



# Lexical storage and morphological segmentability in speech production

New evidence from English derivational affixes

Simon David Stein Ingo Plag







# Frequency and duration



# Frequency and duration

# Lexical frequency

How often does a word occur in a language?

## Acoustic duration

How long do we pronounce linguistic units?





## Frequency and duration

# Lexical frequency

How often does a word occur in a language?

#### **Acoustic duration**

How long do we pronounce linguistic units?

## Usual assumption:

The higher the frequency, the shorter the duration of linguistic units such as words, bases, and affixes.



# Storage in the mental lexicon



## Whole-word storage





complex words are stored unanalyzed



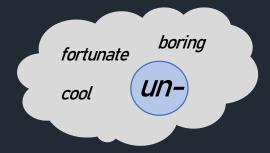
## Whole-word storage





complex words are stored unanalyzed

## Compositional models





morphemes are stored separately



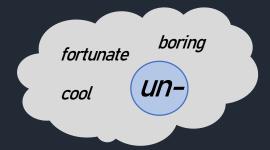
#### Whole-word storage





complex words are stored unanalyzed

## Compositional models





morphemes are stored separately

#### Dual-route models





both morphemes and complex words are stored



## Storage in the mental lexicon

#### Whole-word storage



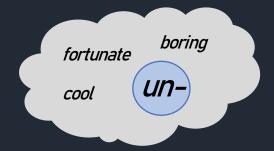


complex words are stored unanalyzed



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





## Segmentability

#### Dual-route models



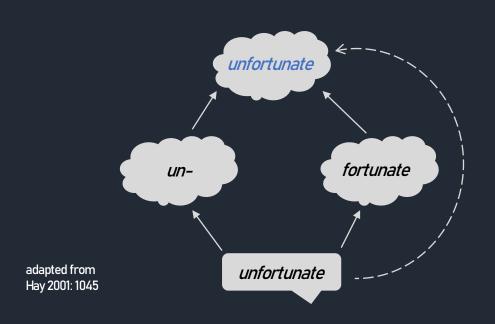


both morphemes and complex words are stored





## Segmentability



Dual-route models



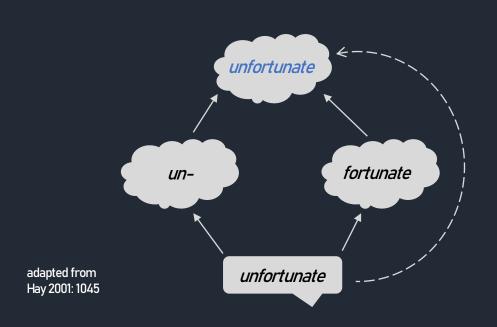


both morphemes and complex words are stored





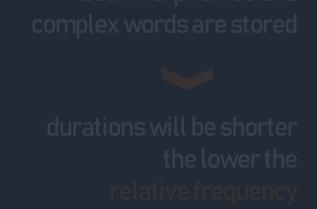
## Segmentability





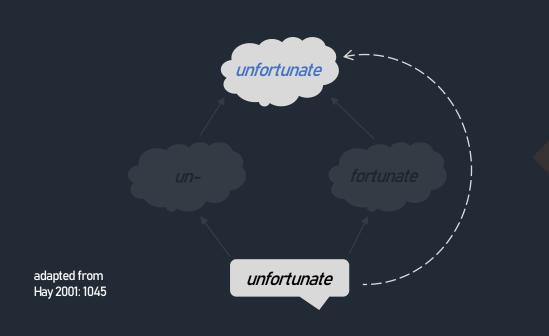
Word	Frequency	S
fortunate	6000	
unfortunate	6915	
boring	7483	
unboring	4	

Segmentability	Prediction
low	shorter duration
high	longer duration





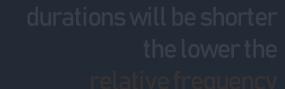
## Segmentability





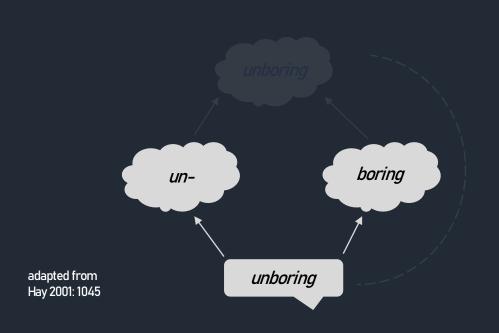
Word	Frequency	Segmentability	Prediction	
fortunate	6000	I.v.	shorter	
unfortunate	6915	low	duration	
	4			

complex words are stored





## Segmentability





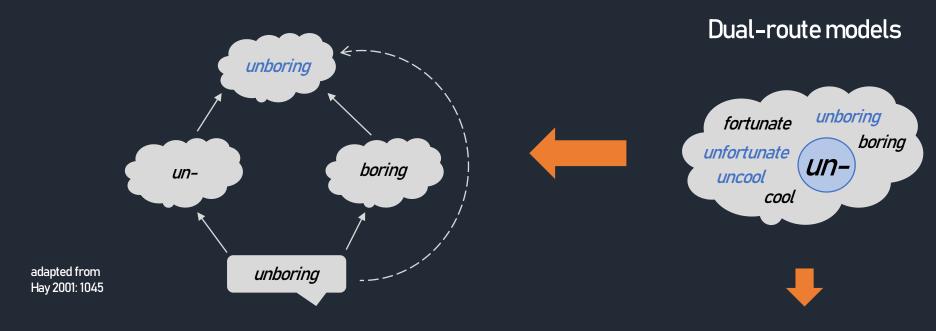
Word	Frequency	Segmentability	Prediction	
boring	7483	himb	longer duration	
unboring	4	high		

complex words are stored

the lower the relative frequency



## Segmentability



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

both morphemes and complex words are stored





#### Whole-word storage



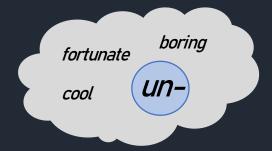


complex words are stored unanalyzed



durations will be shorter the higher the word frequency

## Compositional models



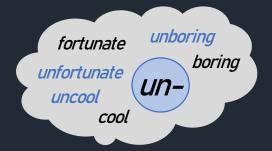


morphemes are stored separately



durations will be shorter the higher the base frequency

#### Dual-route models





both morphemes and complex words are stored





# Previous research



#### 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



#### 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* 



#### 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



#### 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



# Present study

# Hypothesis1

Higher word frequency → shorter duration



# Present study

# Hypothesis1

Higher word frequency → shorter duration

# Hypothesis 2

Higher base frequency → shorter duration



## Present study

# Hypothesis1

Higher word frequency → shorter duration

## Hypothesis 2

Higher base frequency → shorter duration

## Hypothesis 3

≈ more segmentability

Higher relative frequency - longer duration



## Present study

## Hypothesis1

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 collection

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



Da	ata collection	Affixes	N
>	AudioBNC	-ness	364
>	Forced Alignment	-less	216
>	Praat textgrids	pre-	118
>	manual cleaning	-wise	289
	of results	-ize	476
		-ation	3979



Da	nta collection	Affixes	N
>	AudioBNC	-ness	364
>	Forced Alignment	-less	216
>	Praat textgrids	pre-	118
>	manual cleaning	-wise	289
	of results	-ize	476
		-ation	3979

## Modeling

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



Data collection	Affixes	N	Modeling
<ul> <li>AudioBNC</li> <li>Forced Alignment</li> <li>Praat textgrids</li> <li>manual cleaning         of results</li> </ul>	-ness -less pre- -wise -ize -ation	364 216 118 289 476 3979	<ul> <li>multiple linear regression in R using lm-function</li> <li>variable transformations</li> <li>trimming of datasets</li> <li>backwards exclusion of non-significant variables</li> </ul>

# Responses

- word duration
- > affix duration
- base duration



#### Data collection

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

#### Affixes N

- -*ness* 364
- -*less* 216
- *pre* 118
- -*wise* 289
- *-ize* 476
- *-ation* 3979

#### Modeling

- multiple linear regression in Rusing lm-function
- > variable transformations
- trimming of datasets
- backwards exclusion of non-significant variables

#### Responses

- word duration
- affix duration
- base duration

#### **Predictors**

- word frequency
- base frequency
- relative frequency



#### Data collection

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

#### Affixes N

- *-ness* 364
- *-less* 216
- *pre* 118
- -*wise* 289
- *-ize* 476
- *-ation* 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

#### **Predictors**

- > word frequency
- base frequency
- relative frequency

#### Covariates

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



#### Data collection

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

#### Affixes

N

- -ness 364
- -*less* 216 *pre* 118
- *wise* 289
- *-ize* 476
- *-ation* 3979

#### Modeling

- multiple linear regression in Rusing 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: 54 models

#### **Predictors**

- > word frequency
- base frequency
- relative frequency

#### Covariates

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

# Results



# Frequency and segmentability effects



# Frequency and segmentability effects

affix	pre-		
duration	word	affix	base
word frequency			
base frequency			
relative frequency			

p < .001 expected direction



affix	pre-			-ness			
duration	word	affix	base	word	affix	base	
word frequency							
base frequency							
relative frequency							

p < .001 expected direction



affix	pre-			-ness			-ize		
duration	word	affix	base	word	affix	base	word	affix	base
word frequency									
base frequency									
relative frequency									

p < .001

p < .001

expected direction unexpected direction



affix	pre-			-ness			-ize		
duration	word	affix	base	word	affix	base	word	affix	base
word frequency									
base frequency									
relative frequency									

affix	-wise		
duration	word	affix	base
word frequency			
base frequency			
relative frequency			

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



affix	pre-			-ness			-ize		
duration	word	affix	base	word	affix	base	word	affix	base
word frequency									
base frequency									
relative frequency									

affix	-wise			-less			
duration	word	affix	base	word	affix	base	
word frequency							
base frequency							
relative frequency							

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



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

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



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

F F

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

Are the differences related to ...



#### Prefixes vs. suffixes

affix	pre-								
duration	word	affix	base						
word frequency									
base frequency									
relative frequency									

word frequency									
base frequency									
relative frequency									

p ·

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

Are the differences related to ...

the type of affix?



#### Prefixes vs. suffixes

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



p < .001 p < .001

relative frequency

word frequency

base frequency

expected direction unexpected direction

Are the differences related to ...

the type of affix?



#### Affix length

	pre-	pre-								
	word	affix	base							
word frequency										
base frequency										
relative frequency										
				-less						
				word	affix	base				
word frequency										
base frequency										

Are the differences related to ...

the type of affix?

the affix length?

p < .001

p < .001

expected direction

unexpected direction



#### Affix length

word frequency								
base frequency								
relative frequency								

	-wise			-less		-ation			
	word	affix	base	word		word	affix	base	
word frequency									
base frequency									
relative frequency									

p <

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

Are the differences related to ...

the type of affix? the affix length?





x

the type of affix?

the affix length?

the segmentation?

#### Manual resegmentation

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

Are the differences related to ...

p < .001

p < .001

expected direction

unexpected direction



#### Manual resegmentation

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



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

Are the differences related to ...

the type of affix? the affix length? the segmentation?





#### The prosodic hierarchy

- U Phonological utterance
- Intonation phrase
- Phonological phrase
- ( Prosodic word
- Foot
- Syllable



#### The prosodic hierarchy

- U Phonological utterance
- IP Intonation phrase
- Ф Phonological phrase
- **ω** Prosodic word
- Foot
- Syllable

#### 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



#### The prosodic hierarchy

- U Phonological utterance
- IP Intonation phrase
- Ф Phonological phrase
- $(\omega)$  Prosodic word
- Foot
- **o** Syllable

#### 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.

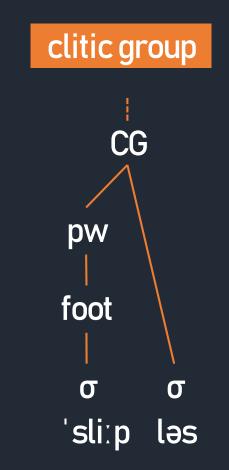


# pword-forming

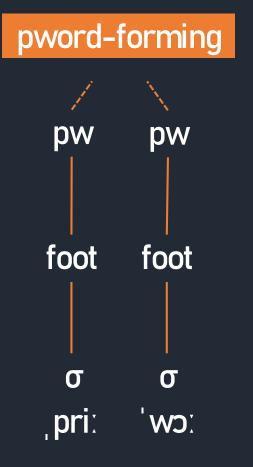


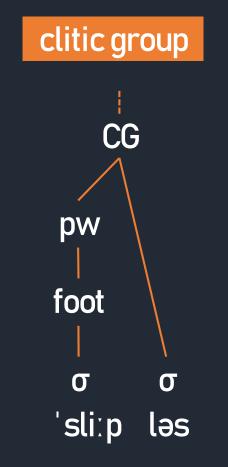






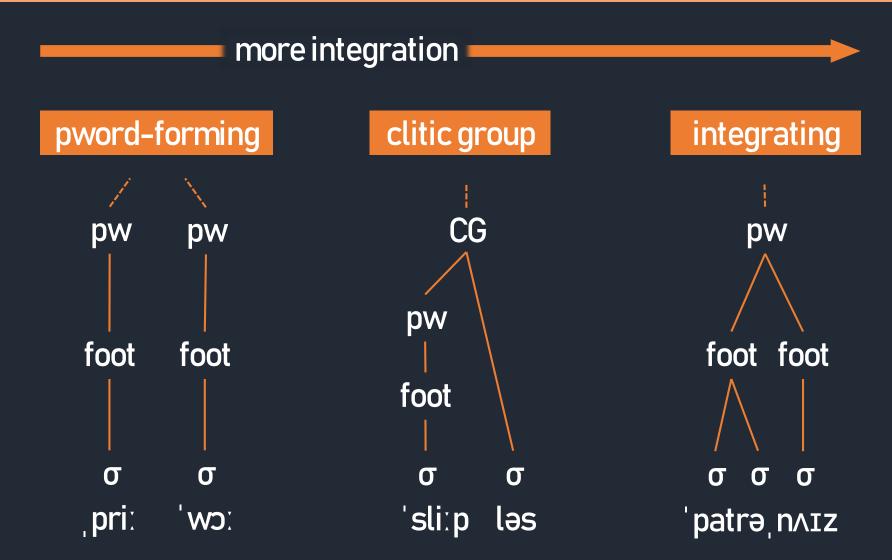




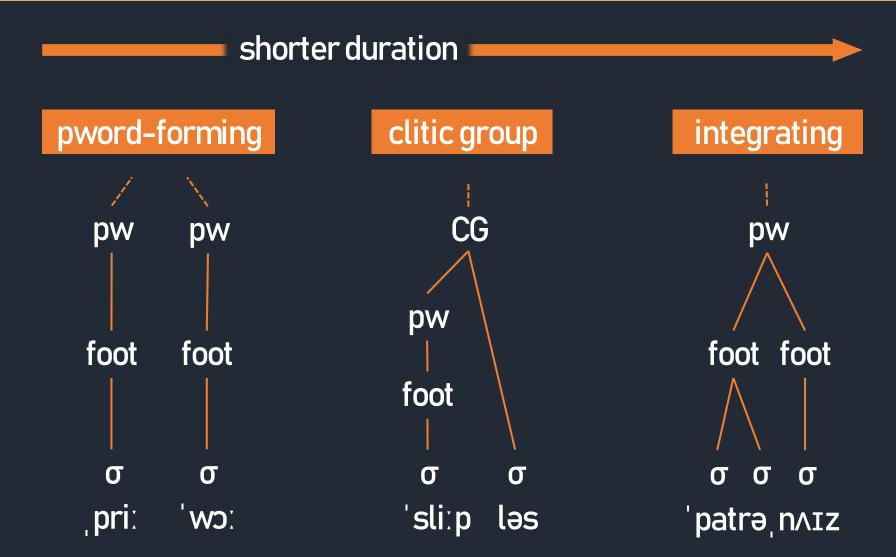






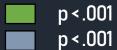








affix	pre-			-ness			-ize			
duration	word	affix	base	word	affix	base	word	affix	base	
word frequency										
base frequency										
relative frequency										
affix	-wise			-less			-ation			
affix duration	-wise	affix	base	-less word	affix	base	-ation	affix	base	
		affix	base		affix	base		affix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



expected direction unexpected direction

Are the differences related to ...

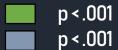
the type of affix? the affix length? the segmentation? prosodic structure?

×



#### Type of prosodic integration

category	prosodic word			clitic group			integrates			
affix	pre-			-ness			-ize			
duration	word	affix	base	word	affix	base	word	affix	base	
word frequency										
base frequency										
relative frequency										
affix	-wise			-less			-ation			
duration	word	affix	base	word	affix	base	word	affix	base	
word frequency										
base frequency										
relative frequency										



expected direction unexpected direction

Are the differences related to ...

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

×

×



the segmentation? prosodic structure?

category	prosoc	lic word		clitic group			integrates			
affix	pre-									
duration	word	affix	base							
word frequency										
base frequency										
relative frequency										
affix	-wise									
duration	word	affix	base							
word frequency										
base frequency										
relative frequency										
	cted direct pected dir		Are	the differences related to			the type of affix?			



#### Type of prosodic integration

prosodic word			clitic g	roup		integrates			
			-ness						
			word	affix	base				
			-less						
			word	affix	base				
	pre- word -wise	pre- word affix  -wise	pre- word affix base	preness word affix base word  -wise -less	preness word affix base word affix  -wise -less	reness word affix base word affix base  -wise -less	re- word affix base word affix base word  -wise -less -ation	reness -ize  word affix base word affix base word affix  -wise -less -ation	



p < .001

expected direction unexpected direction Are the differences related to ...

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

x



#### Type of prosodic integration

category	prosodic word			clitic group			integrates			
							-ize			
							word	affix	base	
word frequency										
base frequency										
relative frequency										
							-ation			
							word	affix	base	
word frequency										
base frequency										



p < .001

p < .001

expected direction unexpected direction

Are the differences related to ...

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

×

×



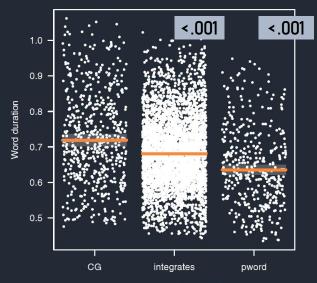
- > Additional predictor: type of prosodic integration
- > Additional covariate: number of timing slots
- $\rightarrow$  N=5450



#### Meta-model including all affixes

- > Additional predictor: type of prosodic integration
- > Additional covariate: number of timing slots
- N = 5450

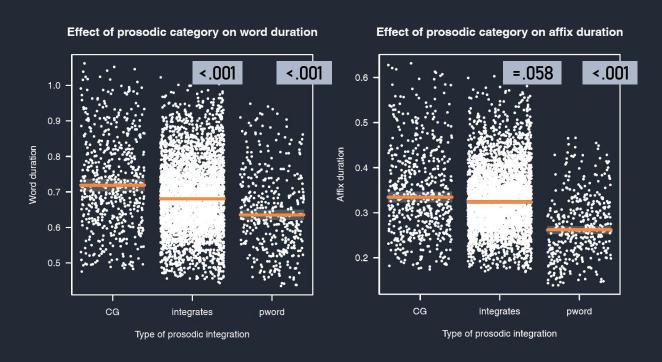
#### Effect of prosodic category on word duration



Type of prosodic integration

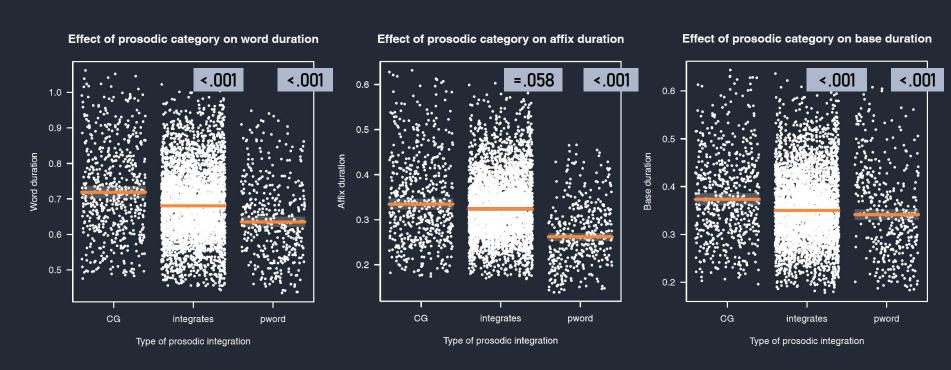


- > Additional predictor: type of prosodic integration
- > Additional covariate: number of timing slots
- > N=5450



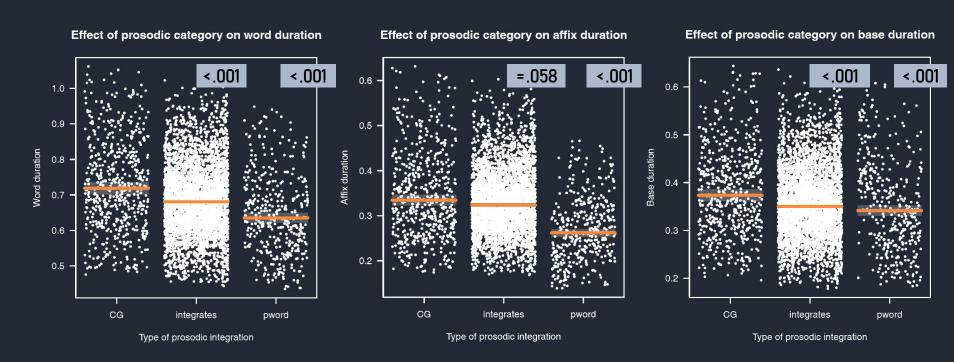


- > Additional predictor: type of prosodic integration
- > Additional covariate: number of timing slots
- > N = 5450





- Additional predictor: type of prosodic integration
- > Additional covariate: number of timing slots
- > N = 5450
- > This does not support the predictions of pword integration.



# Conclusion



# Summary

#### Conclusion



#### 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.



#### 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).

# Conclusion



# Discussion

#### Conclusion



#### Discussion

Our findings imply that ...

> morphological structure can at least partly influence the phonetic output.

### Conclusion



#### Discussion

### 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.

#### Conclusion



#### Discussion

### 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.





- > 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.



- 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.



- 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.



- > 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/.

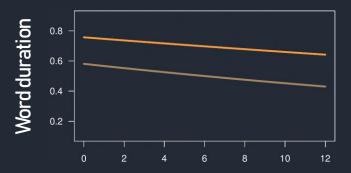


> Vitevitch, Michael S., & Luce, Paul A. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36.3: 481–487.



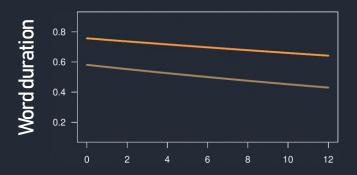


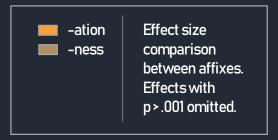
# Log word frequency



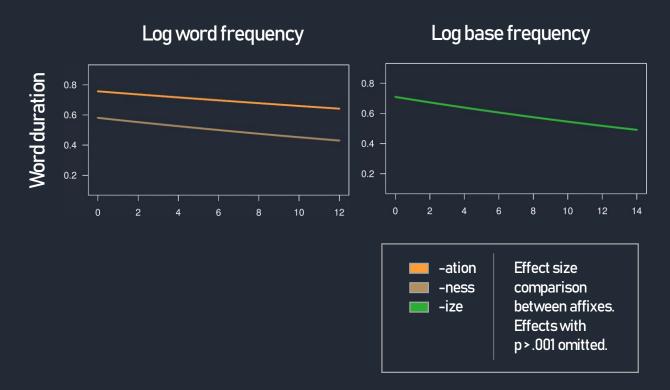


# Log word frequency

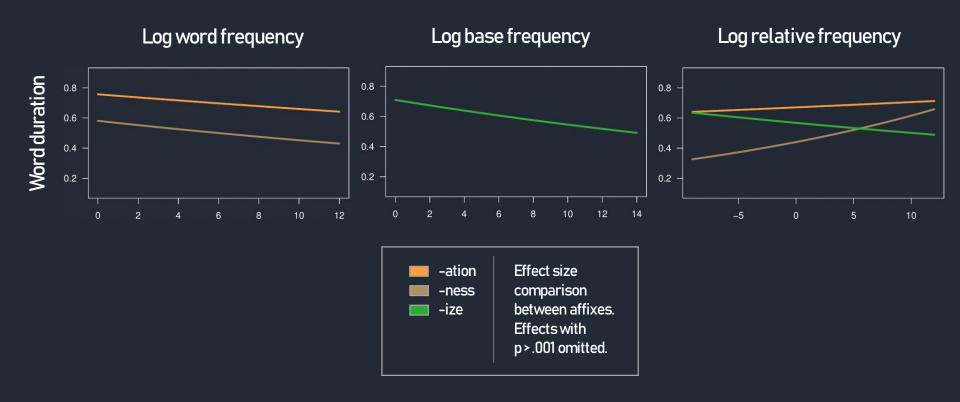




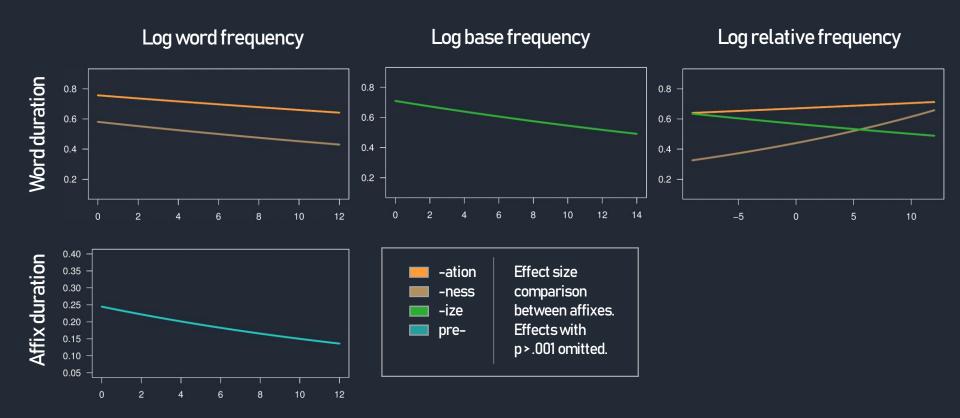




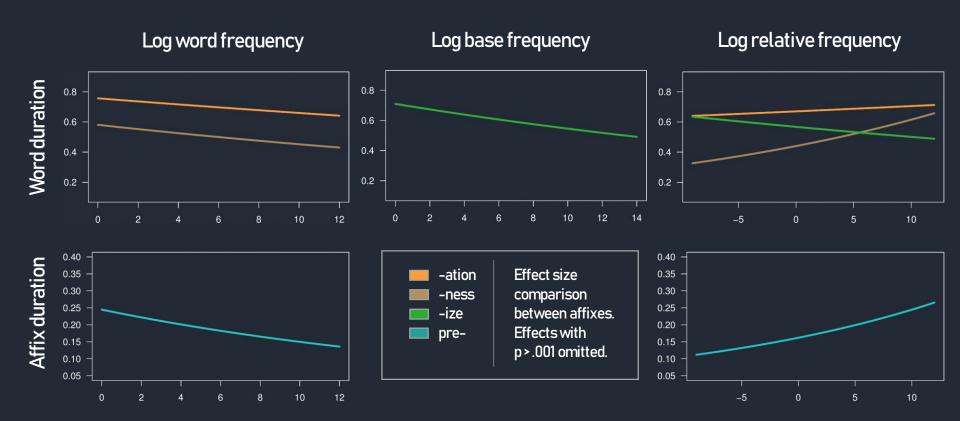
















10







