

Morpho-phonetics and hierarchical morphological structure: The case of triconstituent compounds

Ingo Plag

Heinrich-Heine-Universität Düsseldorf

Collaborators

Gero Kunter (U Düsseldorf)

Funding

Strategischer Forschungsförderfonds Heinrich-Heine-Universität Düsseldorf

Research questions

- How does paradigmatic and syntagmatic morphological structure affect the articulatory, acoustic and phonological properties of complex words?
- What do the phonological and phonetic properties of complex words reveal about the morphological structure of these words and about their paradigmatic relationships?
- What are the implications for the organization of the mental lexicon and for models of morpho-phonology, of lexical processing, of speech production and speech perception?

Some answers and a new question

- We saw that morphological structure can affect the acoustic properties of complex words in unexpected ways
- Morphological information is predictive of gemination duration (*in-*, *in-* and *un-*) and of suffix duration (final S and D in English)
- Implications for the organization of the mental lexicon and for models of morpho-phonology, of lexical processing, of speech production and speech perception
- So far, only **bi-morphemic** words have been analyzed. How does **hierarchical morphological structure** manifest itself in speech? Does it?

The problem

How much morphology is in the speech signal with words that have more than two morphemes?

- Trimorphemic words
 - base-suffix-suffix (left-branching): *[fool-ish]-ness, [princ-ess]-hood*
 - prefix-prefix-base (right-branching): *anti-[super-hero], non-[trans-sexual]*
 - prefix-base-suffix (left-branching): *[super-hero]-less, [inter-view]-er*
 - prefix-base-suffix (right-branching): *non-[conform-ist], anti-[soci-al]*
- Affix boundaries may differ in strength:

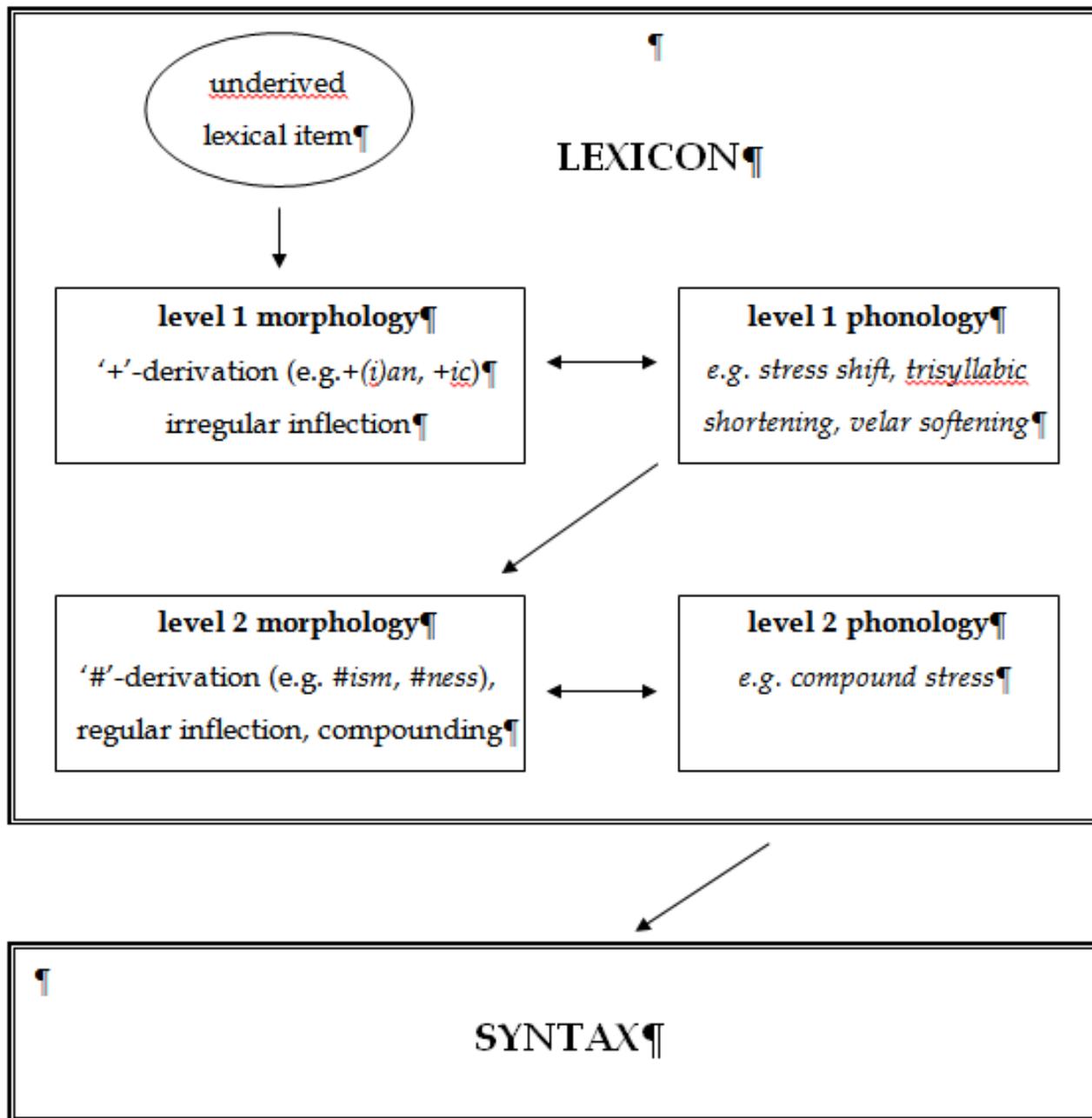
[[base + X] # Y]

(e.g. Chomsky & Halle 1968)

- Lexical Phonology (e.g. Mohanan 1981, Kiparsky 1982, Kaisse & Shaw 1985)

Main questions

- Is there a larger system governing the interaction of phonology and morphology?
- Can phonological processes refer to morphological structure?
- Can any morphological process refer to any phonological structure, or are there limits?
- What is the possible phonological influence of affixes on stems?
- Is morpheme order constrained (if so, by what)?



Level ordering

Level I suffixes:

+al, +ate, +ic, +ion, +ity, +ive, +ous

Level I prefixes:

be+, con+, de+, en+, in+, pre+, re+, sub+

Level II suffixes:

#able, #er, #ful, #hood, #ist, #ize, #less, #ly, #ness, #wise

Level II prefixes:

anti#, de#, non#, re#, sub#, un#, semi#

Properties of affixes by level

- Level 1 affixes tend to be of foreign origin ('Linate').
Level 2 affixes are mostly Germanic.
- Level 1 affixes can attach to bound roots and to words.
level 2 affixes attach to words only.
- Level 1 affixes tend to be phonologically more integrated into their base (stress shifts and other morpho-phonological alternations).
level 2 suffixes do not affect their bases phonologically.
- Level 1 affixes are generally less productive than stratum 2 affixes.
- Level 2 affixes attach outside level 1 affixes

Problems...

Affix ordering is too simple

- No predictions within levels
- Double membership of many affixes: wrong predictions across levels
- Empirically highly unsatisfactory

What constrains affix ordering?

- Selectional restrictions (Fabb 1988, Plag 1999, Hay & Plag 2004)
- 'Complexity-based Ordering' (Hay & Plag 2004, Plag & Baayen 2009)

Selectional restrictions

What has been analyzed as would-be stratal behavior falls out automatically from the phonological, morphological and semantic properties of affixes. (e.g. Plag 1998)

These properties must be stated anyway to account for the particular behavior of a given affix.

No further stratal apparatus is necessary.

Example: verb-deriving *-en* (*deepen*) only attaches to monosyllables ending in an obstruent, hence suffixed adjectives are not eligible bases, hence *-en* cannot combine with such suffixes.

Complexity-based ordering

Basic idea

The morphological separability of affix and base is a graded phenomenon, with far-reaching consequences for affix stacking.

“While some affixes basically tolerate no internal structure, others will tolerate structure to some minimum degree. The degree of internal structure tolerated by an affix is . . . determined by how much structure that affix, itself, creates. Phrased in terms of processing, **an affix which can be easily parsed out should not occur inside an affix which can not.**” (Hay, 2002:527–528, my emphasis)

Graded morphological complexity: How?

Relative frequency as an indicator of juncture strength

- The higher the frequency of the derived word in relation to the base word, the less likely is decomposition.
- The lower the frequency of the derived word in relation to the base word, the more important the role of the constituents becomes.

government (BNC lemma freq. 66894) vs. *govern* (BNC lemma freq. 2626)

→ whole-word bias for *government*

discernment (BNC 61) vs. *discern* (BNC 452)

→ strong bias for constituent-driven processing

Graded morphological complexity: How?

Hapax-conditioned productivity as an indicator of juncture strength

\mathcal{P} : the number of hapax legomena, i.e. the words that occur only once in the corpus with a given affix, divided by the number of tokens with that affix.

$$\mathcal{P} = \frac{n_1}{N}$$

Productive processes:

Large numbers of low frequency words and small numbers of high frequency words, hence strong boundaries.

Back to affix ordering

Affixes that are stronger in processing will occur outside affixes that are weaker (Hay 2002, Burzio 1994)

Hay/Plag (2004): Hierarchy of juncture strength

Given a hierarchy of suffixes: *(base)- X-Y-Z-A-B-C-D*

possible: *base-A-B, base-X-A-C, base-Y-Z-A*

impossible: *base-A-Z, base-Y-A-Z, base-X-A-Y*

Largely born out by the data (210 combinations, 15 suffixes)

Plag & Baayen (2009): 31 affixes, 930 combinations

Affix ordering in English

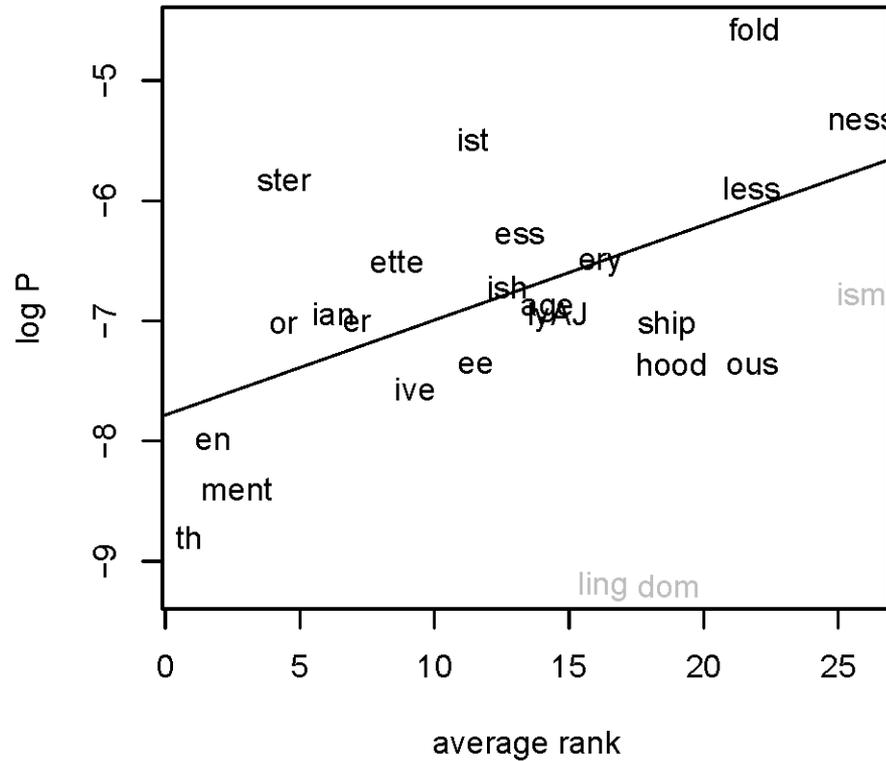
Plag & Baayen (*Language* 85.1, 2009)

	age	aryA	aryN	dom	ee	en	er	ery	ess	ette	fold	fulA	fulN	hood	ian	ish	ism	ist	ive	less	ling	lyAJ	lyAV	ment	ness	or	ous	ship	ster	th	wise
age	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1	1	0	1	1	0	0	1
aryA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0
aryN	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	1	0	0	0	0	0	1	1	0	0	1
dom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ee	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
en	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
er	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1
ery	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
ess	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1
ette	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1
fold	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
fulA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1
fulN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ian	1	1	0	1	0	0	1	1	1	0	0	0	0	1	0	1	1	1	0	1	1	0	0	0	0	1	1	1	0	1	1
ish	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1
ism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ist	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
ive	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
less	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
ling	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lyAJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lyAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ment	0	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1
ness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
or	1	0	0	1	0	0	0	0	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0	0	0	1	1	0	0	1
ous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ship	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ster	1	0	0	1	0	0	0	0	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
th	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
wise	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

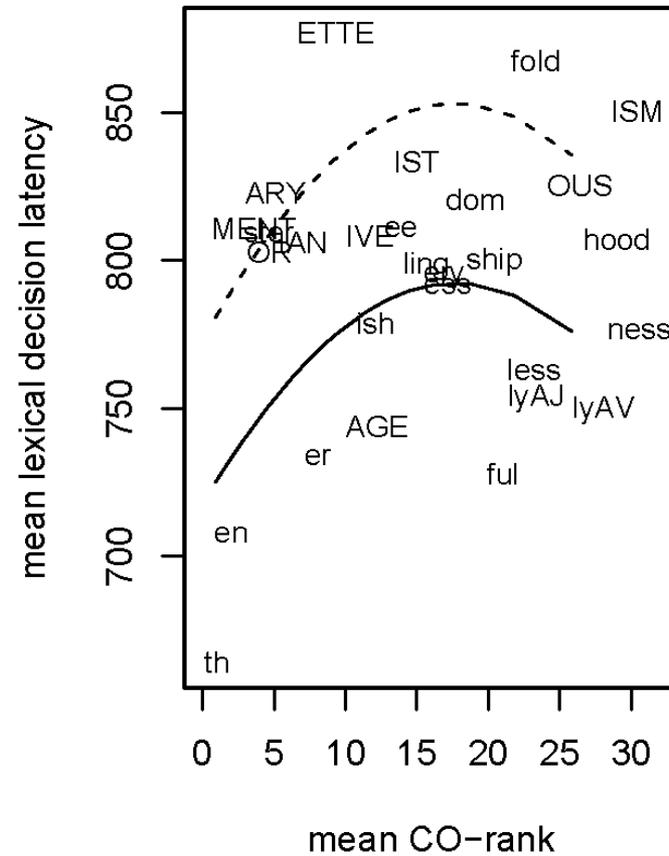
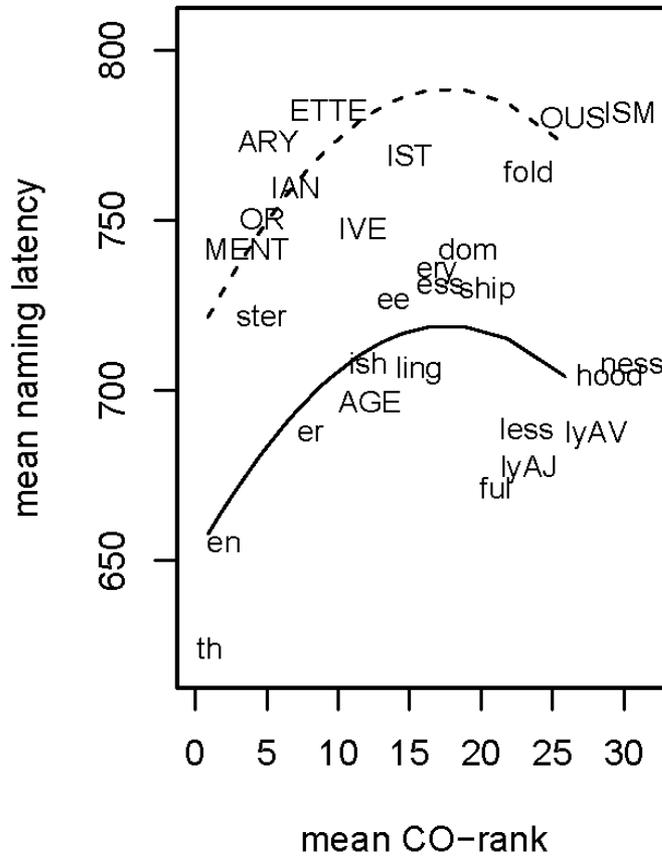
Affix ordering in English: Plag & Baayen 2009

	th	en	ment	aryN	or	ster	ian	er	ette	ish	ive	aryA	ee	ist	age	ess	ery	lyAJ	ling	ship	hood	dom	less	fold	fulA	ous	ism	ness	lyAV	fulN	wise
th	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
en	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ment	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
aryN	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
or	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ster	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ian	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
er	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ette	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ish	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ive	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
aryA	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ee	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ist	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
age	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ess	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ery	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
lyAJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
ling	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ship	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
hood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
dom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
less	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
fold	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
fulA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
ous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
ism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
ness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
lyAV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
fulN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wise	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Correlation of affix rank with productivity



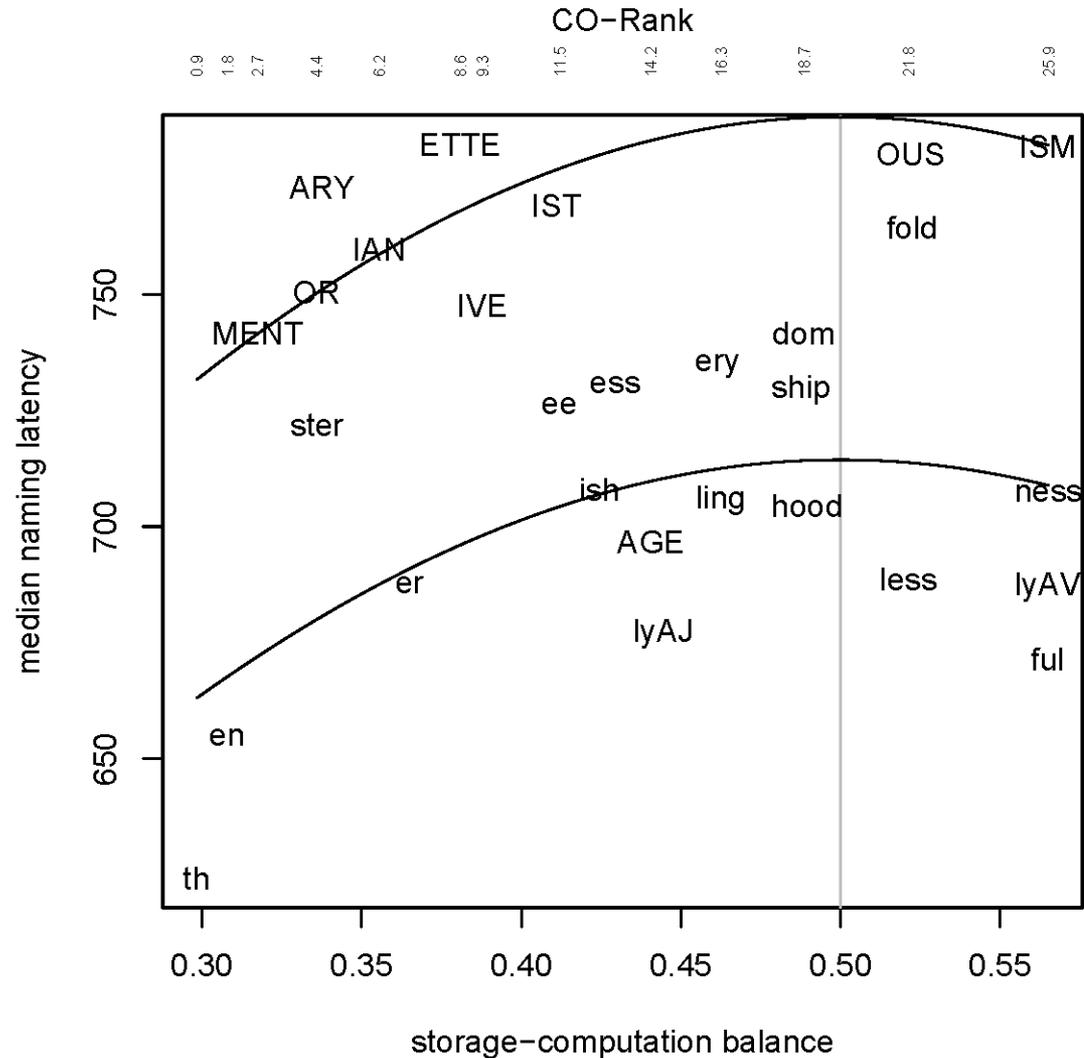
Correlation of affix rank with processing



$$(1) \quad y = a + bx + cx^2$$

$$(2) \quad y = a' + c' \cdot \underbrace{(1 - x')}_{\text{computation}} \cdot \underbrace{(x')}_{\text{storage}}, \quad (c' > 0)$$

Correlation of affix rank with processing: storage vs. computation



Summary

- Affix ordering hierarchy **is** predictive for the average processing complexity of words of a specific morphological category.
- Segmentability correlates with affix rank.
- Morphological integration goes together with phonological/phonetic **integration** (cf. Chomsky & Halle 1968, Kiparsky 1982, Dressler 1985)
- What are the correlates of this integration, and is hierarchical morphological organization reflected in phonetic properties?

Evidence?

Phonetic evidence

- Acoustic duration of the same segments at an affix boundary is shorter than at a compound boundary (Sproat 1993 and Sproat & Fujimura 1993)
- Acoustic properties of affixes vary according to segmentability (Hay 2007)
- Acoustic duration varies according to morphological status: plural vs. singular stems (Kemps et al. 2005a, 2005b)
- Acoustic properties of different /s/ morphemes vary (Plag, Homann & Kunter 2014)

Lack of evidence

- More than one morphological boundary?
- Influence of embedding?

Research question

Can durational properties of complex words tell us something about their internal constituency?

Embedded Reduction Hypothesis

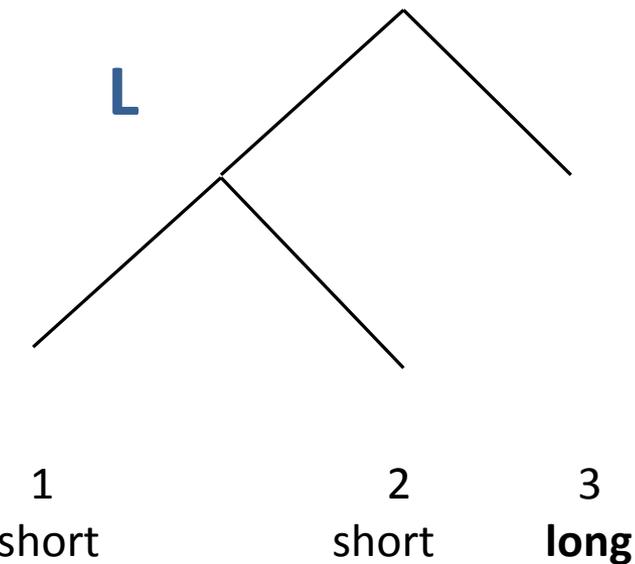
In a complex word with more than two constituents, embedded forms show more phonetic reduction than forms at higher derivational levels

The hypothesis

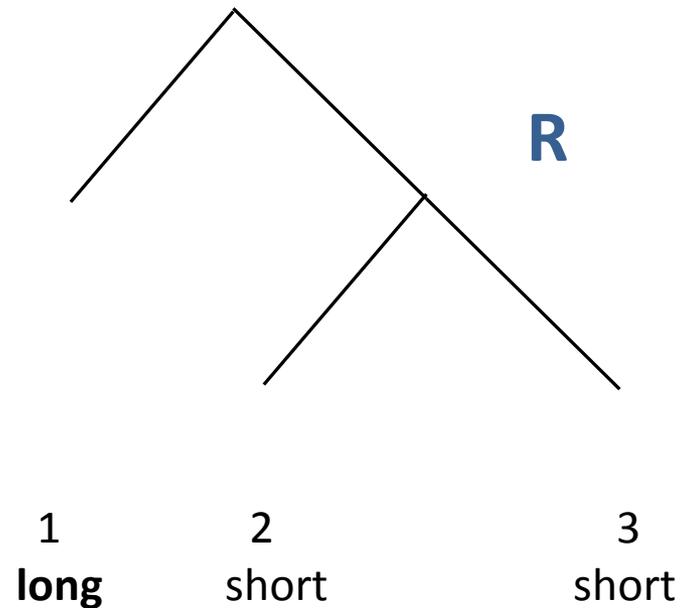
NNN compounds

Embedded forms show more phonetic reduction than forms at higher derivational levels

[weather station] data



team [locker room]



A complication

- Predictability of a word in its context influences its duration (e.g. Jurafsky et al. 2001, Pluymaekers et al. 2005, 2010).
- N-gram frequency:
 - N1-N2 bigram frequency
 - N2-N3 bigram frequency
 - 'embedded' bigram frequency: [N1 N2] N3
N1 [N2 N3]
 - 'cross-boundary' bigram frequency:
[N1 N2] N3
N1 [N2 N3]
- Bigram frequency is also a correlate of lexical integration, hence indirectly also of morphological structure

This study: Triconstituent compounds

- Test the embedded reduction hypothesis with triconstituent NNN compounds
- Use data from an experiment by Kösling (2013) and Kösling et al. (2014)
- Production experiment with 12 speakers
- 40 NNN compounds, embedded in declarative sentences, 40 filler sentences
- 20 left-branching, 20 right-branching
- Different combinations of branching and stress

L/N1 *[coffee table] designer*
 [day care] center

R/N2 *adult [jogging suit]*
 business [credit card]

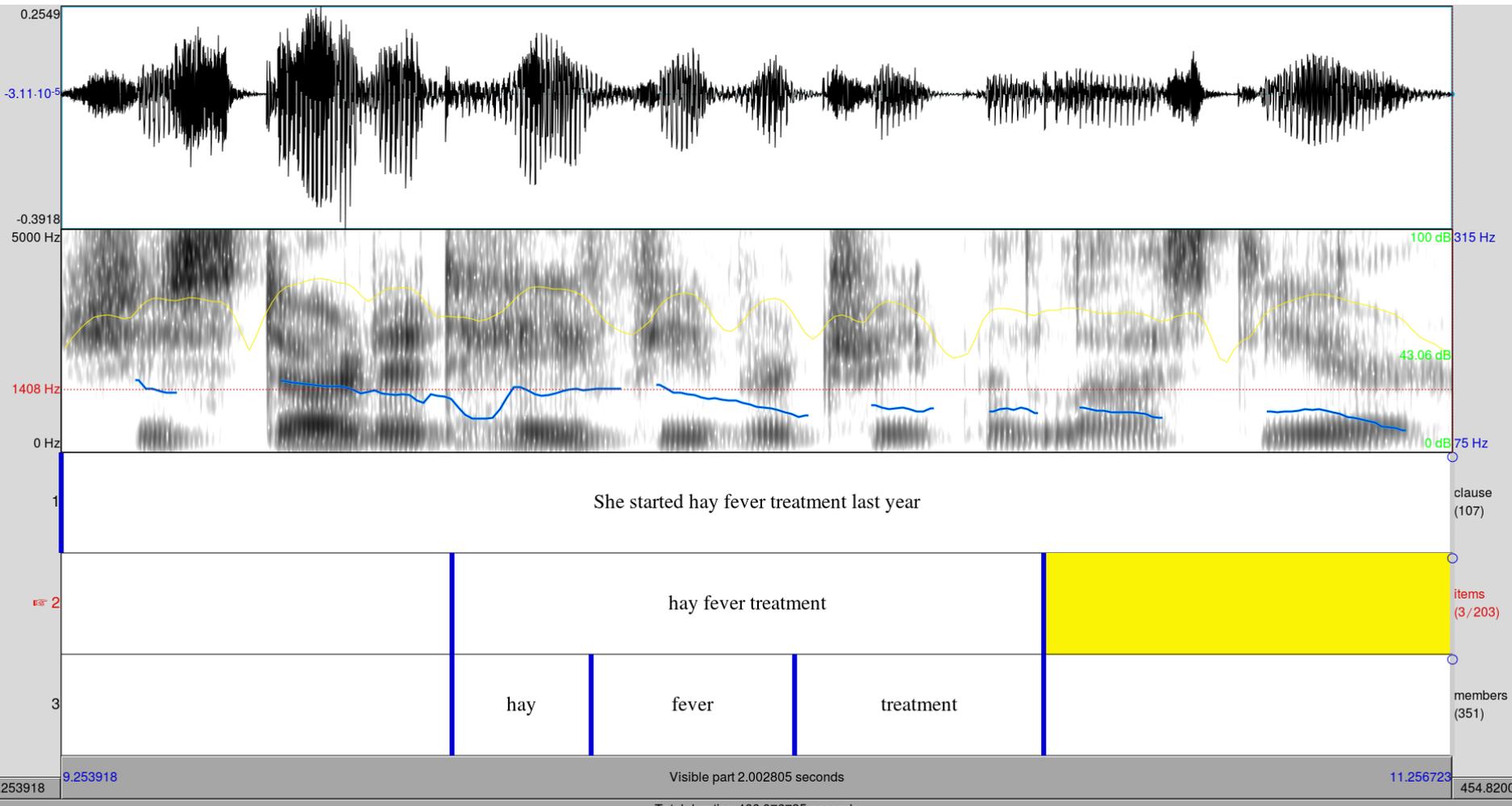
L/N2 *[city hall] restoration*
 [cotton candy] maker

R/N3 *baby [lemon tea]*
 company [internet page]

- most prominent constituent is embedded: works against hypothesis

The data

hay fever treatment



1 She started hay fever treatment last year

2 hay fever treatment

3 hay fever treatment

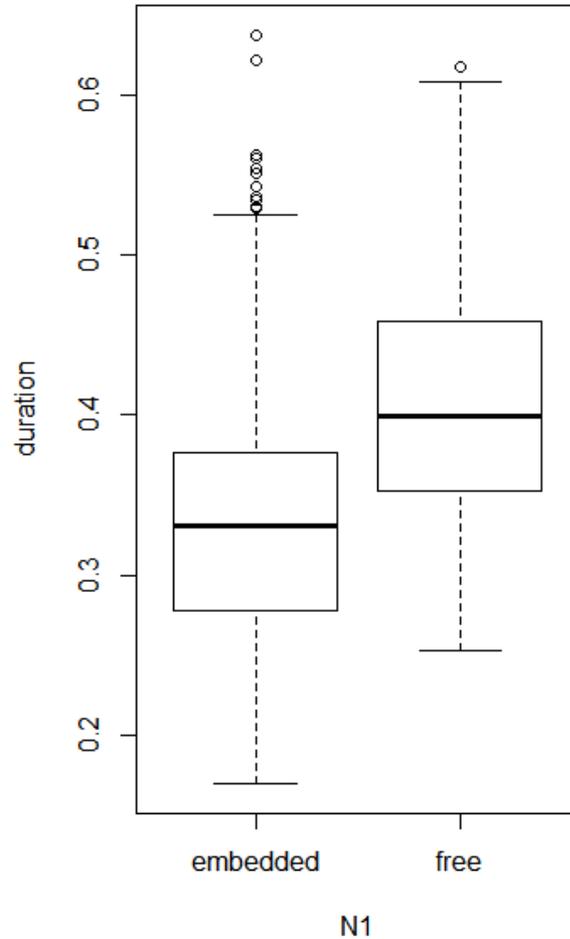
clause (107)

items (3/203)

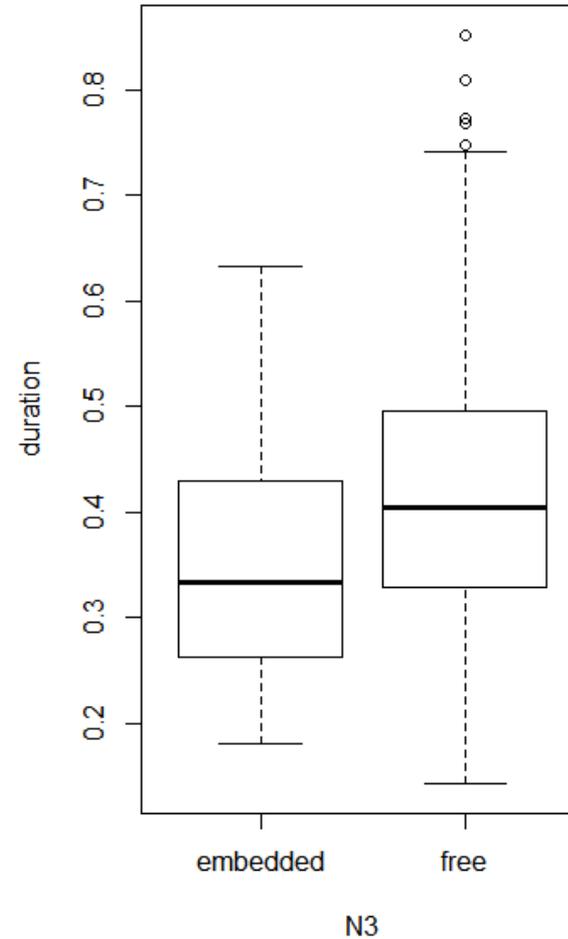
members (351)

Results: first impression

N1 duration by branching

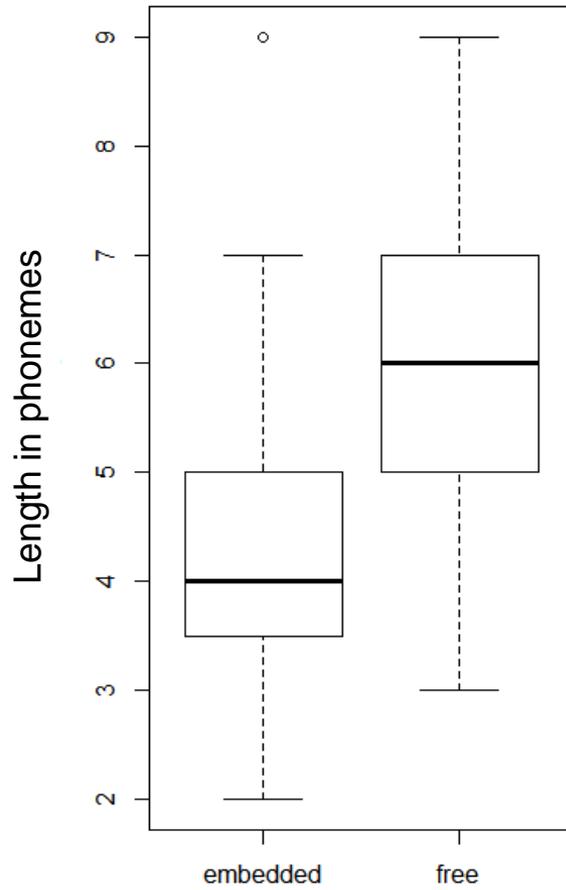


N3 duration by branching



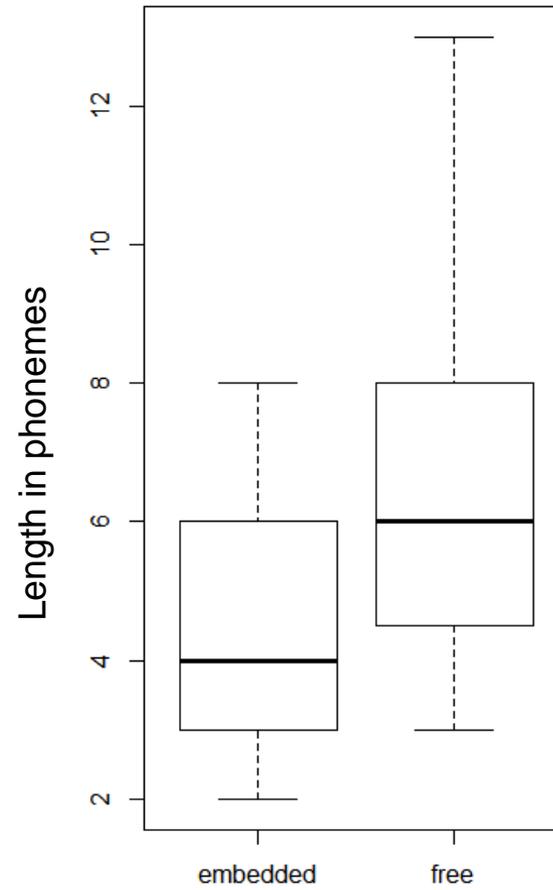
Problem: Covariates

Length in phonemes by embedding



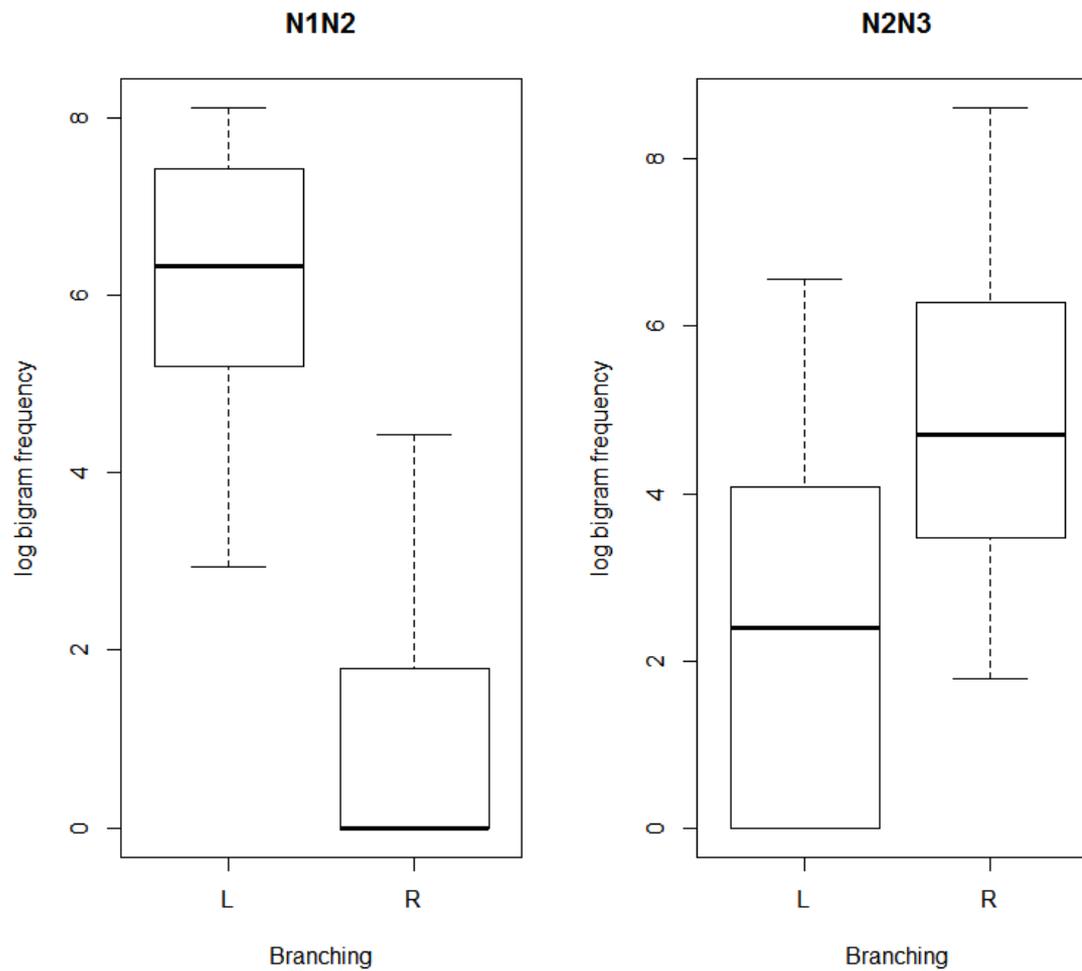
N1

Length in phonemes by embedding



N3

A possible confound: Bigram frequency



Analysis

Mixed effects multiple linear regression

- look at the effect of one variable while holding all others constant
- statistically control for random variables

Dependent variable

DURATION OF CONSTITUENT (in ms, token-based, N= 1431)

Fixed effects (and their interactions)

BRANCHING (*left or right*)

CONSTITUENT (*N1, N2 or N3*)

Embedded bigram frequency

Cross-boundary bigram frequency

CONSTITUENT FREQUENCY (= log of COCA frequency)

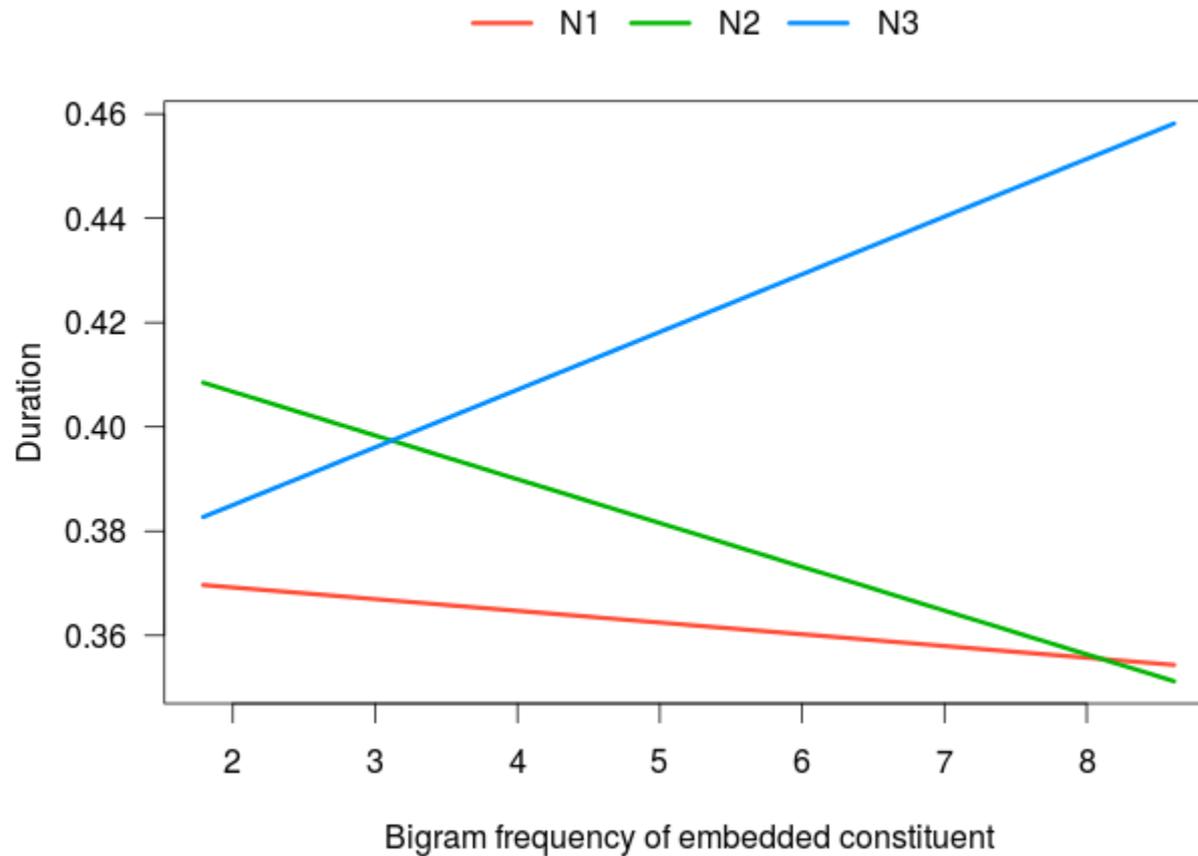
CONSTITUENT LENGTH (in number of phonemes)

ACCENT ON CONSTITUENT (*yes or no*, see Kösling et al. 2014)

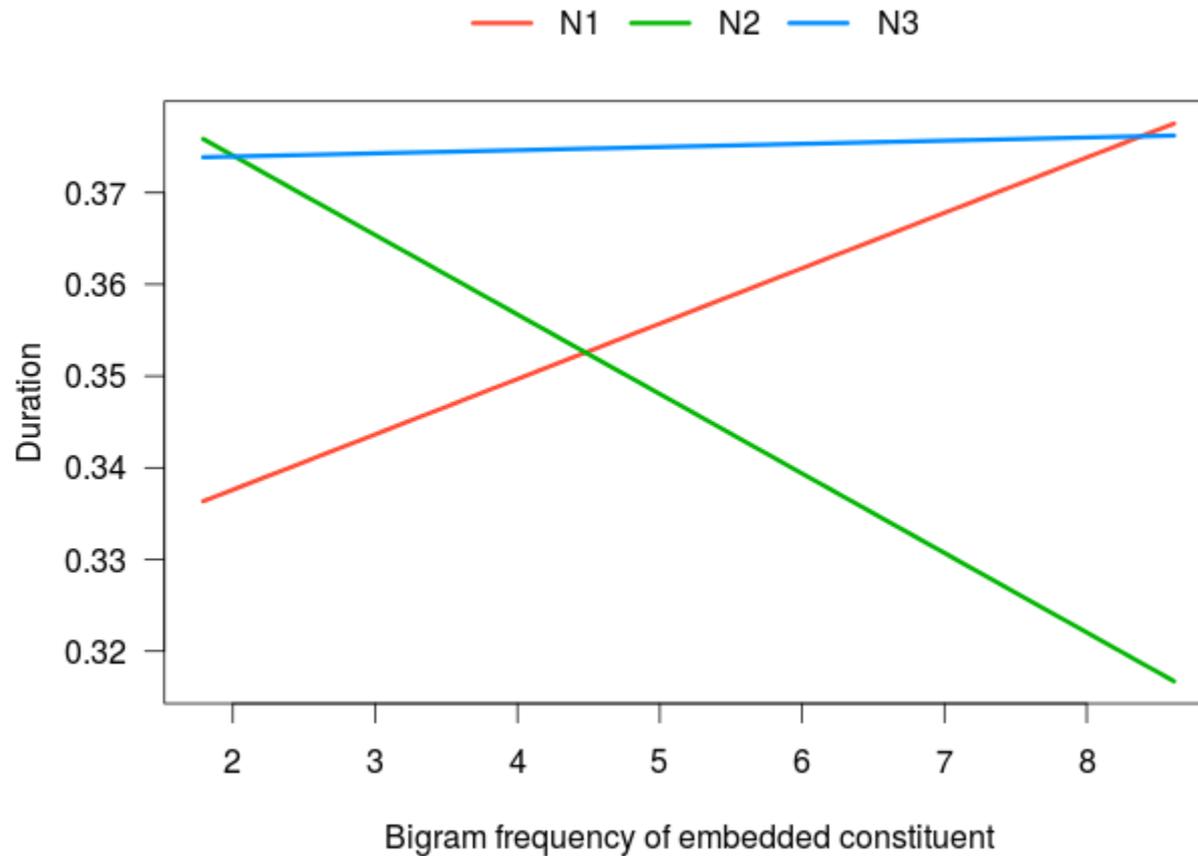
Random effect

SPEAKER

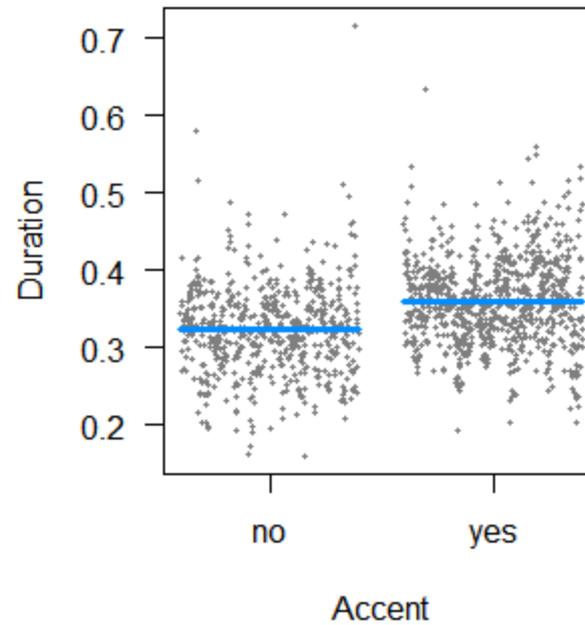
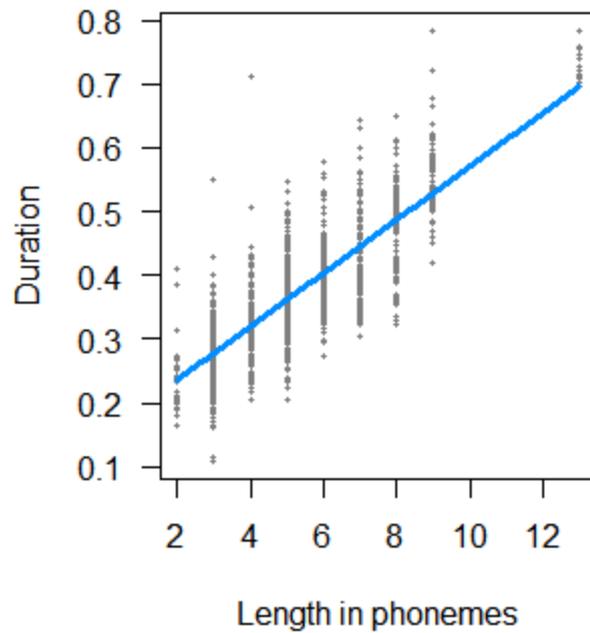
Results: left-branching compounds



Results: right-branching compounds

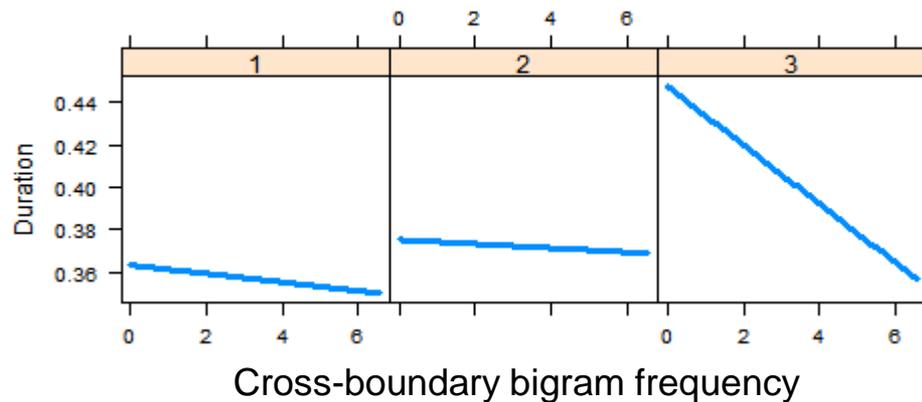


Results: covariates

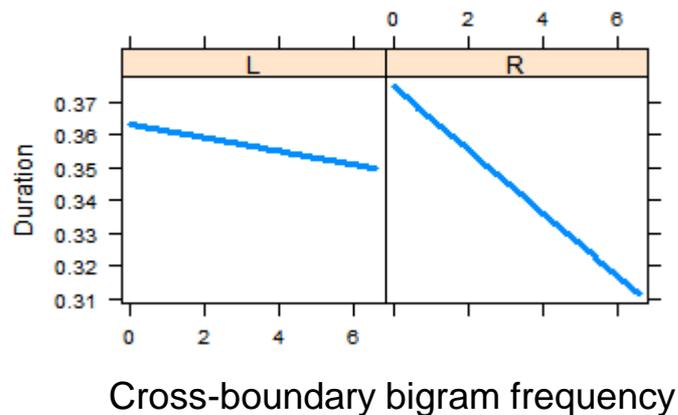


Covariates: cross-boundary bigram frequency

Interaction with constituent:



Interaction with branching:



Summary

Left-branching compounds

- Shortening effect of **embedding** on N1 and N2 (but only for mid to large embedded bigram frequencies)
- Shortening effect of **embedded bigram frequency** on N2
- No effect of **embedded bigram frequency** on N1
- Lengthening effect of embedded bigram frequency effect on N3

Right-branching compounds

- No **direct** effect of **embedding**
 - Shortening effect of **embedded bigram frequency** on N2
 - No effect of **embedded bigram frequency** on duration of N3
 - Lengthening effect of embedded bigram frequency effect for N1
-
- Covariates have expected effects

Generalizations

- Lengthening effect of embedded bigram frequency on non-embedded constituent
- Shortening effect of embedded bigram frequency on N2,
- But no effect on the other embedded constituent

What does it all mean?

- Mixed results concerning the Embedded Reduction Hypothesis
- More complex picture: Embedded reduction through the back door

New hypothesis

The weaker the morphological boundary within the complex constituent, the more salient the boundary between the complex constituent and the third constituent (as marked by the durational properties of the cross-boundary constituents).

To take home

- Durational properties of complex words tell us something about their internal constituency
- ‘Post-lexical’ phonology needs to be reconceived: phonetic detail reflects morphological structure.
- Phonetic detail of morphologically complex words must have a place in models of speech production

Future research

- Replicate the observed effects for compounds and derived words
- Develop new models of phonology-morphology interaction

Thank you very much for your attention!

Acknowledgements

- We thank the *Strategische Forschungsförderfonds* of Heinrich-Heine-Universität Düsseldorf for financial support.