Heinrich Heine
Universität Düsseldorf

DFG Research Unit FOR 2373
Spoken Morphology: Phonetics and phonology of complex words

Morphological boundaries and stem duration in English

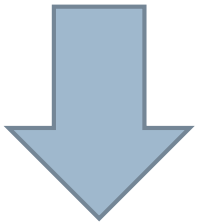
Paradigm uniformity effects in conversational speech

Marie Engemann, Ingo Plag & Julia Zimmermann

P&P Conference Düsseldorf, 27th September 2019

Introduction

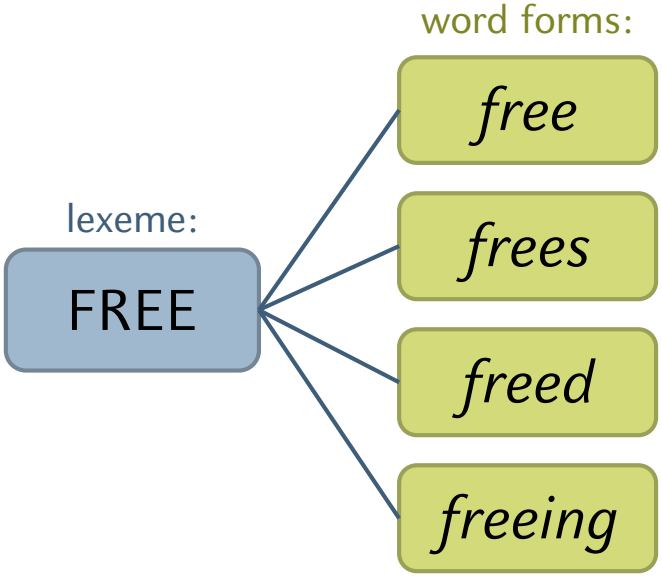
- phonetic duration may be affected by morphological structure
  - different types of word-final [s, z] have different durations (Plag, Homann & Kunter 2017; Seyfarth et al. 2017; Tomaschek et al. 2019; Plag et al. 2019)
  - stems of words ending in [s, z] also have longer durations if these are complex words (Seyfarth et al. 2017)



What is a possible cause of these differences?

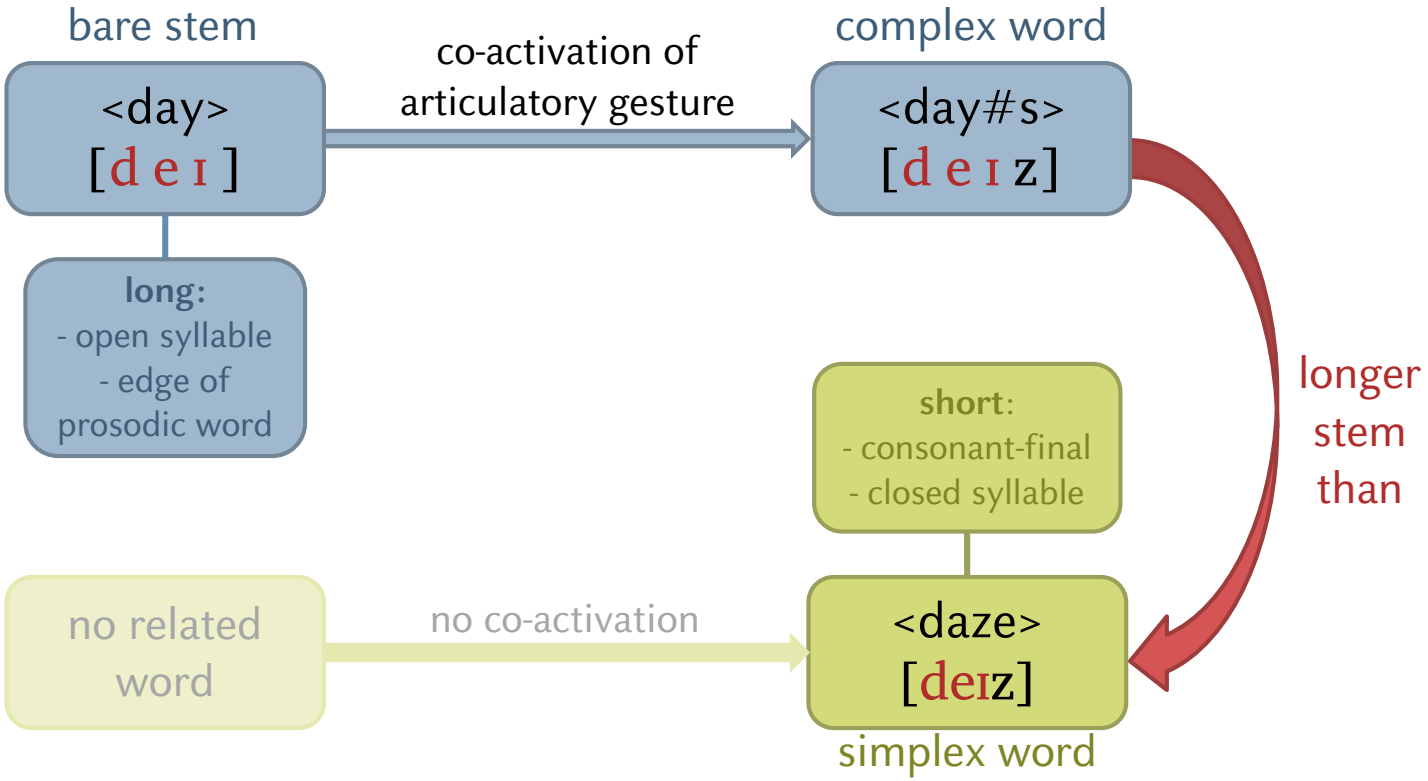
Paradigm uniformity (1/4)

- paradigm uniformity effect may cause lengthening of complex stems (Seyfarth et al. 2017)
- What is paradigm uniformity?
- morphological paradigm consists of set of morphologically related forms
  - e.g. inflectional paradigms contain all word forms of a lexeme



Paradigm uniformity (2/4)

- paradigm uniformity effects arise if morphologically complex form is influenced by paradigm members



Paradigm uniformity (3/4)

- note on terminology:
  - <day> [d e ɪ] bare stem
  - <day#s> [d e ɪ z] pural stem
- we will refer to the corresponding string of sounds in monomorphemic words as ‘monomorphemic stem’
  - <daze> [deɪz] monomorphemic stem

Paradigm uniformity (4/4)

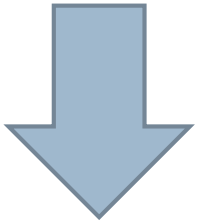
- Seyfarth et al. (2017) found that categorically, stems of complex words ending in [s, z] are longer than stems of simplex words
  - *days* is longer than *daze*
  - categorical paradigm uniformity
- they further predicted that a stronger representation (~higher frequency) of the stem leads to an even longer duration
  - found no relation between frequency and duration
  - gradient paradigm uniformity

The present study

- we expand Seyfarth et al. (2017) because:
  - their results only partly confirm paradigm uniformity
  - they did not discuss their null results for gradient paradigm uniformity
- additionally, we address these problems:
  - they used phonetically matched dialogues with embedded homophones to emulate natural speech
    - we use natural speech from a corpus
  - they recruited speakers of North American English
    - we use New Zealand English

Hypothesis

- stems of plural words are longer than stems of non-morphemic words before [z]
  - a) in corpus data
  - b) for New Zealand English



categorical paradigm uniformity effect

Data

- QuakeBox Corpus (Walsh et al. 2013) recorded in Christchurch, New Zealand
  - monologues in which speakers share their experiences in the earthquakes in 2010/2011
- dataset was originally used for the study of the durations of word-final S (Zimmermann 2016)
- we use subset to investigate paradigm uniformity (487 tokens, 74 types)
  - included only words that are
    - monosyllabic and ending in /z/
    - monomorphemic or plural
    - have final /z/ preceded by vowel

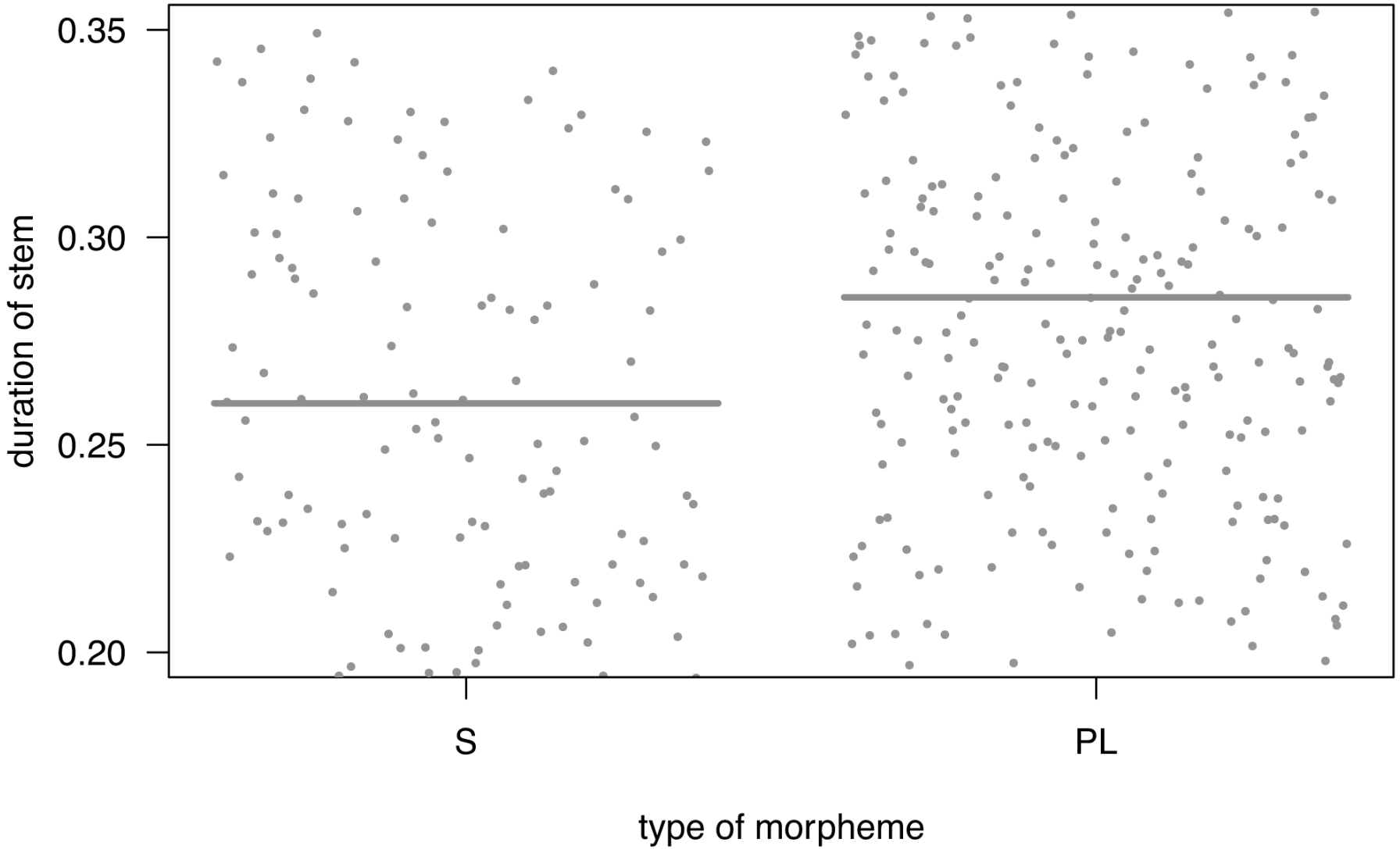
Statistical modeling (1/2)

- multiple linear regression modeling in R (R Core Team 2015)
- response variable: **stem duration**
- predictor variable: **type of morpheme**
  - monomorphemic or plural
- 6 covariates:
  - number of phonemes
  - word form frequency
  - speech rate
  - position within sentence
  - voicing ratio
  - age of speaker

Results: Categorical Paradigm Uniformity (1/4)

- significant effect of type of morpheme on duration of the stem in the expected direction ($p < 0.0005$)
- plural stems are about 25 milliseconds longer than stems of monomorphemic words

Results: Categorical Paradigm Uniformity (2/4)



Results: Categorical Paradigm Uniformity (3/4)

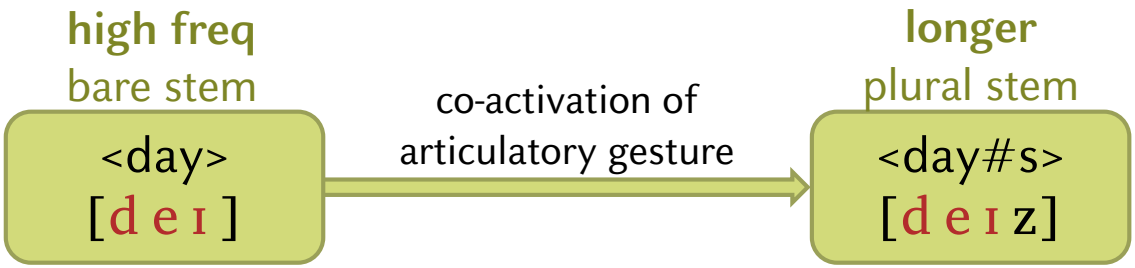
- covariates show significant effects and behave in expected direction
  - higher word form frequency = shorter stem
  - faster speech rate = shorter stem
  - higher number of phonemes = longer stem
  - higher voicing ratio = shorter stem
  - older speakers speak slower

Results: Categorical Paradigm Uniformity (4/4)

- we find support for categorical paradigm uniformity
- What is the relation between frequency and duration?
  - our results: higher word form frequency = shorter duration
  - Seyfarth et al. (2017) predicted: stronger representation (~higher frequency) of stem leads to even longer duration

Hypothesis 2

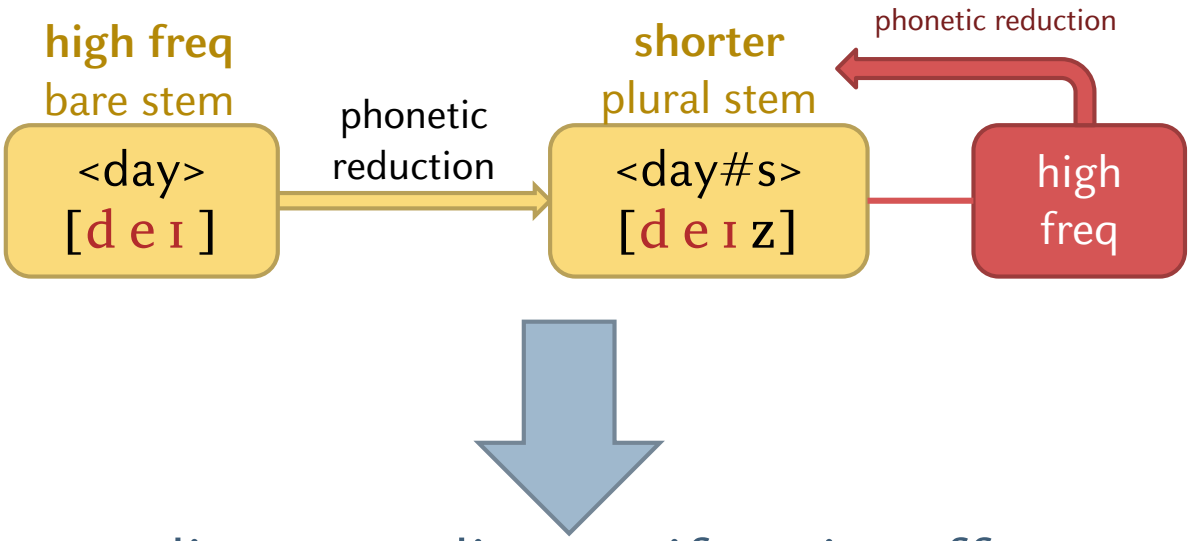
- higher frequency of bare stem leads to longer duration of plural stem



gradient paradigm uniformity effect
due to **strength of activation**

Alternative: Hypotheses 3

- a) higher frequency of bare stem leads to shorter duration of plural stem
- b) higher frequency of plural word-form leads to shorter duration of plural stem



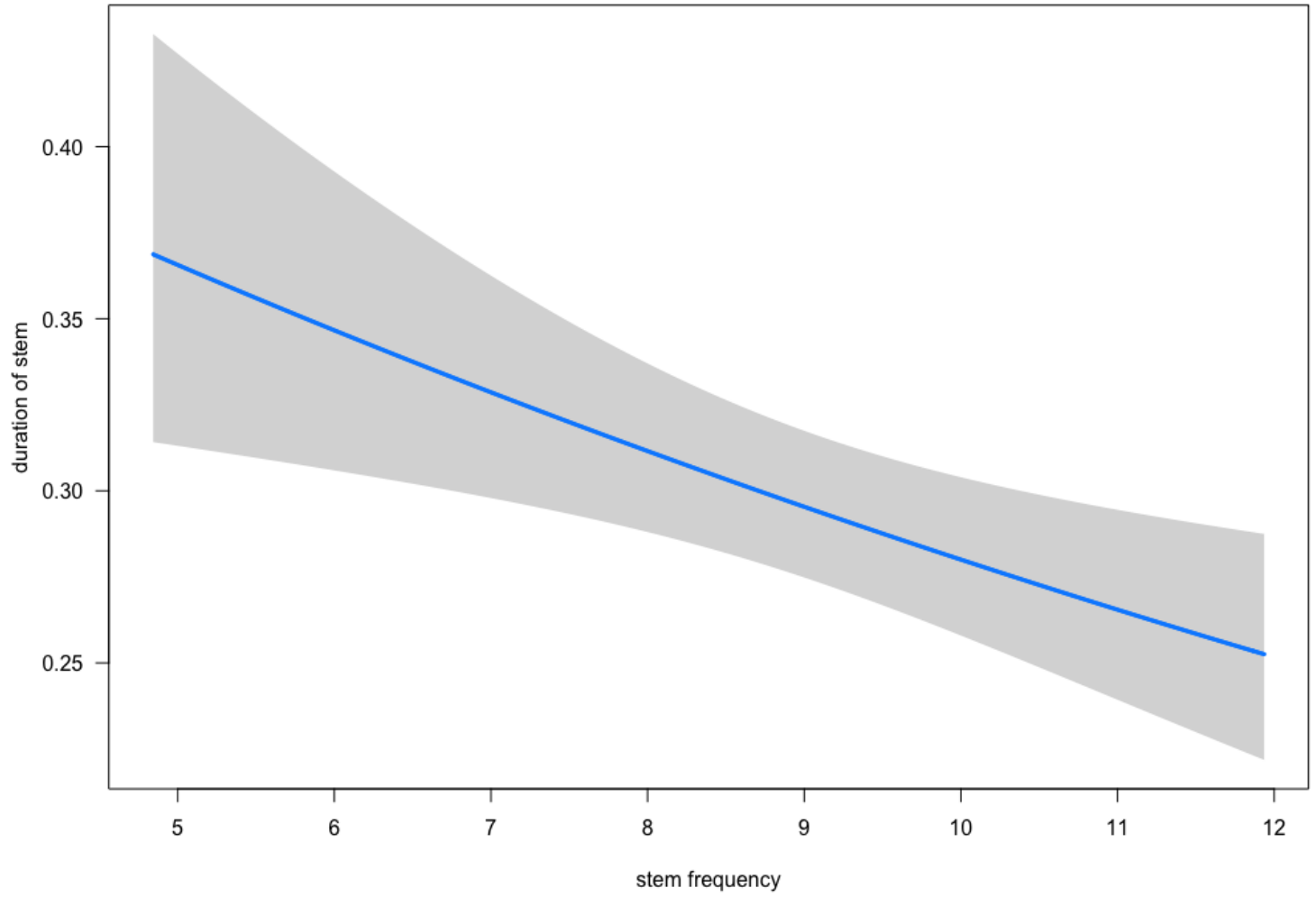
gradient paradigm uniformity effect
due to **phonetic reduction**

Methodology

- sub-dataset with only plural words (324 tokens, 40 types)
- response variable: **stem duration**
- 2 different models with predictor variables: **stem frequency**, **word form frequency**

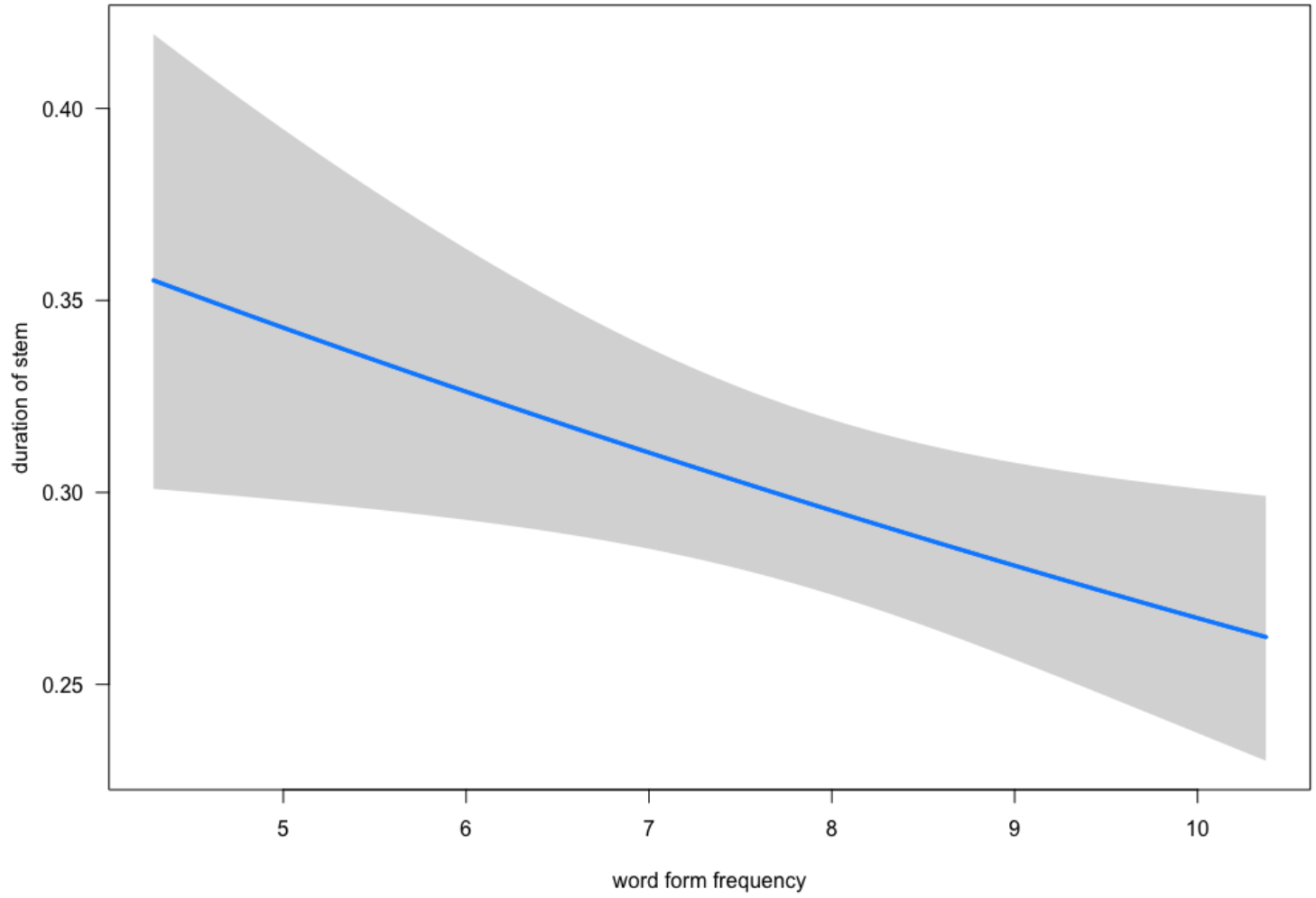
Results: Gradient Paradigm Uniformity (1/4)

- **stem frequency**: more frequent bare stem causes shorter duration of plural stem



Results: Gradient Paradigm Uniformity (3/4)

- **word form frequency**: higher frequency of the plural form causes shorter plural stem durations



Results: Gradient Paradigm Uniformity (3/4)

- our results refute hypothesis 2 and confirm hypotheses 3:
  - H2: higher bare stem frequency ≠ longer plural stem duration
  - H3a: higher bare stem frequency = shorter plural stem duration
  - H3b: higher word form frequency = shorter plural stem duration

Results: Gradient Paradigm Uniformity (4/4)

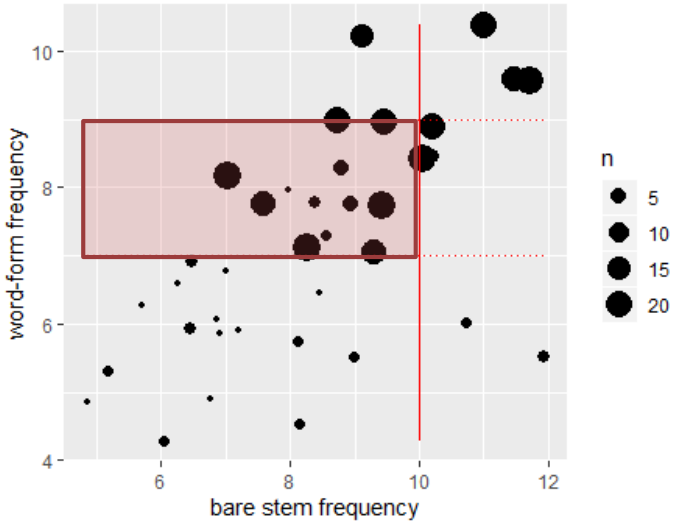
- we do not find evidence for gradient paradigm uniformity due to strength of activation
- we find evidence for gradient paradigm uniformity due to **phonetic reduction**

Correlation of frequencies

- word form frequency and bare stem frequency correlate positively in our data set (*rho*=0.61, $p<0.001$, Spearman test)
- both correlate negatively with duration
- shorter plural forms with rising stem frequency might be a masked word-form frequency effect

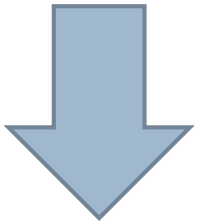
Unmasking word form frequency effects (1/2)

- we created a model in which frequencies do not correlate
  - we chose a narrow word-form frequency band in the middle of the distribution that had many observations
- observations with log word-form frequencies between 7 and 9, and log bare stem frequency of less than 10
- 164 observations (against 314 in the previous analysis)



Unmasking word form frequency effects (2/2)

- significant effect of bare stem frequency on the duration of the plural stem
  - dataset without correlation of word form frequency
  - dataset without effect of word form frequency on duration



strong support for H3:
phonetic reduction

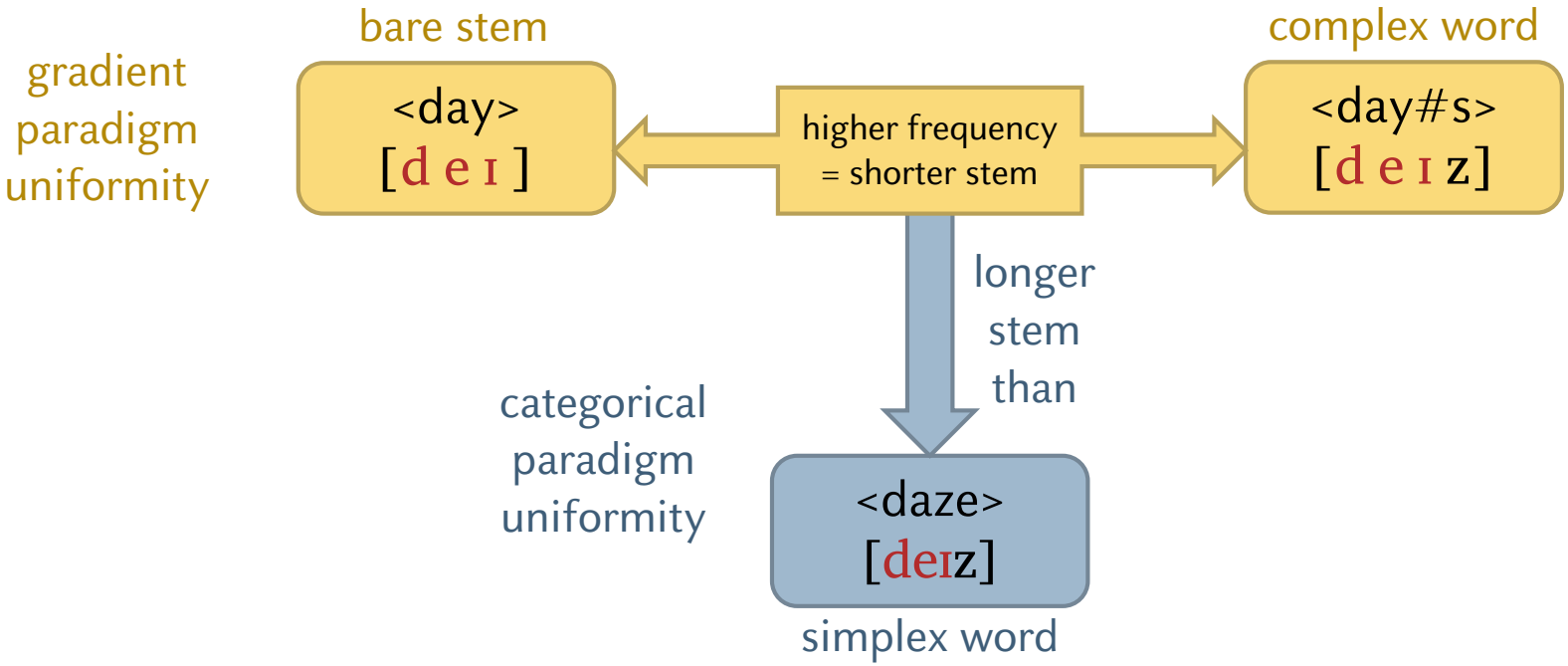
Conclusion (1/3)

- we tested three predictions following from work on paradigm uniformity on corpus data of New Zealand English (Seyfarth et al. 2017)
- support for hypothesis 1: categorical paradigm uniformity
  - plural stems are about 25 ms longer than monomorphemic stems
  - results are in line with Seyfarth et al. (2017) who found that complex stems are 18ms longer
  - robust effect across different types of data and varieties

Conclusion (2/3)

- no support for hypothesis 2: gradient paradigm uniformity due to strength of activation
  - contrary to what Seyfarth et al. predicted
- support for hypothesis 3: gradient paradigm uniformity due to phonetic reduction
  - general reduction effect of high frequency forms on paradigm members

Conclusion (3/3)



Thank you for your attention!

References (1/2)

- Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Gabor Grothendieck & Peter Green. 2017. *lme4: Linear Mixed-Effects Models using "Eigen" and S4*. https://cran.r-project.org/web/packages/lme4/index.html (24 October, 2017).
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2015. *The Oxford Reference Guide to English Morphology*. Oxford University Press.
- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111. doi:10.1016/j.jml.2008.06.003.
- Caselli, Naomi K., Michael K. Caselli & 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.
- Cohen, Clara. 2014. Probabilistic reduction and probabilistic enhancement. *Morphology* 24(4). 291–323. doi:10.1007/s11525-014-9243-y.
- Gahl, Susanne. 2008. "Time" and "thyme" are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3). 474–496.
- Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. *Frequency and the emergence of linguistic structure* (Typological Studies in Language, Vol. 45), 229–254. Amsterdam, Netherlands: John Benjamins Publishing Company. doi:10.1075/tsl.45.13jur.
- Kemps, Rachel J. J. K., Mirjam Ernestus, Robert Schreuder & Harald Baayen. 2005. Prosodic cues for morphological complexity: the case of Dutch plural nouns. *Memory & Cognition* 33(3). 430.
- Kemps, Rachel J. J. K., Lee H. Wurm, Mirjam Ernestus, Robert Schreuder & Harald Baayen. 2005. Prosodic cues for morphological complexity in Dutch and English. *Language And Cognitive Processes* 20(1–2). 43–73. doi:10.1080/01690960444000223.
- Lohmann, Arne. 2018a. Time and thyme are <small class="caps">not</small> homophones: A closer look at Gahl's work on the lemma-frequency effect, including a reanalysis. *Language* 94(2). e180–e190. doi:10.1353/lan.2018.0032.
- Lohmann, Arne. 2018b. Cut (n) and cut (v) are not homophones: Lemma frequency affects the duration of noun–verb conversion pairs. *Journal of Linguistics* 54(4). 753–777. doi:10.1017/S0022226717000378.

References (2/2)

- Lõo, Kaidi, Juhani Järvikivi, Fabian Tomaschek, Benjamin V. Tucker & R. Harald Baayen. 2018. Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology* 28(1). 71–97. doi:10.1007/s11525-017-9318-7.
- Plag, Ingo, Sonia Ben Hedia, Arne Lohmann & Julia Zimmermann. 2019. An <s> is an <s'>, or is it? Plural and genitive-plural are not homophonous. *submitted for publication*.
- Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53(1). 181–216.
- Pluymaekers, Mark, Mirjam Ernestus, R. Harald Baayen & Geert Booij. 2010. Morphological effects on fine phonetic detail: The case of Dutch-igheid. (Ed.) C Fougeron, B Kühnert, M D'Imperio & N Vallée. *Laboratory phonology* 10. 511–531.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. (Version 3.2.1). Vienna, Austria. https://www.R-project.org.
- Roettger, Timo B. 2019. Researcher degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1). 1. doi:10.5334/labphon.147.
- Seyfarth, Scott, Marc Garellek, Gwendolyn Gillingham, Farrell Ackerman & Robert Malouf. 2017. Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience* 33(1). 32–49.
- Tomaschek, Fabian, Ingo Plag, Mirjam Ernestus & Harald R. Baayen. 2019. Modeling the duration of word-final S in English with Naive Discriminative Learning. *submitted to Journal of Linguistics*.
- Walsh, Liam, Jen Hay, Derek Bent, Jeanette King, Paul Millar, Viktoria Papp & Kevin Watson. 2013. The UC QuakeBox Project: Creation of a community-focused research archive. https://ir.canterbury.ac.nz/handle/10092/15635 (20 November, 2018).
- Zimmermann, Julia. 2016a. Morphological Status and Acoustic Realization: Findings from NZE. In C Carignan & M.D. Tyler (eds.), *Proceedings of the 16th Australasian International Conference on Speech Science and Technology*, 6–9. Sydney: University of Western Sydney.
- Zimmermann, Julia. 2016b. Morphological Status and Acoustic Realization: Findings from NZE. Heinrich Heine Universität Düsseldorf.