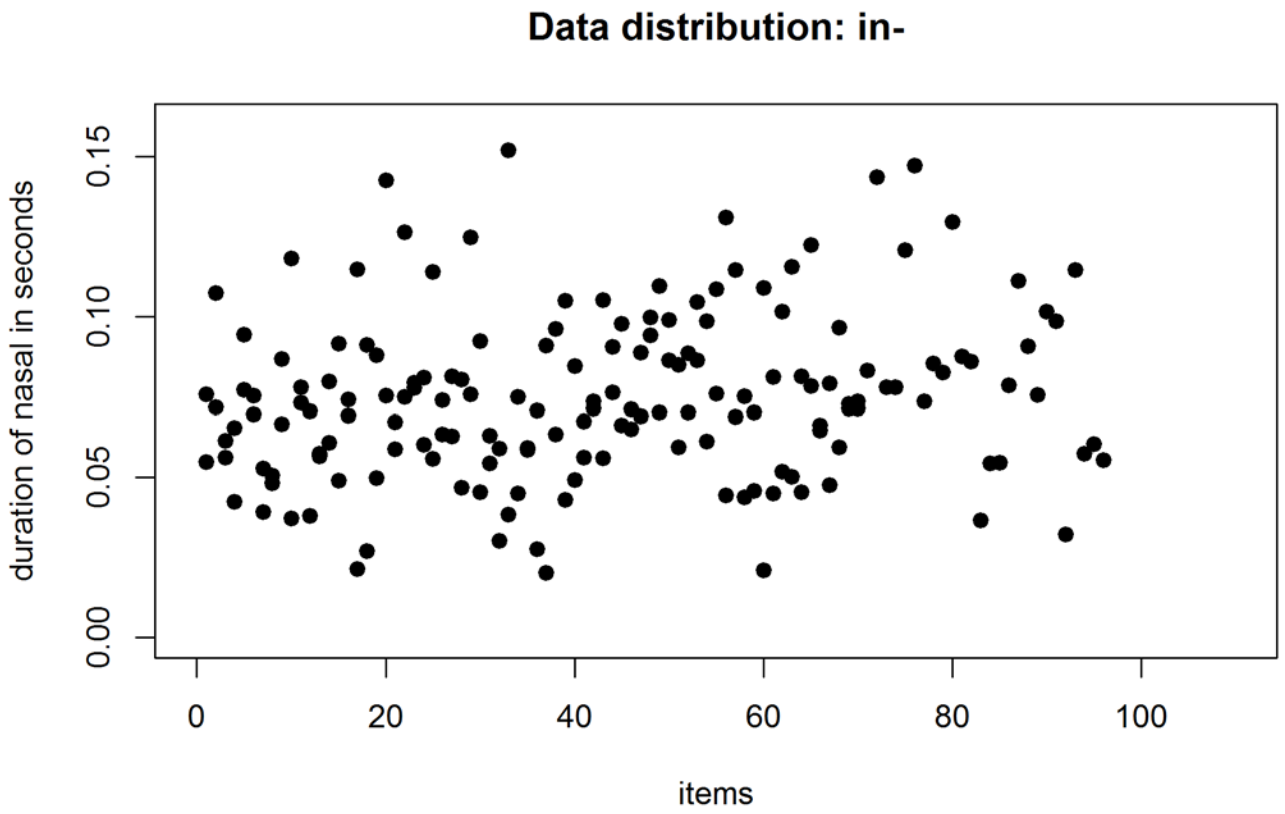
# Overview of the *un*-data



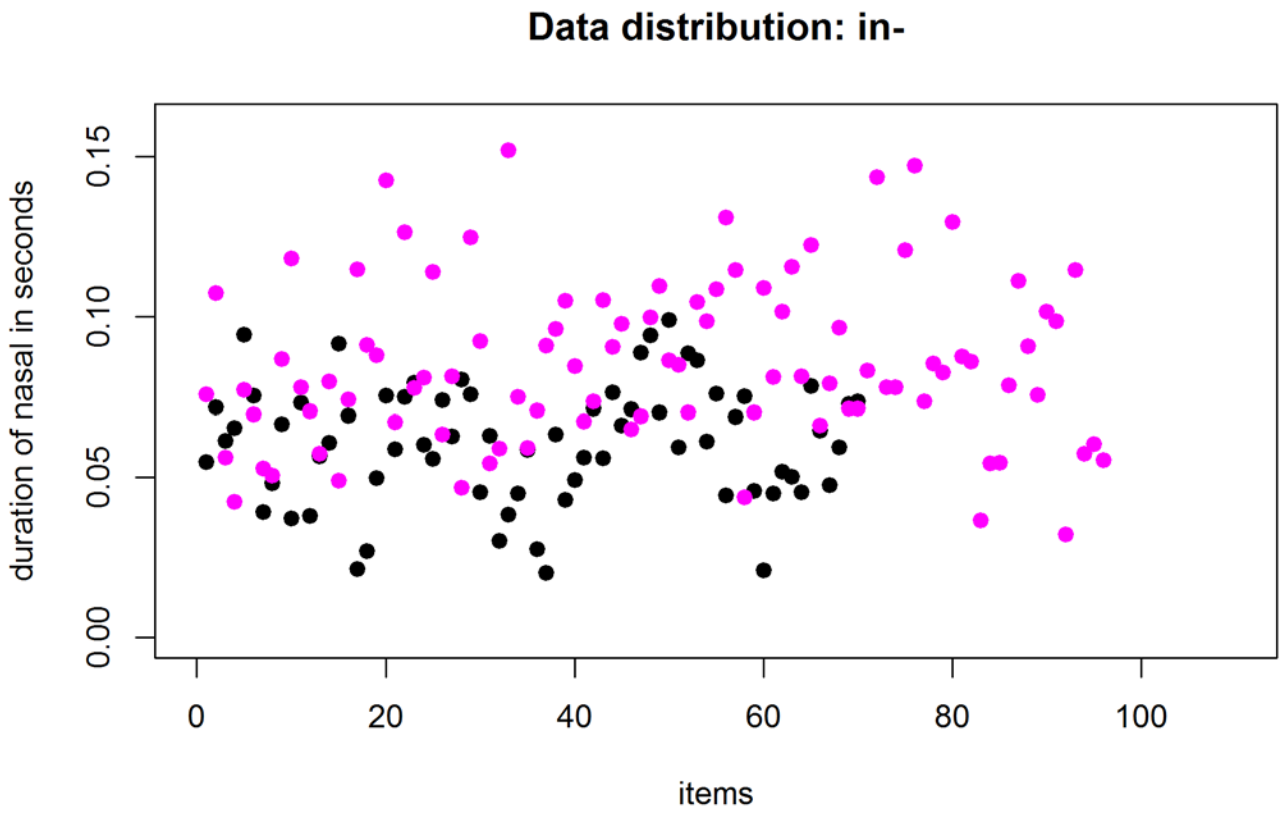
# Overview of the *un*-data



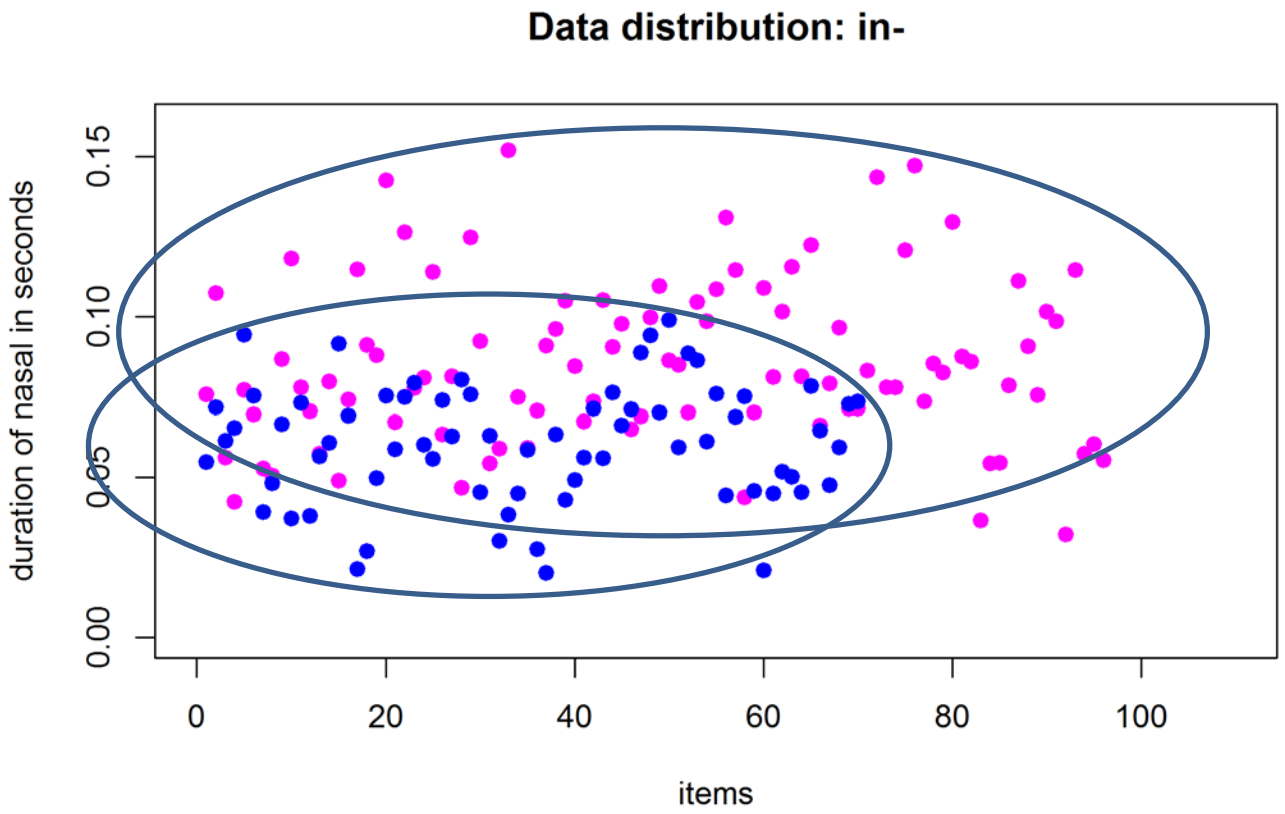
# Overview of the *in*-data



# Overview of the *in*-data



# 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 *un*

**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

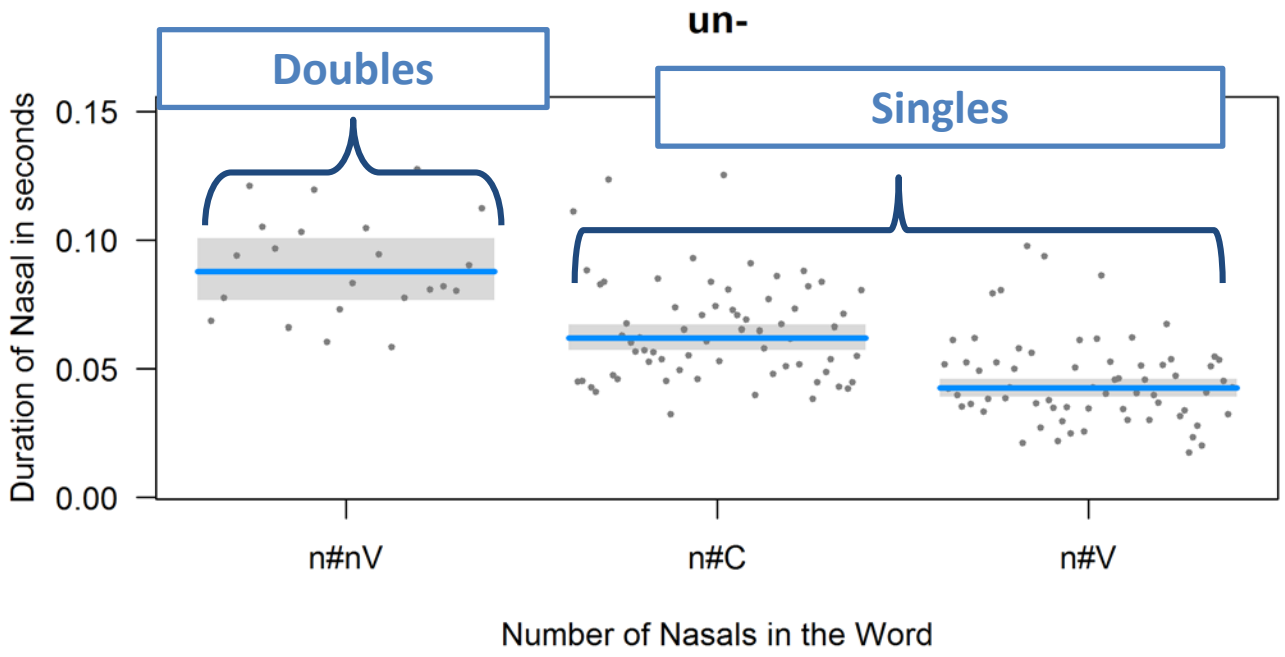
**Covariates:** Word Form Frequency, Relative Frequency, Preceding Segment Duration, Number of Segments in the Word

# Results: Model 2 *in-*

**Semantic Transparency and Type of Affix:**

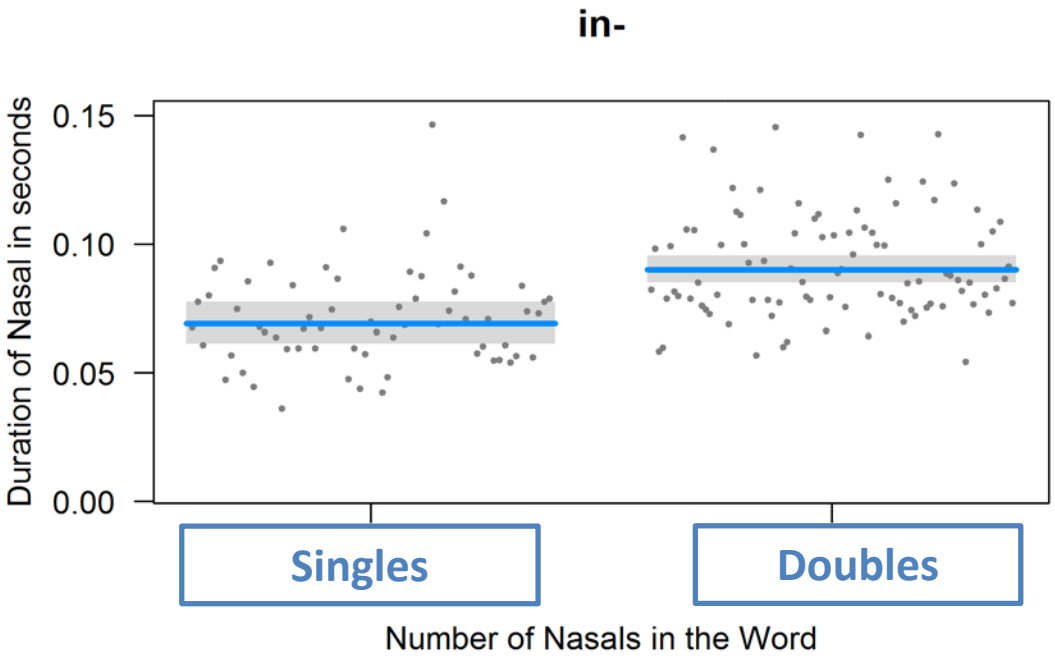
| Type of affix / Semantic Transparency | **In-Locative** | **In-Negative** |
|---|---|---|
| **Transparent** | *immigrate* | *immature* |
| **Opaque** | *imply* | *impunity* |

# Results: Model 1 *un-*




- **Doubles are longer than singles → Gemination**

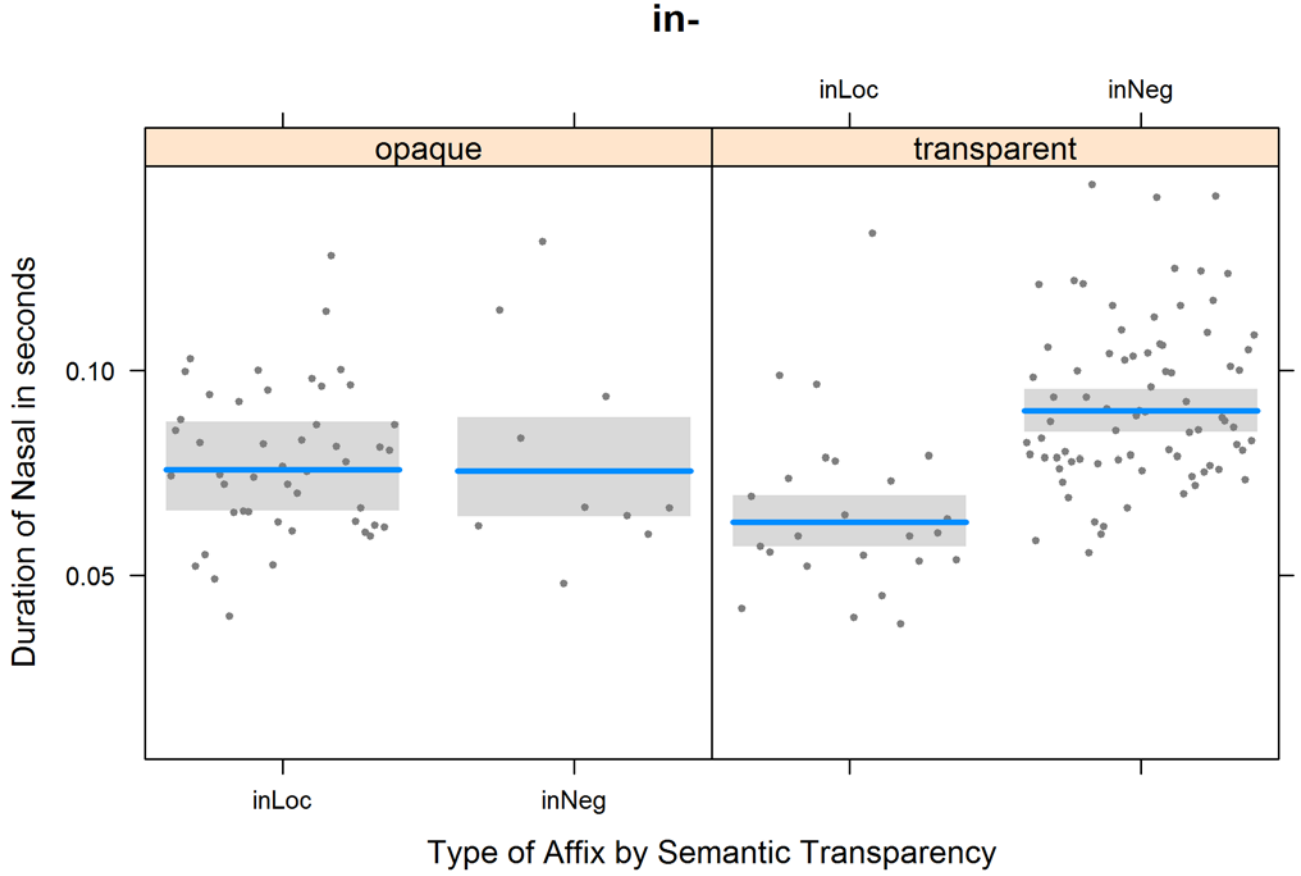
# Results: Model 2 *in-*




- **Doubles are longer than singles → 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 *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


Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/

Cho, T. (2001). Effects of Morpheme Boundaries on Intergestural Timing: Evidence from Korean. Phonetica, 58(3), 129–162.

Cohen-Goldberg, Ariel M. (2013): Towards a theory of multimorphemic word production: The heterogeneity of processing hypothesis. In: Language and Cognitive Processes 28 (7), S. 1036–1064.

Cruttenden, Alan; Gimson, Alfred Charles (2014): Gimson's pronunciation of English. 8th ed. London, New York: Routledge.

Godfrey, John J.; Holliman, Edward (1997): Switchboard-1 Release 2. [Philadelphia, Pa.]: Linguistic Data Consortium.

Hay, J. (2007). The phonetics of 'un'. In J. Munat (Ed.), Studies in functional and structural linguistics: v. 58. Lexical creativity, texts and contexts (pp. 39–57). Amsterdam, Philadelphia: J. Benjamins Pub. Co.

Kiparsky, Paul (1982): Lexical morphology and phonology. In: Linguistics in the morning calm. Selected papers from SICOL-1981. Unter Mitarbeit von The linguistic society of Korea. Seoul, Korea: Hanshin Pub. Co.

Mohanan, K. P. (1986). The theory of lexical phonology. Studies in natural language and linguistic theory: [v. 6]. Dordrecht, Boston, Norwell, MA: D. Reidel Pub. Co.; Sold and distributed in the U.S.A. and Canada by Kluwer Academic.

Oh, Grace E.; Redford, Melissa A. (2012): The production and phonetic representation of fake geminates in English. In: Journal of Phonetics 40 (1), S. 82–91.

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Italian Journal of Linguistics / Rivista di Linguistica.*,

R Development Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

# 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

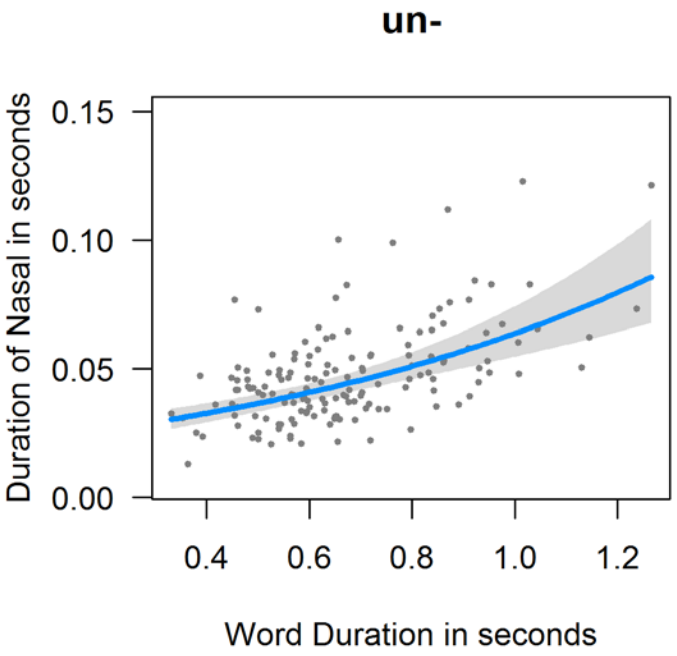
# What is gemination?

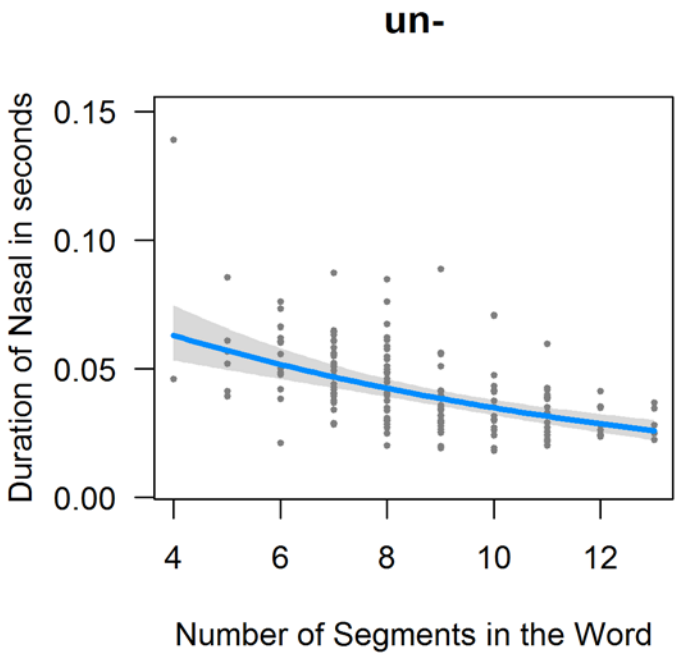
| Level | Italian | English |
|---|---|---|
| Phonetics | [pap:a] | [ʌ**nn**ætrəl]?<br>[ʌ**n:**ætrəl]?<br>[ʌ**n**ætrəl]? |
| Orthography | pappa | unnatural |

# Data Distribution

| Segmental Sequence | Example | Number of Types | Number of Tokens |
|---|---|---|---|
| Un+n | *Unnecessary* | 5 | 22 |
| Un+C | *Unfit* | 51 | 66 |
| Un+V | *Unable* | 43 | 67 |
| **Total** | | 99 | 155 |
| Im+m | *Immemorial* | 22 | 94 |
| Im+p/b | *impossible* | 64 | 65 |
| **Total** | | 86 | 159 |

# Results: Model 1 *Un-*

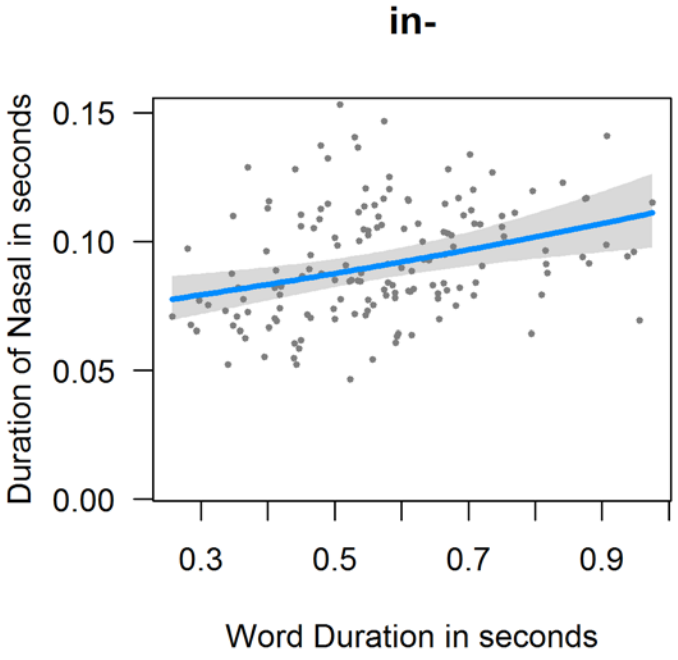


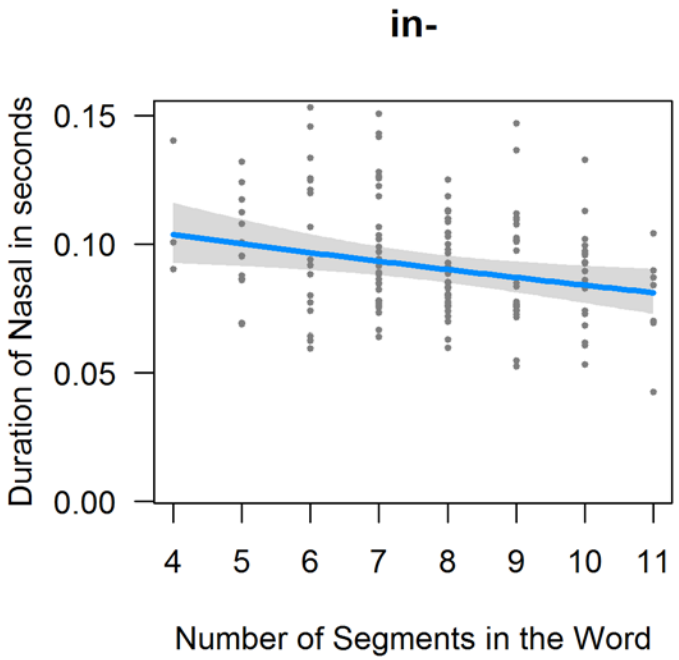




**Effects of Covariates**

- Longer words → longer consonant duration
- More segments in a word → shorter consonant duration

# Results: Model 2 *Im-*




**Effects of Covariates**

- Longer words → longer consonant duration
- More segments in a word → shorter consonant duration

# Model 1: un

```
Call:
lm(formula = log(AbsDurCon) ~ TransitionType + WordDur + NoSegWord,
    data = unComplex1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.89180 -0.21009  0.00323  0.21242  0.83250

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:
     Min       1Q   Median       3Q      Max
-0.64335 -0.15523 -0.01938  0.15791  0.75189

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
```