



The phonetics of derivation

Segmentability effects on the acoustic duration of affixed words

Simon David Stein
Ingo Plag

DFG Deutsche
Forschungsgemeinschaft
FOR2373



Frequency and duration

Lexical frequency

How often does a linguistic unit occur in a language?

higher



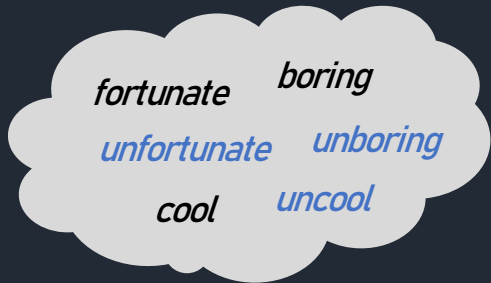
Acoustic duration

How long do we pronounce linguistic units?

shorter

Storage in the mental lexicon

Whole-word storage

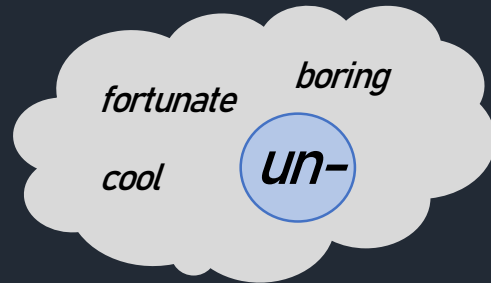


complex words are stored unanalyzed



durations will be shorter the higher the **word frequency**

Compositional models

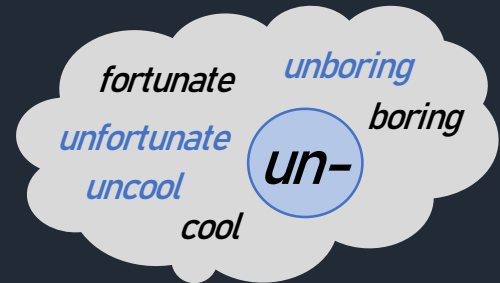


morphemes are stored separately



durations will be shorter the higher the **base frequency**

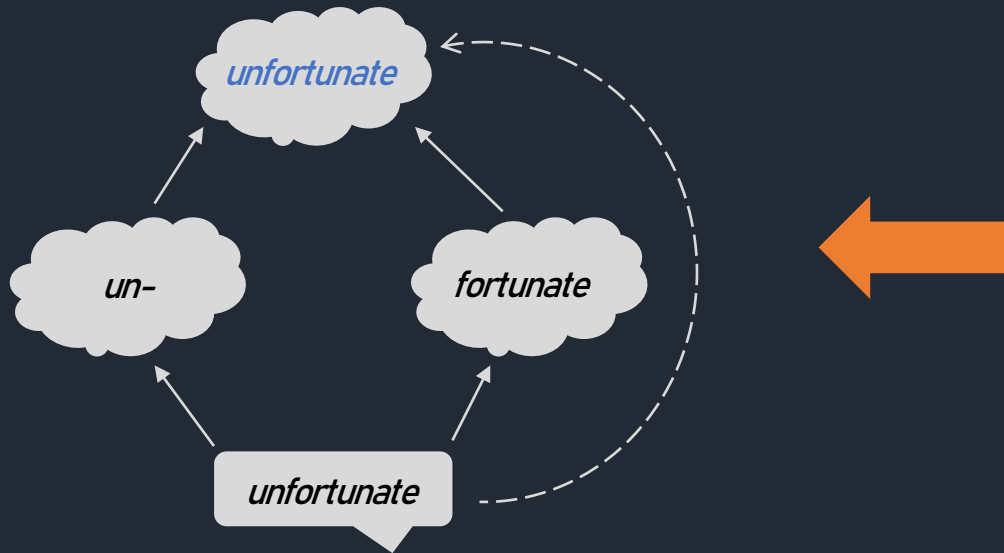
Dual-route models



both morphemes and complex words are stored

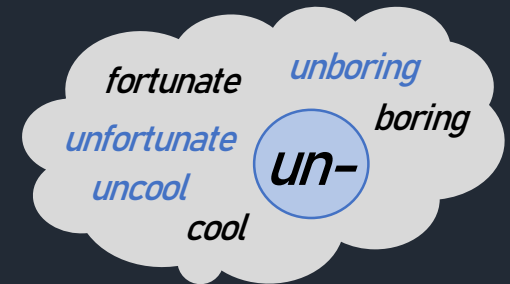


durations will be shorter the lower the **relative frequency**



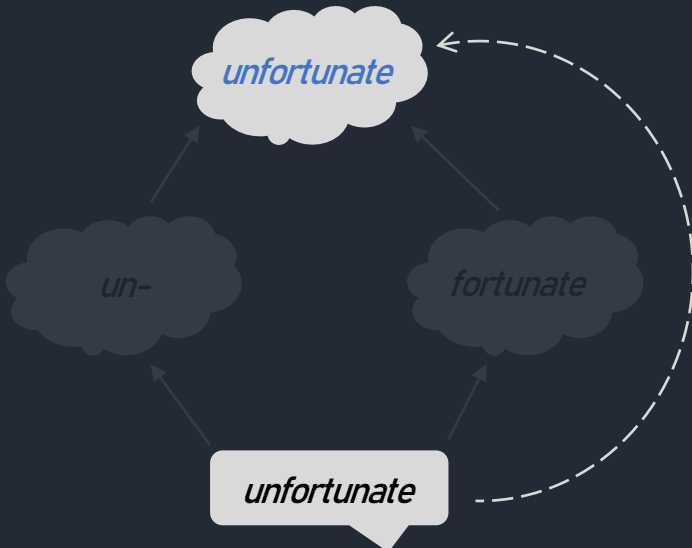
adapted from
Hay 2001: 1045

Dual-route models



both morphemes and
complex words are stored

durations will be shorter
 the lower the
 relative frequency



adapted from Hay 2001: 1045

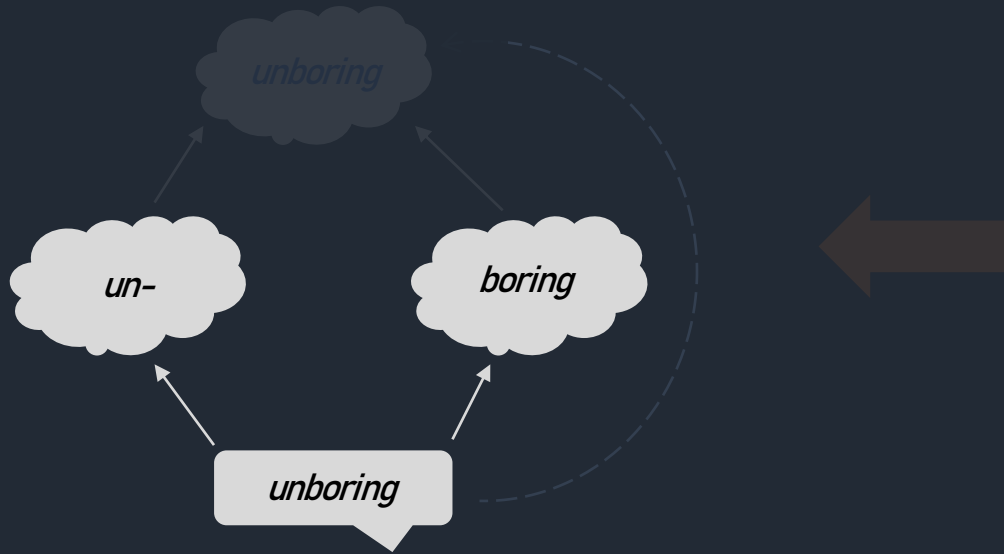
Word	Frequency	Segmentability	Prediction
fortunate	6000	low	shorter duration
unfortunate	6915		
boring	7483	high	longer duration
unboring	4		

Dual-route models



both morphemes and complex words are stored

durations will be shorter the lower the relative frequency



adapted from Hay 2001: 1045

Word	Frequency	Segmentability	Prediction
fortunate	6000	low	shorter duration
unfortunate	6915		
boring	7483	high	longer duration
unboring	4		

Dual-route models



both morphemes and complex words are stored

durations will be shorter the lower the relative frequency

Previous research

Caselli et al. 2016

- › inflectional suffixes *-ing*, *-ed*, and *-s*
- › evidence for both whole-word storage and composition
 - › higher base frequency → shorter word duration
 - › higher word frequency → shorter word duration

Hay 2003, 2007

- › segmentability effects for *un-* and *-ly*

Plag and Ben Hedia 2018

- › segmentability effects for *un-* and *dis-*
- › null effects for negative *in-*, locative *in-*, and *-ly*

**Contradictory evidence:**

Why do the frequency measures sometimes show and sometimes not show effects?

Present study

Hypothesis 1

Higher word frequency → shorter duration of word, base, and affix

Hypothesis 2

Higher base frequency → shorter duration of word, base, and affix

Hypothesis 3

Higher relative frequency → longer duration of word, base, and affix
≈ more segmentability

Data and measurement

Data collection

- › AudioBNC
- › Forced Alignment
- › Praat textgrids
- › manual cleaning of results

Affixes N

<i>-ness</i>	363	<i>pre-</i>	123
<i>-less</i>	216	<i>dis-</i>	689
<i>-wise</i>	289	<i>un-</i>	960
<i>-ize</i>	476	<i>in-</i>	342
<i>-ation</i>	3979		

Modeling

- › multiple linear regression in R using lm-function
- › variable transformations
- › trimming of datasets
- › backwards exclusion of non-significant variables

Responses

- › word duration
- › affix duration
- › base duration
- › **separate models for durations and frequencies: 81 models**

Predictors

- › word frequency
- › base frequency
- › relative frequency

Covariates

- › speech rate
- › number of syllables
- › biphone probability sum
- › bigram frequency

Frequency and segmentability effects

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency	green	white	green	white	white	white	green	white	green
base frequency	white	white	white	green	white	green	white	white	green
relative frequency	green	white	green	blue	white	blue	green	white	green
affix	-less			pre-			-wise		
word frequency	white	white	white	white	green	white	white	white	green
base frequency	white	white	white	white	white	white	white	white	white
relative frequency	white	white	white	white	green	white	white	white	green
affix	dis-			un-			in-		
word frequency	white	green	white	green	green	white	white	white	white
base frequency	white	white	white	white	white	white	white	white	white
relative frequency	white	white	white	white	white	white	white	white	white

■ p < .001 expected direction
■ p < .001 unexpected direction

Frequency and segmentability effects

duration	word	affix	base	word	affix	base	word	affix	base
affix		-ness			-ize			-ation	
word frequency	Green	Green					Green	Green	
base frequency									
relative frequency	Green	Green					Green	Green	
affix		-less			pre-			-wise	
word frequency				Green				Green	
base frequency									
relative frequency				Green				Green	
affix		dis-			un-			in-	
word frequency		Dark Green		Dark Green	Dark Green				
base frequency									
relative frequency									

■ p < .001 expected direction
■ p < .001 unexpected direction

Prefixes vs. suffixes

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency	green	yellow	green	yellow	yellow	yellow	green	yellow	green
base frequency	yellow	yellow	yellow	green	yellow	green	yellow	yellow	green
relative frequency	green	yellow	green	light blue	yellow	light blue	green	yellow	green
affix	-less			pre-			-wise		
word frequency	yellow	yellow	yellow	dark grey	dark grey	dark grey	yellow	yellow	green
base frequency	yellow	yellow	yellow	dark grey	dark grey	dark grey	yellow	yellow	yellow
relative frequency	yellow	yellow	yellow	dark grey	dark grey	dark grey	yellow	yellow	green
affix	dis-			un-			in-		
word frequency	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey
base frequency	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey
relative frequency	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey	dark grey

suffixes

- green p < .001
- light blue p < .001

expected direction
unexpected direction

Are the differences related to ... the type of affix?



Prefixes vs. suffixes

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency									
base frequency									
relative frequency									
affix	-less			pre-			-wise		
word frequency									
base frequency									
relative frequency									
affix	dis-			un-			in-		
word frequency									
base frequency									
relative frequency									

prefixes

■ p < .001
■ p < .001

expected direction
 unexpected direction

Are the differences related to ... the type of affix?



Affix length

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency	green	yellow	green	yellow	yellow	yellow	green	yellow	green
base frequency	yellow	yellow	yellow	green	yellow	green	yellow	yellow	green
relative frequency	green	yellow	green	blue	yellow	blue	green	yellow	green
affix	-less			pre-			-wise		
word frequency	yellow	yellow	yellow	yellow	green	yellow	yellow	yellow	green
base frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow
relative frequency	yellow	yellow	yellow	yellow	green	yellow	yellow	yellow	green
affix	dis-			un-			in-		
word frequency	yellow	green	yellow	green	green	yellow	yellow	yellow	yellow
base frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow
relative frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow

■ p < .001
■ p < .001

expected direction
 unexpected direction

Are the differences related to ...

the type of affix?
 the affix length?



Affix length

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency	dark green	grey	dark green	grey	grey	grey	dark green	grey	dark green
base frequency	grey	grey	grey	dark green	grey	dark green	grey	grey	grey
relative frequency	dark green	grey	dark green	grey	grey	grey	dark green	grey	dark green
affix	-less			pre-			-wise		
word frequency	grey	grey	grey	grey	dark green	grey	grey	grey	dark green
base frequency	grey	grey	grey	grey	grey	grey	grey	grey	grey
relative frequency	grey	grey	grey	grey	dark green	grey	grey	grey	dark green
affix	dis-			un-			in-		
word frequency	grey	dark green	grey	light green	light green	light yellow	light yellow	light yellow	light yellow
base frequency	grey	grey	grey	light yellow	light yellow	light yellow	light yellow	light yellow	light yellow
relative frequency	grey	grey	grey	light yellow	light yellow	light yellow	light yellow	light yellow	light yellow

around 100-150 ms

light green p < .001
 light blue p < .001

expected direction
 unexpected direction

Are the differences related to ...



the type of affix?
 the affix length?



Affix length

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency									
base frequency									
relative frequency									
affix	-less			pre-			-wise		
word frequency									
base frequency									
relative frequency									
affix	dis-			un-			in-		
word frequency									
base frequency									
relative frequency									

around 250-300 ms

 p < .001
 p < .001

expected direction
 unexpected direction

Are the differences related to ...

the type of affix?
 the affix length?

✗
 ✗

Manual resegmentation

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency	green	yellow	green	yellow	yellow	yellow	green	yellow	green
base frequency	yellow	yellow	yellow	green	yellow	green	yellow	yellow	green
relative frequency	green	yellow	green	blue	yellow	blue	green	yellow	green
affix	-less			pre-			-wise		
word frequency	yellow	yellow	yellow	yellow	green	yellow	yellow	yellow	green
base frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow
relative frequency	yellow	yellow	yellow	yellow	green	yellow	yellow	yellow	green
affix	dis-			un-			in-		
word frequency	yellow	green	yellow	green	green	yellow	yellow	yellow	yellow
base frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow
relative frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow

■ p < .001
■ p < .001

expected direction
 unexpected direction

Are the differences related to ...

- the type of affix? ✗
- the affix length? ✗
- the segmentation?

Manual resegmentation

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency									
base frequency									
relative frequency									
affix	-less			pre-			-wise		
word frequency									
base frequency									
relative frequency									
affix	dis-			un-			in-		
word frequency									
base frequency									
relative frequency									

- p < .001
- p < .001
- p < .01

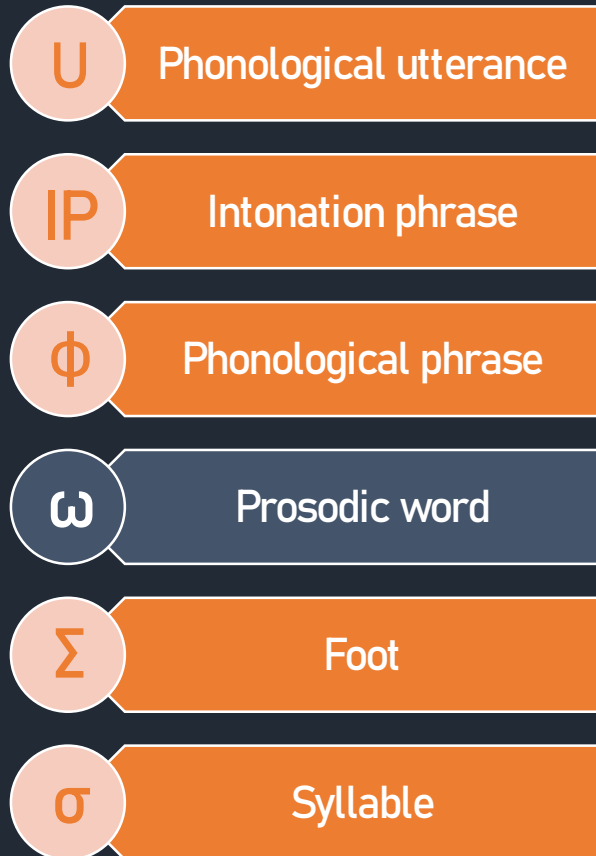
expected direction
 unexpected direction
 weaker effect

Are the differences related to ...

- the type of affix? ✗
- the affix length? ✗
- the segmentation? ✗

Type of prosodic integration

The prosodic hierarchy



Type of prosodic integration



pword-forming



clitic group



integrating



Type of prosodic integration



pword-forming



clitic group





integrating



Type of prosodic integration




duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency									
base frequency									
relative frequency									
affix	-less			pre-			-wise		
word frequency									
base frequency									
relative frequency									
affix	dis-			un-			in-		
word frequency									
base frequency									
relative frequency									

prosodic words

 p < .001
 p < .001

expected direction
 unexpected direction

Are the differences related to ...

- the type of affix? 
- the affix length? 
- the segmentation? 
- prosodic structure?

Type of prosodic integration

duration	word	affix	base	word	affix	base	word	affix	base	
affix	-ness							-ation		
word frequency	green	yellow	green							
base frequency	yellow	yellow	yellow							
relative frequency	green	yellow	green							
affix	-less				pre-			-wise		
word frequency	yellow	yellow	yellow		dark green					
base frequency	yellow	yellow	yellow							
relative frequency	yellow	yellow	yellow		dark green				dark green	
affix	dis-				un-			in-		
word frequency		dark green		dark green	dark green					
base frequency										
relative frequency										

clitic groups

green p < .001
 blue p < .001

expected direction
 unexpected direction



Are the differences related to ...

- the type of affix? x
- the affix length? x
- the segmentation? x
- prosodic structure? x

Type of prosodic integration





duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency									
base frequency									
relative frequency									
affix	-less			pre-			-wise		
word frequency									
base frequency									
relative frequency									
affix	dis-			un-			in-		
word frequency									
base frequency									
relative frequency									

integrating

 p < .001
 p < .001

expected direction
 unexpected direction

Are the differences related to ...

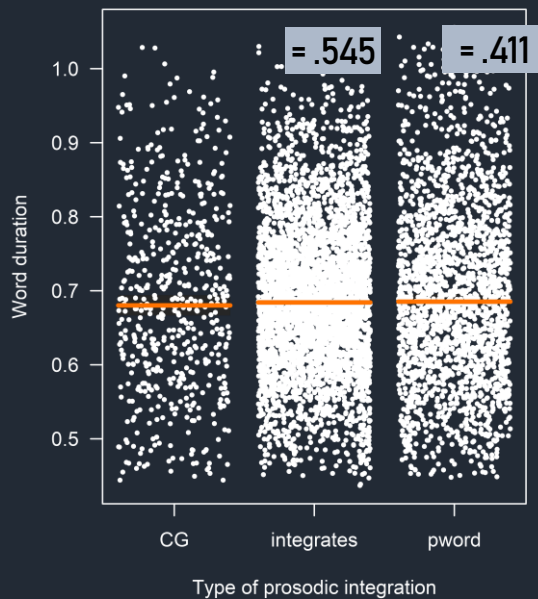
- the type of affix? 
- the affix length? 
- the segmentation? 
- prosodic structure? 

Type of prosodic integration

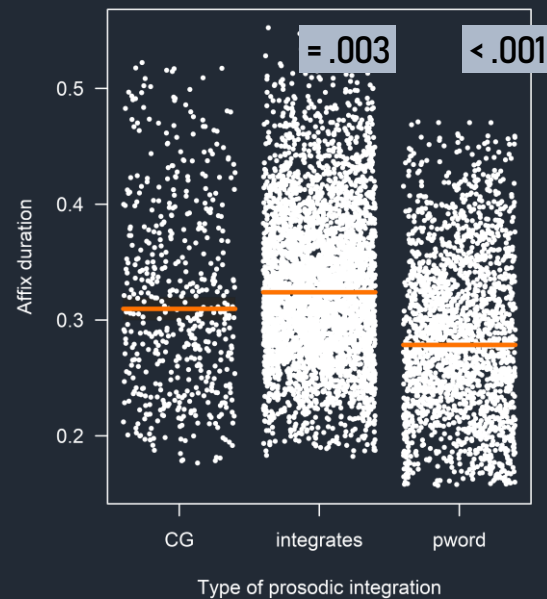
Meta-model including all affixes

- › Additional predictor: **type of prosodic integration**
- › Additional covariate: **number of timing slots**
- › **N = 7441**
- › **This does not support the predictions of pword integration.**

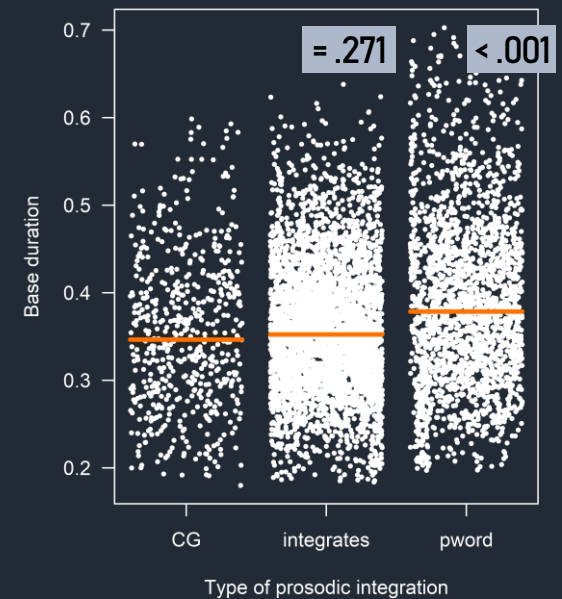
Effect of prosodic category on word duration



Effect of prosodic category on affix duration



Effect of prosodic category on base duration



In sum, we have a mixed picture.

- › Some results are in line with Caselli et al. 2016:
 - › All three frequency measures **can** independently predict duration.
 - › This is evidence for both types of storage in the mental lexicon, as well as for segmentability effects.

- › However, there are also null effects, which require explanation.
 - › So far, we cannot attribute the differences to:
 - › the domain of durational measurement (word, affix, base)
 - › the type of affix (prefix, suffix)
 - › the prosodic category (pword, clitic group, integrating).

Our findings imply that ...

- › morphological structure can at least partly influence the phonetic output.
- › models that prohibit post-lexical access of morphological information (e.g. Kiparsky 1982, Levelt et al. 1999, Bermúdez-Otero 2018) should be revised.
- › we need to investigate further factors that might cause frequency effects to surface or to not surface.

Thank you for listening.

- › Ben Hedia, Sonia. 2018. *Gemination and Degemination in English Affixation: Investigating the Interplay between Morphology, Phonology and Phonetics*. Ph.D. dissertation: Heinrich-Heine-Universität Düsseldorf.
- › Bermúdez-Otero, Ricardo. 2018. Stratal Phonology. In S. J. Hannahs & Anna Bosch (eds.), *Routledge handbook of phonological theory*, 100–143. London: Routledge.
- › Boersma, Paul & David J. M. Weenik. 2014. Praat: Doing phonetics by computer (Version 5.4.04). Computer program. <http://www.praat.org/>.
- › Caselli, Naomi K, Michael K. Caselli, and Ariel M. Cohen-Goldberg. 2016. Inflected words in production: Evidence for a morphologically rich lexicon. *The Quarterly Journal of Experimental Psychology* 69.3: 432–454.

Thank you for listening.

- › Coleman, John, Ladan Baghai-Ravary, John Pybus & Sergio Grau. 2012. *Audio BNC: The audio edition of the Spoken British National Corpus*. Phonetics Laboratory, University of Oxford. <http://www.phon.ox.ac.uk/AudioBNC>.
- › Davies, Mark. 2008–. *The Corpus of Contemporary American English: 450 million words, 1990–present*. <http://corpus.byu.edu/coca/>.
- › Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39.6: 1041–1070.
- › Hay, Jennifer. 2003. *Causes and consequences of word structure*. New York, London: Routledge.
- › Hay, Jennifer. 2007. The phonetics of *un*. In Judith Munat (ed.), *Lexical creativity, texts and contexts*, 39–57. Amsterdam & Philadelphia: John Benjamins.

Thank you for listening.

- › Hildebrandt, Kristine A. 2015. The prosodic word. In John R Taylor (ed.), *The Oxford Handbook of the Word*. Oxford: Oxford University Press.
- › Kiparsky, Paul. 1982. Lexical morphology and phonology. In In-Seok Yang (ed.), *Linguistics in the morning calm: Selected papers from SICOL*, 3–91. Seoul: Hanshin.
- › Levelt, William J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1: 1–38.
- › Plag, Ingo & Sonia Ben Hedia. 2018. The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin & Esme Winter-Froemel (eds.), *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and Ludicity*. Berlin & New York: de Gruyter Mouton.

Thank you for listening.

- › Raffelsiefen, Renate. 1999. Diagnostics for prosodic words revisited: The case of historically prefixed words in English. In Tracy A. Hall & Ursula Kleinhenz (eds.), *Studies of the phonological word*. 133–201. Amsterdam, Philadelphia: Benjamins.
- › Raffelsiefen, Renate. 2007. Morphological word structure in English and Swedish: The evidence from prosody. In Geert Booij, Luca Ducceschi, Bernard Fradin, Ernesto Guevara, Angela Ralli & Sergio Scalise (eds.), *Online Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*, Fréjus, 15–18 September 2005, 209–268.
- › R Core Team 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. <http://www.R-project.org/>.

Thank you for listening.

- › Vitevitch, Michael S., & Luce, Paul A. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36.3: 481–487.

Informativity

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency	green	yellow	green	yellow	yellow	yellow	green	yellow	green
base frequency	yellow	yellow	yellow	green	yellow	green	yellow	yellow	green
relative frequency	green	yellow	green	blue	yellow	blue	green	yellow	green
affix	-less			pre-			-wise		
word frequency	yellow	yellow	yellow	yellow	green	yellow	yellow	yellow	green
base frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow
relative frequency	yellow	yellow	yellow	yellow	green	yellow	yellow	yellow	green
affix	dis-			un-			in-		
word frequency	yellow	green	yellow	green	green	yellow	yellow	yellow	yellow
base frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow
relative frequency	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow	yellow

■ p < .001 expected direction
■ p < .001 unexpected direction

Are the differences related to ...
 the type of affix? ✗
 the affix length? ✗
 the segmentation? ✗
 prosodic structure? ✗
 affix informativity?

Informativity

Measured in two ways:

Semantic information load score

5-point Likert scales coded for:

- › clearness of semantic meaning
- › type of base: free vs. bound root
- › semantic transparency
- › productivity



Affix-specific semantic
segmentability hierarchy

H: The higher the semantic information load, the longer the duration.

Conditional affix probability C_{aff}

Affix probability given preceding word:

SUFFIX EXAMPLE		PREFIX EXAMPLE		
A	B	A	B	C
<i>random</i>	<i>ize</i>	<i>her</i>	<i>pre-</i>	<i>...</i>



$$C_{aff} = \frac{Freq(AB)}{Freq(A)}$$

H: The higher the conditional affix probability, the shorter the duration.

Informativity: Semantic information load score

duration	word	affix	base	word	affix	base	word	affix	base	
affix	-ness							-ation		
word frequency	green	yellow	green							
base frequency	yellow	yellow	yellow							
relative frequency	green	yellow	green							
affix	-less				pre-			-wise		
word frequency										
base frequency										
relative frequency										
affix	dis-			un-			in-			
word frequency		dark green		green	green	yellow				
base frequency				yellow	yellow	yellow				
relative frequency				yellow	yellow	yellow				

high information load

green p < .001
blue p < .001

expected direction
unexpected direction



Are the differences related to ...

- the type of affix? ✗
- the affix length? ✗
- the segmentation? ✗
- prosodic structure? ✗
- affix informativity? ✗

Informativity: Semantic information load score






duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
word frequency									
base frequency									
relative frequency									
affix	-less			pre-			-wise		
word frequency									
base frequency									
relative frequency									
affix	dis-			un-			in-		
word frequency									
base frequency									
relative frequency									

low information load

 p < .001
 p < .001

expected direction
 unexpected direction

Are the differences related to ...

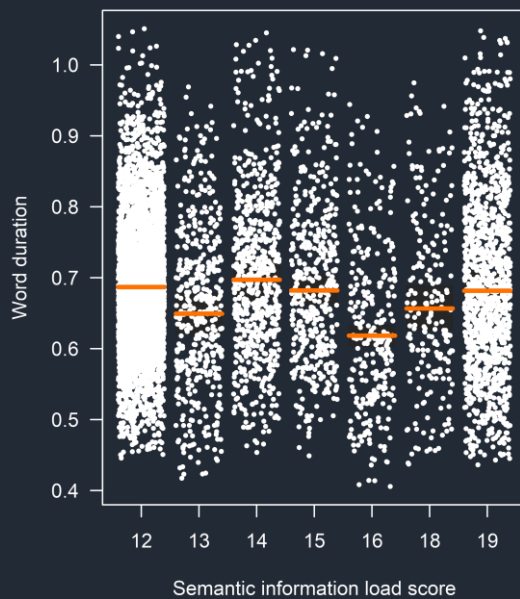
- the type of affix? 
- the affix length? 
- the segmentation? 
- prosodic structure? 
- affix informativity? 

Informativity: Semantic information load score

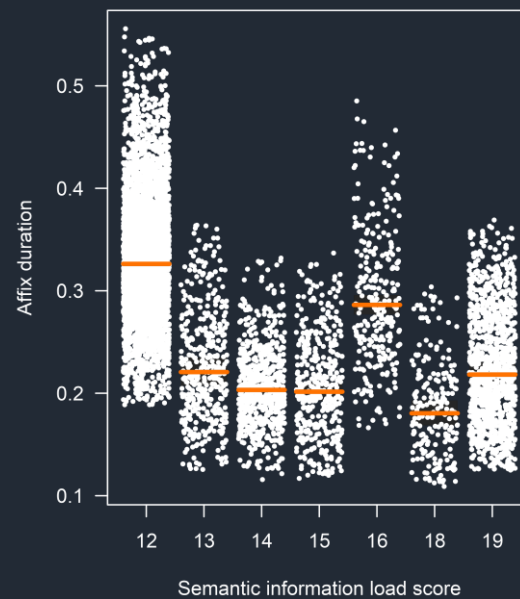
Meta-model including all affixes

- › Additional predictor: **semantic information load score**
- › Additional covariate: **number of timing slots**
- › **N = 7441**
- › **This does not support the predictions of semantic information load.**

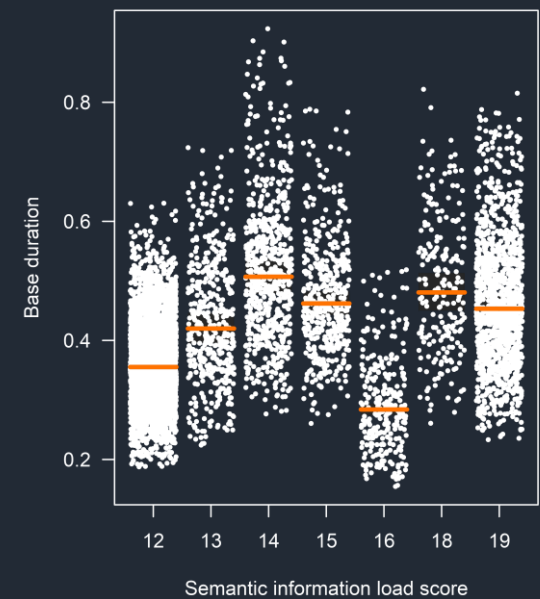
Effect of information load on word duration



Effect of information load on affix duration



Effect of information load on base duration



Informativity: Conditional affix probability

duration	word	affix	base	word	affix	base	word	affix	base
affix	-ness			-ize			-ation		
affix probability									

affix	-less		pre-			-wise		
affix probability								

affix	dis-		un-			in-		
affix probability								

$p < .001$

positive correlation

Are the differences related to ...

the type of affix? ✗

the affix length? ✗

the segmentation? ✗

prosodic structure? ✗

affix informativity? ✗

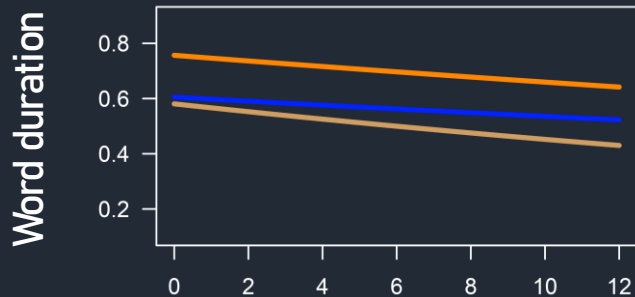
Updated summary

In sum, we have a mixed picture.

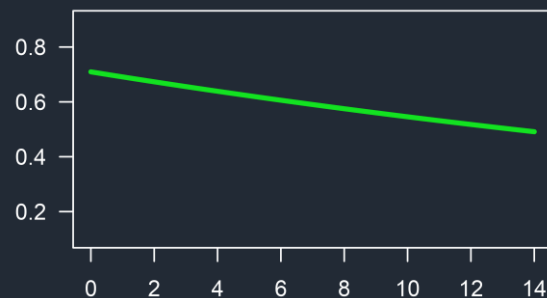
- › Some results are in line with Caselli et al. 2016:
 - › All three frequency measures **can** independently predict duration.
 - › This is evidence for both types of storage in the mental lexicon, as well as for segmentability effects.

- › However, there are also null effects, which require explanation.
 - › So far, we cannot attribute the differences to:
 - › the domain of durational measurement (word, affix, base)
 - › the type of affix (prefix, suffix)
 - › the prosodic category (pword, clitic group, integrating)
 - › **the informativity of the affix (information load, probability).**

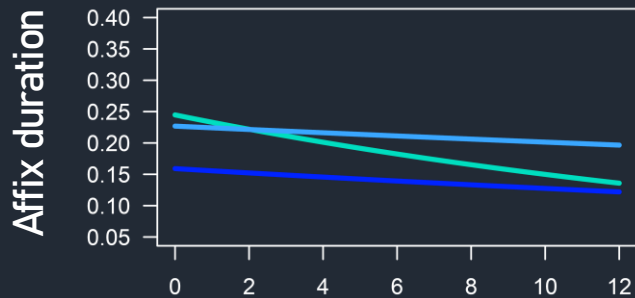
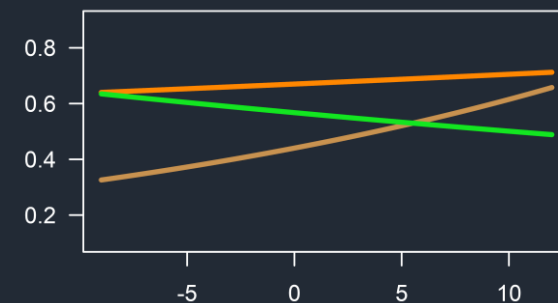
Log word frequency



Log base frequency

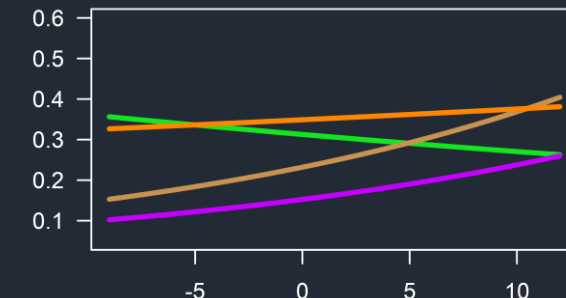
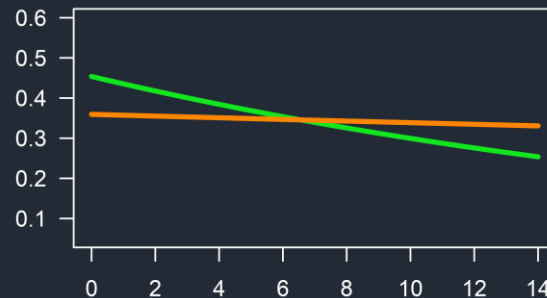
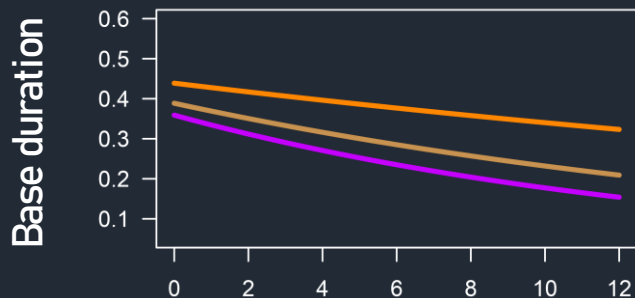
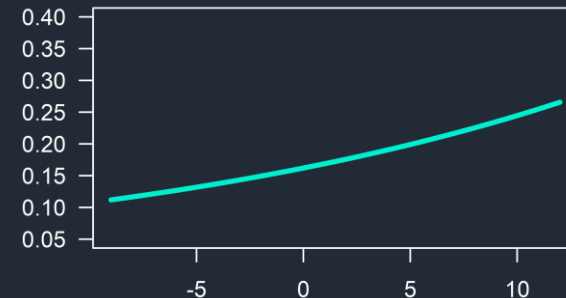


Log relative frequency



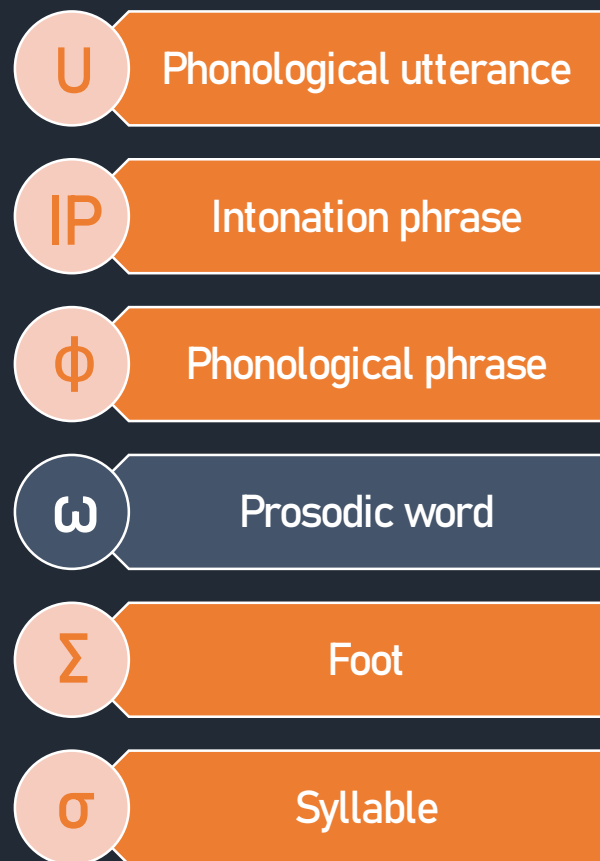
■ -ation ■ -ize ■ pre-
■ -ness ■ dis- ■ -wise
■ un-

Effect size comparison between affixes. Effects with $p > .001$ omitted.



Prosodic word diagnostics

The prosodic hierarchy



Some pword-diagnostics

- > onset or coda conditions, LOI-violations
- > ambisyllabicity
- > stress and relative prominence
- > trisyllabic laxing, vowel reduction
- > minimal word requirements
- > compositionality, type of base

Morpho-prosodic alignment

- > A morpheme **cannot** include multiple pwords, but a pword **can** include multiple morphemes.

Type of prosodic integration



pword-forming



clitic group



integrating

