Comprehension difficulty renders the *-er* suffix in generic masculines longer than in specific masculines

Dominic Schmitz

Heinrich Heine University Düsseldorf, Germany

div-ling talk series, Winter 2024/2025

Suphonemic differences

- previous research found durational differences where established theories of speech production do not expect them (e.g. Kiparsky 1982, Levelt et al. 1999)
  - homophonous free and bound (pseudo-)stems (e.g. Seyfarth et al. 2017)

    frees vs. *freeze*
  - homophonous prefixes (e.g. Ben Hedia & Plag 2017)

    impossible vs. *implant* (negative vs. locative)
  - types of /s/ (e.g. Plag et al. 2017, Schmitz et al. 2021)

    bus vs. *cats* vs. *cat's* (non-morphemic vs. suffix vs. clitic)
- similar phonology + different morphology = differences in phonetics
- similar phonology + similar morphology = ???

Specific and generic masculines in German

- in German, masculine role nouns with feminine counterparts can be used generically, i.e. independent of a referent's gender (e.g. Kotthoff & Nübling, 2024)

*Tim ist **Lehrer** von Beruf.*

'Tim is a teacher by profession.'

specific

*Anna ist **Lehrer** von Beruf.*

'Anna is a teacher by profession.'

generic

- may be further differentiated in terms of 'gender definiteness'

gender-specified

*Anna ist **Lehrer** von Beruf.*

'Anna is a teacher by profession.'

definite

not gender-specified

*Mein Kind ist **Lehrer** von Beruf.*

'My child is a teacher by profession.'

indefinite

Research questions

RQ 1

Does the semantic difference between **specific** and **generic** masculines lead to subphonemic durational differences?

RQ 2

Does the semantic difference between **definite** and **indefinite** generic masculines lead to subphonemic durational differences?

RQ 3

If there are durational differences, how can they be accounted for?

Experiment: Reading Task

Part 1

Materials

Items

- **targets**: 20 role nouns ending in the *-er* suffix, i.e. /ɐ/

stereotypically female (Misersky et al., 2014)				
Balletttänzer	*Eiskunstläufer*	*Flugbegleiter*	*Geburtshelfer*	*Haushälter*
Hellseher	*Kosmetiker*	*Pfleger*	*Schneider*	*Verkäufer*
stereotypically male				
Bauarbeiter	*Elektriker*	*Fußballspieler*	*Kranführer*	*Maurer*
Programmierer	*Rennfahrer*	*Reporter*	*Schreiner*	*Wahrsager*

- **fillers**
  - feminine forms of target items, e.g. *Balletttänzerin*, *Bauarbeiterin*
  - used with female referents only

Materials

Contexts

- phrase or sentence introducing the referent
- phrase or sentence containing the target item

specific	***Matteos Vater*** *kann richtig gut nähen.*
	Er *ist* ***Schneider*** *von Beruf.*
indefinite generic	***Mein Kind*** *kann richtig gut nähen.*
	Es *ist* ***Schneider*** *von Beruf.*
definite generic	***Marlenes Mutter*** *kann richtig gut nähen.*
	Sie *ist* ***Schneider*** *von Beruf.*

Materials

Lists

- 4 lists with 40 items, i.e. 30 targets + 10 fillers
- per list:

		type	number
15	5	specific masculine	singular
	5	generic masculine, definite	
	5	generic masculine, indefinite	
15	5	specific masculine	plural
	5	generic masculine, definite	
	5	generic masculine, indefinite	
10	5	specific feminine	singular
	5	specific feminine	plural

- pseudo-randomised: trials with the same item did not directly follow each other

Participants & procedure

Participants

- 40 participants
- L1 German
- age: mean 29.1 years, range: 20 – 64 years

Procedure

- 1 set of context and target phrase/sentence per trial
- instructions: read quietly before reading aloud
- self-paced

Acoustic analysis

- annotation of base and suffix durations in Praat (Boersma & Weenink, 2024)
- utterances with production errors, stutter, laughter were excluded ($n = 87$)
- extraction of durational information via rPraat (Bořil & Skarnitzl, 2016) in R (R Core Team, 2024) ($n = 1113$)
- example: *Geburtshelfer* 'obstetrician'
  - one is a definite generic plural, one is a specific singular




Statistical analysis

- **initial** linear mixed-effects regression model, fitted with *lme4* (Bates et al., 2015)

```
durEr ~
    durBase +                       # duration of the base
    typeOfEr +                      # specific, definite or indefinite generic
    preType + folType +             # type of preceding and following segment
    number + stereotypicality +  # singular/plural, male/female
    speechRate + trialNumber +
    age + gender +
    attGM +                         # attitude towards generic masculines
    (1 | speaker) + (1 | word)
```

- model with **best fit**, found with *lmerTest* (Kuznetsova et al., 2017)

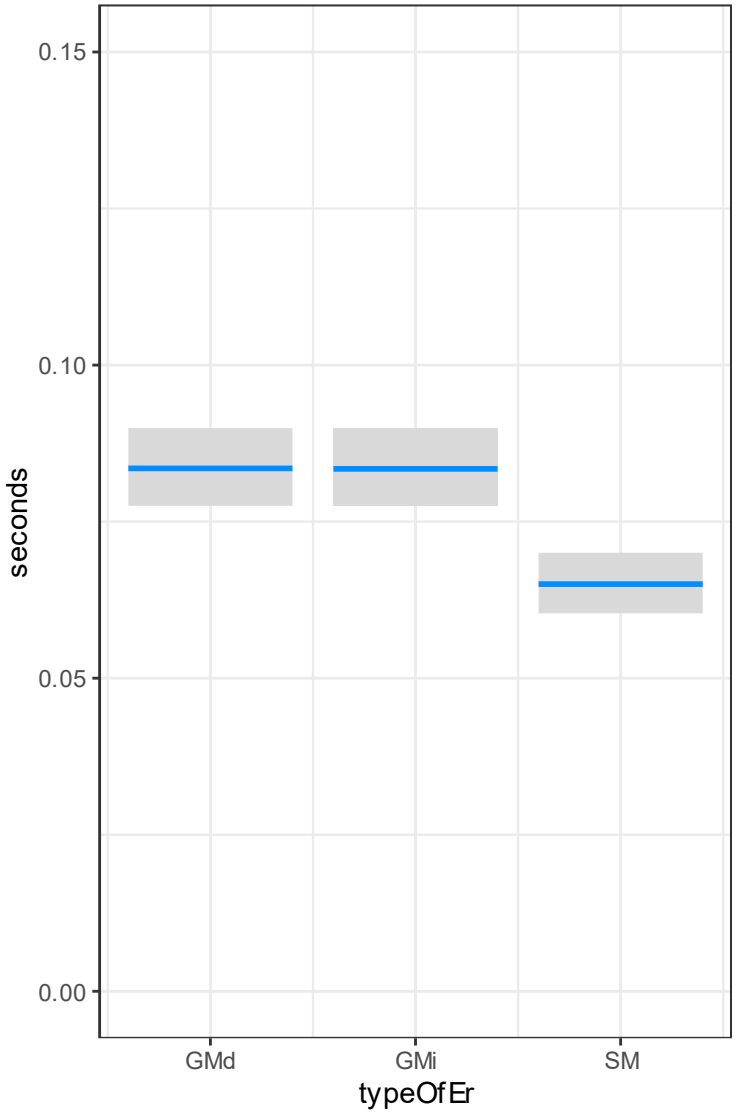
```
durEr ~ typeOfEr + (1 | speaker) + (1 | word)
```

Results

- the type of masculine shows a clearly significant effect, i.e. GMd = GMi > SM

	GMd	GMi	SM
mean (sd)	0.0869 (0.0262)	0.0871 (0.0258)	0.0682 (0.0217)

- the effect size is large with $\eta^2 = 0.2$, with 95% CI of $[0.48, 1.00]$



Discussion

RQ 1

Does the semantic difference between **specific** and **generic** masculines lead to subphonemic durational differences?

→ YES

RQ 2

Does the semantic difference between **definite** and **indefinite** generic masculines lead to subphonemic durational differences?

→ NO

Discussion

RQ 1

Does the semantic difference between **specific** and **generic** masculines lead to subphonemic durational differences?

→ YES

RQ 2

Does the semantic difference between **definite** and **indefinite** generic masculines lead to subphonemic durational differences?

→ NO

RQ 3

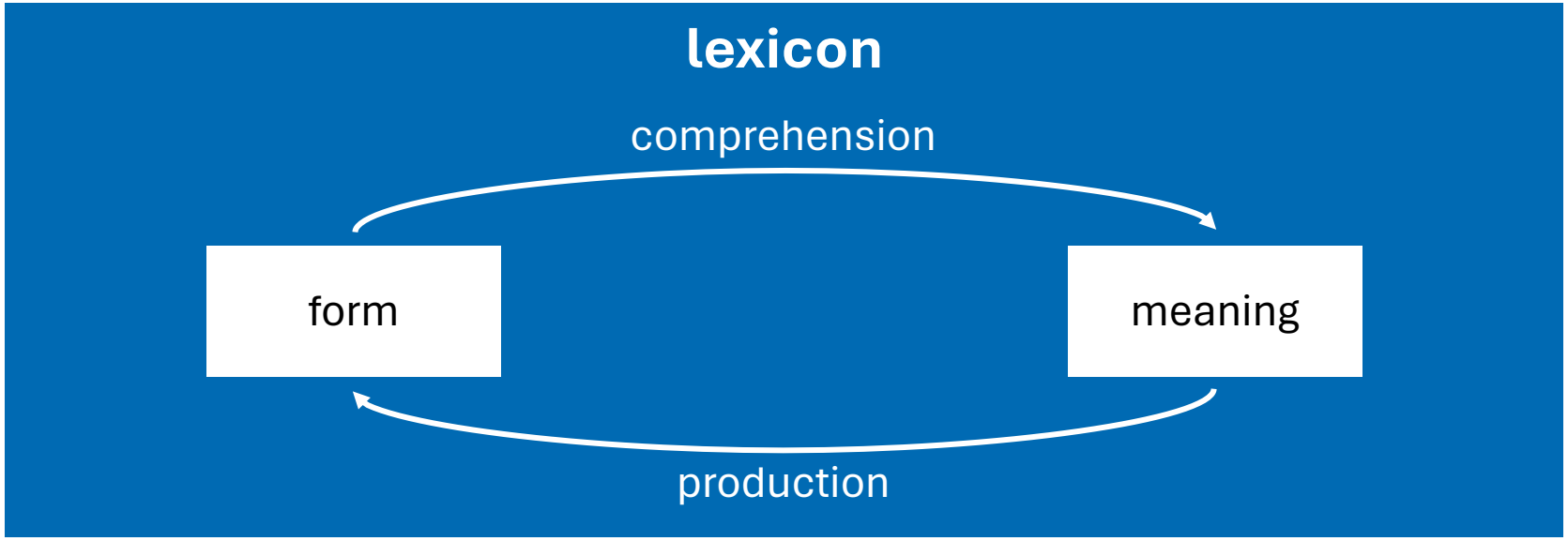
If there are durational differences, how can they be accounted for?

LDL implementation

Part 2

Idea

- model a lexicon with generic masculines, specific masculines, and other entries to gain more detailed insight into the semantic and form features of generic masculines and specific masculines



Lexicon – targets

- similar to the setup in Schmitz et al. (2023)
- **step 1**

  one million sentences per year from 2011 – 2020 from the ‘news’ subcorpus of the Leipzig Corpora Collection (Goldhahn et al., 2012)
- **step 2**

  sample all sentences containing the targets = targets from the experiment
- **step 3**

  get overall frequency of each target
- **step 4**

  sample random sentences according to frequency

Lexicon – targets

- similar to the setup in Schmitz et al. (2023)

frequency	*n* samples	
up to 200	100	*Ballettänzer, Eiskunstläufer, Geburtshelfer, Haushälter, Hellseher, Kosmetiker, Kranführer, Maurer, Wahrsager*
201 – 1000	200	*~~Bauarbeiter~~, Elektriker, Flugbegleiter, Fußballspieler, Programmierer, Schreiner*
1001 – 2000	300	*Pfleger, Rennfahrer*
2001 – 10000	400	*Reporter, ~~Schneider~~, Verkäufer*
10001 – 20000	500	
20001 and more	600	

Lexicon – other entries

- word list based on the sampled sentences containing target words
- ensures that the words that make up the lexicon are actually found ‘in the wild’ with the words we are interested in
- overall, 11745 word-forms

Form

- form is represented by triphones
- based on phonological transcriptions provided by the Python package *epitran* (Mortensen et al., 2018)

	#hɛ	hɛl	…	eɐ#	nɐ#	fɐ#
Hellseher	1	1	…	1	0	0
Schreiner	0	0	…	0	1	0
Verkäufer	0	0	…	0	0	1

Meaning

- embeddings computed with the pre-trained BERT model ‘bert base german cased’ (Devlin et al., 2018)
- **for target words**

  context-dependent embeddings via the sentences from the experiment
- **for all other words**

  given in isolation, i.e. ‘basic’ embeddings straight from the BERT model

Measures

- based on the LDL implementation, the following measures were computed
  - degree of semantic co-activation

    Euclidean norm of a given predicted semantic vector

Measures

- based on the LDL implementation, the following measures were computed
  - degree of semantic co-activation
  - degree of comprehension accuracy

    correlation of input and predicted semantic vector

Measures

- based on the LDL implementation, the following measures were computed
  - degree of semantic co-activation
  - degree of comprehension accuracy
  - semantic neighbourhood density

    mean correlation with predicted vectors of 20 nearest neighbours

Measures

- based on the LDL implementation, the following measures were computed
  - degree of semantic co-activation
  - degree of comprehension accuracy
  - semantic neighbourhood density
  - degree of polysemy

    Shannon entropy of the predicted semantic vector

Measures

- based on the LDL implementation, the following measures were computed
  - degree of semantic co-activation
  - degree of comprehension accuracy
  - semantic neighbourhood density
  - degree of polysemy
  - degree of form co-activation

    Euclidean norm of a given predicted form vector

Measures

- based on the LDL implementation, the following measures were computed
  - degree of semantic co-activation
  - degree of comprehension accuracy
  - semantic neighbourhood density
  - degree of polysemy
  - degree of form co-activation
  - degree of form suffix support

    weight of the final triphone in the predicted form matrix

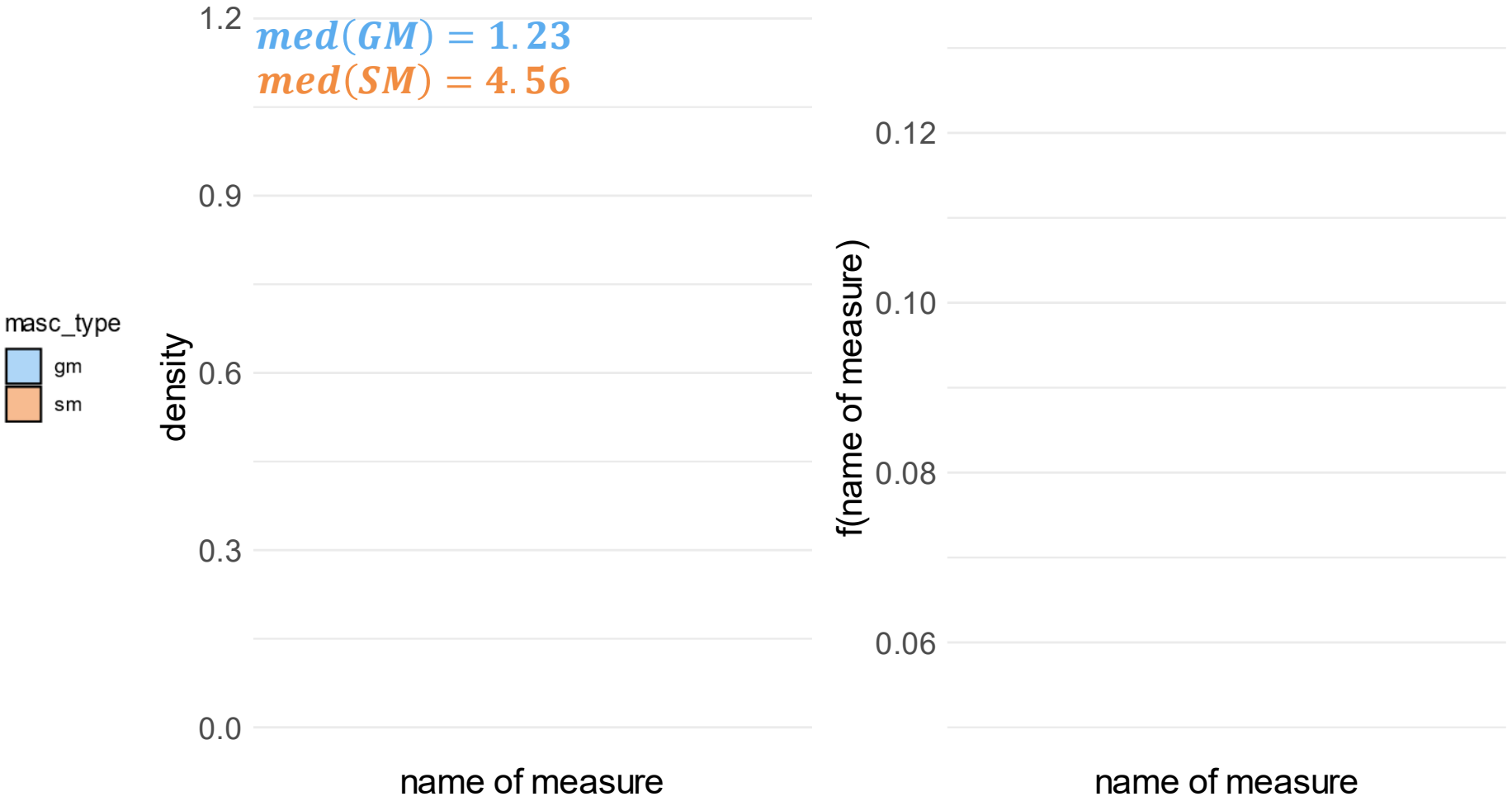
Analysis

- linear mixed effects model similar to the one used for the production experiment but with LDL measures added
- model with **best fit**, found with *lmerTest* (Kuznetsova et al., 2017)

```
dur_er_log ~
        degree of semantic co-activation +
        degree of comprehension accuracy +
        semantic neighbourhood density +
        degree of polysemy +
        (1 | speaker) + (1 | word)
```

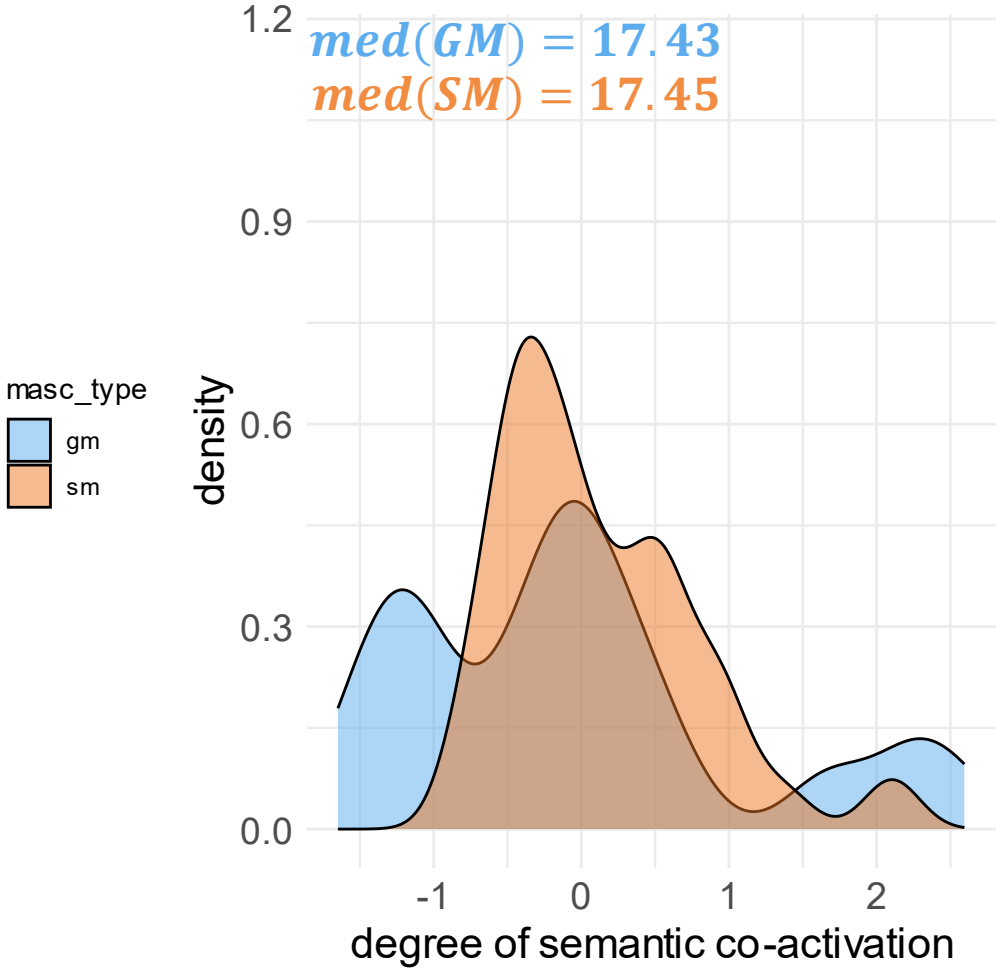
Results

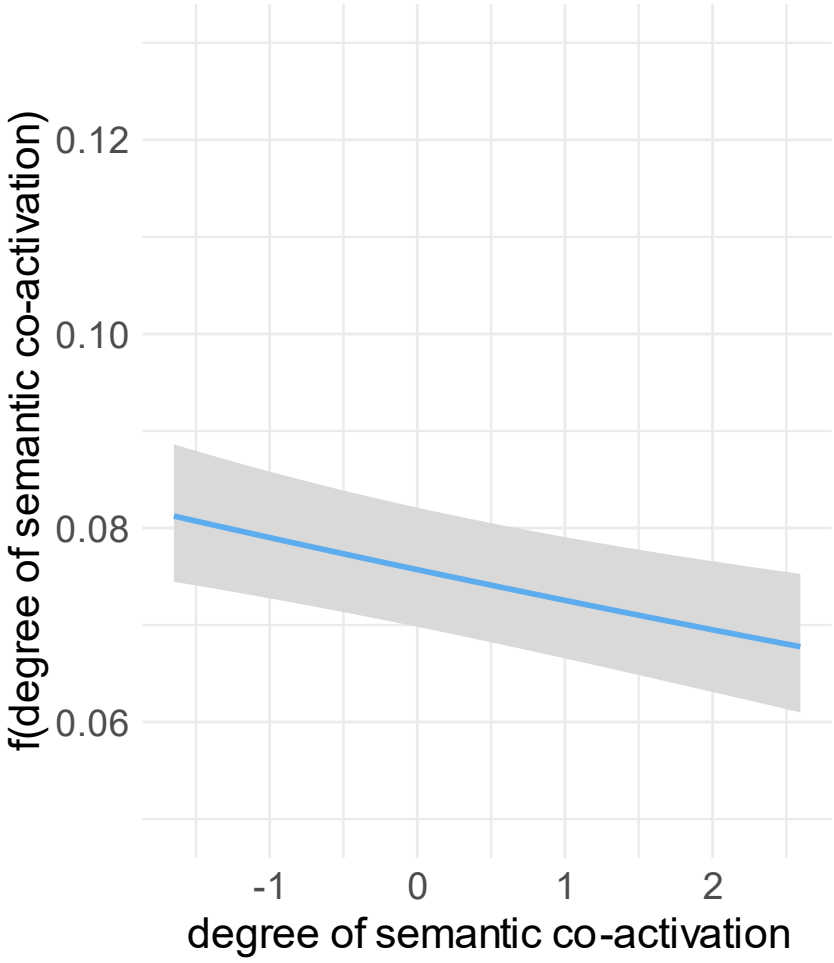
name of measure



Results

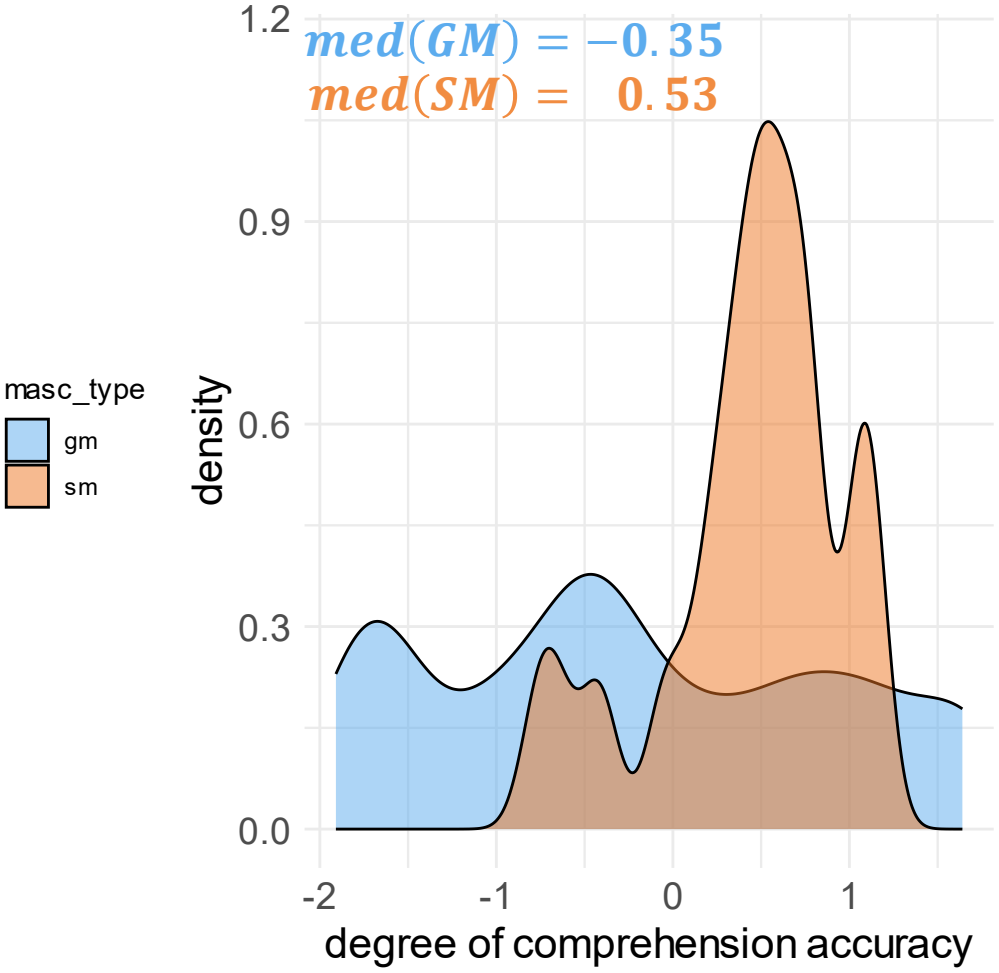
degree of semantic co-activation

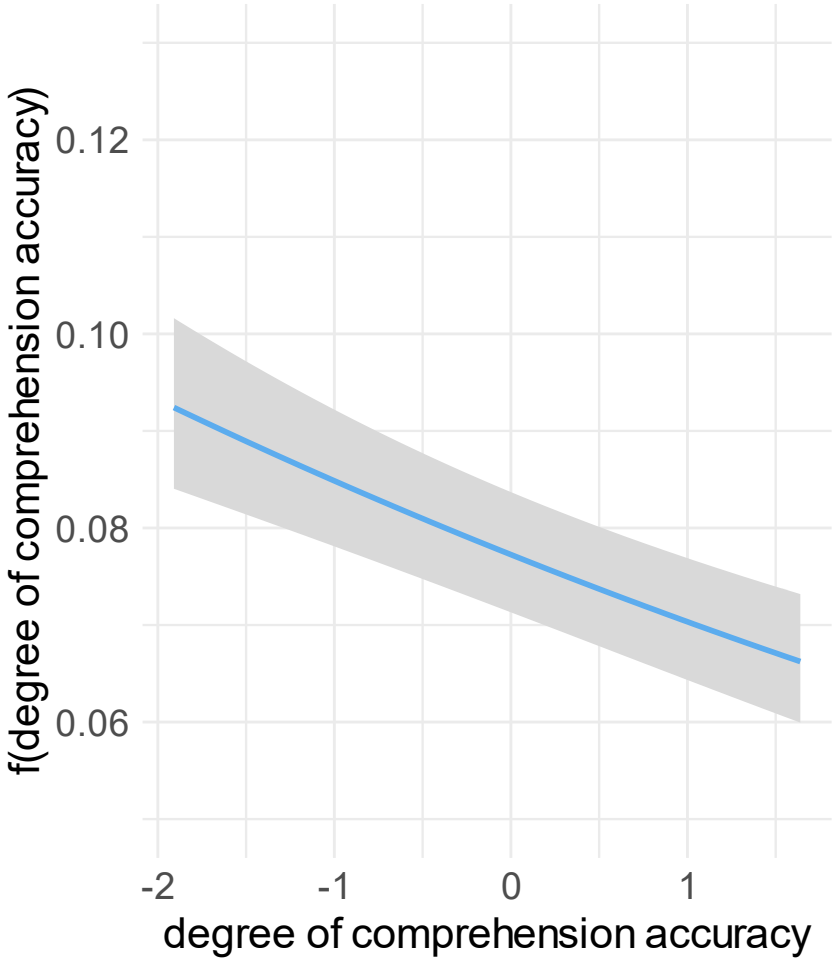




Results

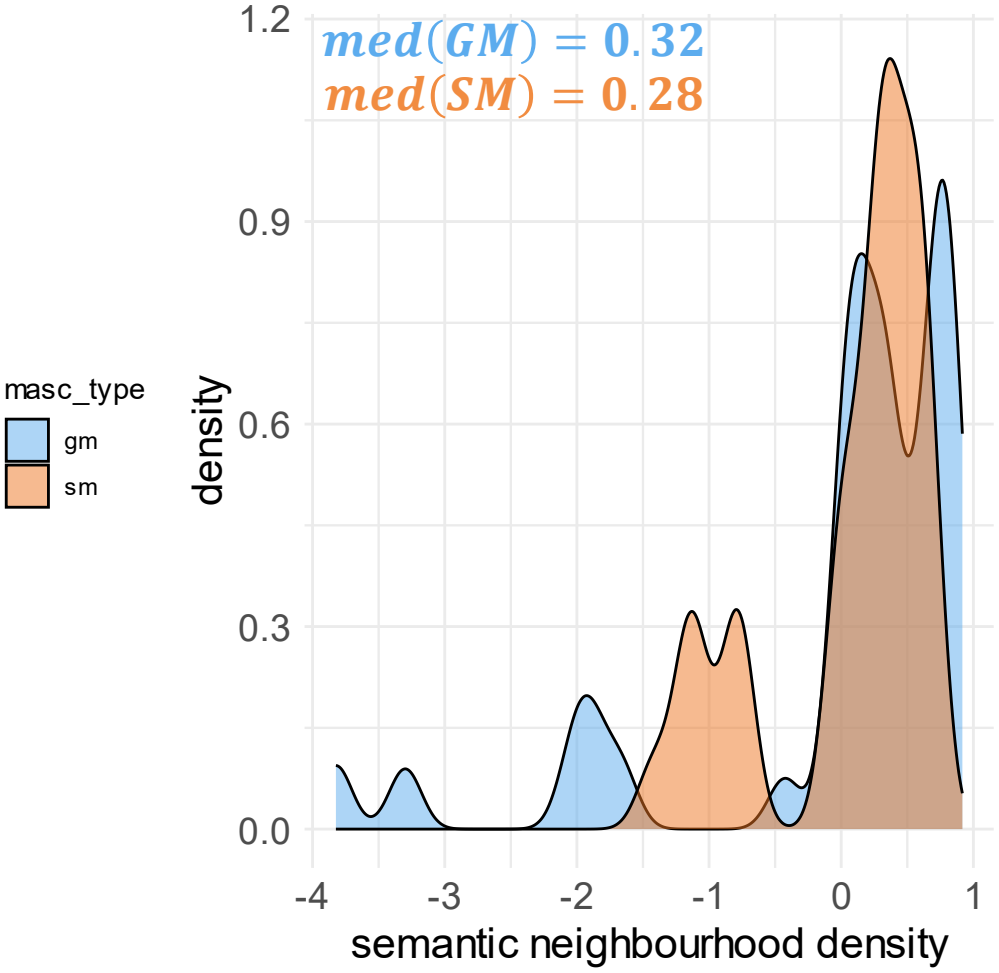
degree of comprehension accuracy

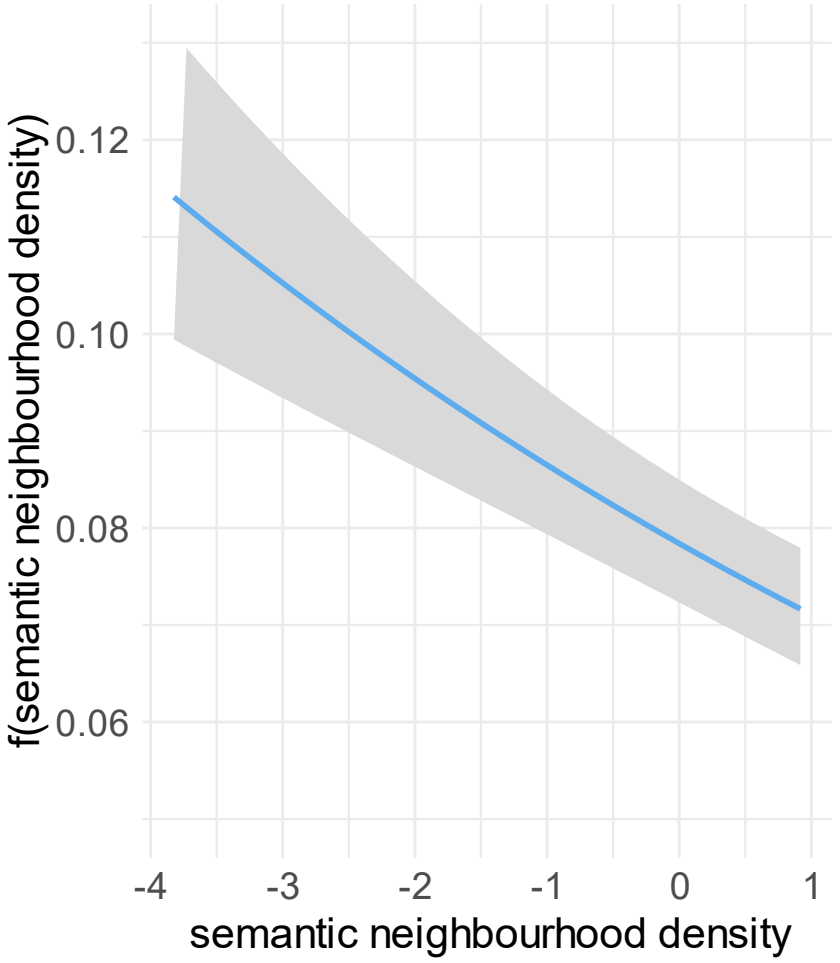




Results

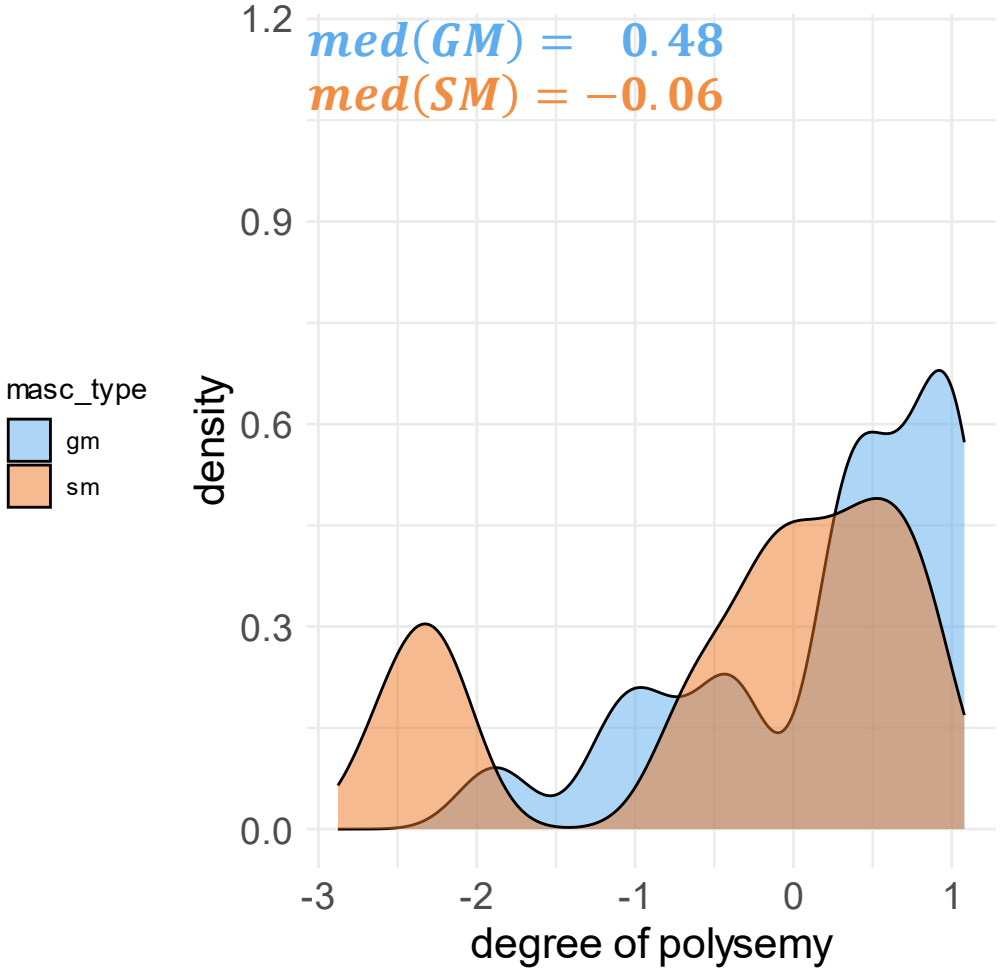
semantic neighbourhood density

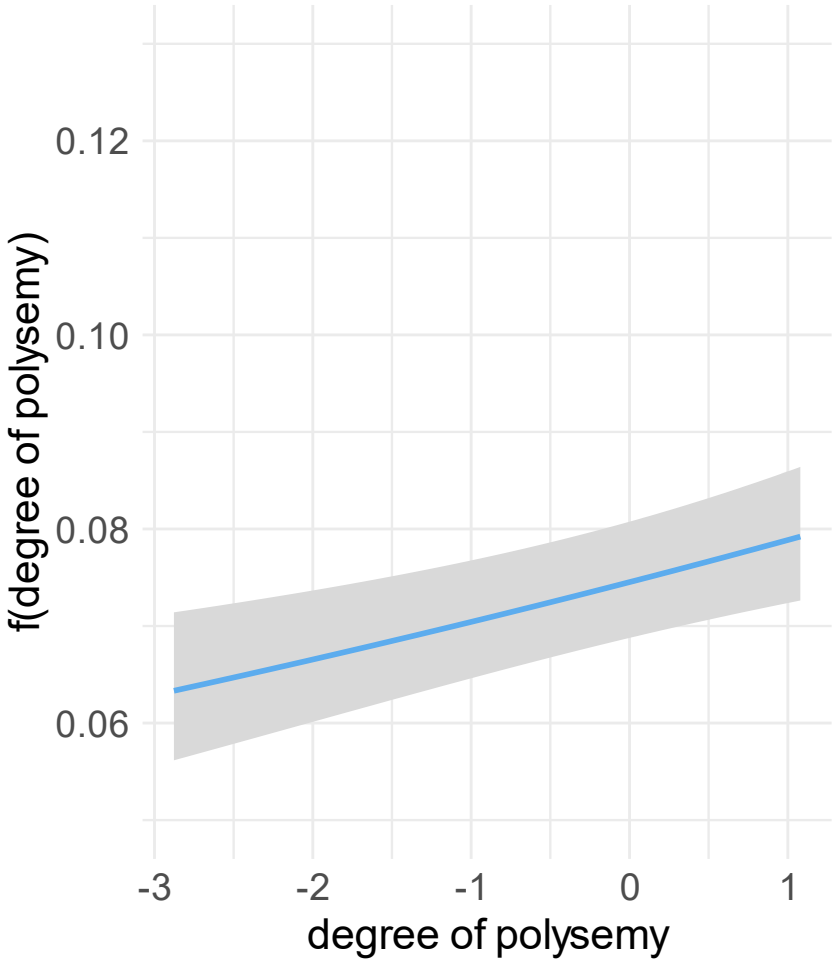




Results

degree of polysemy





Summary

- the higher the degree of semantic coactivation, the **shorter** the /ɐ/

  → general effect ($med(GM) = 17.43, med(SM) = 17.45$)

- the higher the comprehension accuracy, the **shorter** the /ɐ/

  → SM shorter than GM ($med(GM) = -0.35, med(SM) = 0.53$)

- the denser the semantic neighbourhood, the **shorter** the /ɐ/

  → general effect ($med(GM) = 0.32, med(SM) = 0.28$)

- the higher the degree of polysemy, the **longer** the /ɐ/

  → SM shorter than GM ($med(GM) = 0.48, med(SM) = -0.06$)

Discussion

RQ 3

If there are durational differences, how can they be accounted for?

→ specific masculines are **better comprehended** than generic masculines

→ specific masculines are **less polysemous** than specific masculines

→ in line with the idea by Schmitz (2024) that generic masculines come with a higher comprehension effort as they may refer to a wider variety of referents

Conclusion

- the /ɐ/ in generic masculines shows a longer duration than in specific masculines
- the durational difference is not influenced by gender definiteness, stereotypicality or the attitude towards generic masculines
- the potential cause of the durational difference lies in the more polysemous semantics and with that worse comprehension of masculine generics

THANK YOU!

References

Bates, D., **Mächler**, M., **Bolker**, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Ben Hedia, S., & **Plag**, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics*, *62*, 34–49. https://doi.org/10.1016/j.wocn.2017.02.002

Boersma, P., & **Weenink**, D. (2024). *Praat: Doing phonetics by computer* (6.4.07, retrieved 17/03/2024). http://www.praat.org/

Bořil, T., & **Skarnitzl**, R. (2016). *Tools rPraat and mPraat* (pp. 367–374). https://doi.org/10.1007/978-3-319-45510-5_42

Devlin, J., **Chang**, M. W., **Lee**, K., & **Toutanova**, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*, 4171–4186.

Goldhahn, D., **Eckart**, T., & **Quasthoff**, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. *Proceedings of the 8th International Language Resources and Evaluation (LREC'12)*.

Kiparsky, P. (1982). Lexical morphology and phonology. In I. Yang (Ed.), *Linguistics in the morning calm: Selected papers from SICOL1* (pp. 3–91). Hanshin.

Kuznetsova, A., **Brockhoff**, P. B., & **Christensen**, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Nübling, D., & **Kotthoff**, H. (2024). *Genderlinguistik: Eine Einführung in Sprache, Gespräch und Geschlecht*. Narr Francke Attempto.

Levelt, W. J. M., **Roelofs**, A., & **Meyer**, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1–75. https://doi.org/10.1017/S0140525X99001776

Mortensen, D. R., **Dalmia**, S. & **Littell**, P. (2018). Epitran: Precision G2P for many languages. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Plag, I., **Homann**, J., & **Kunter**, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, *53*(1), 181–216. https://doi.org/10.1017/S0022226715000183

R Core Team. (2024). *R: A language and environment for statistical computing* (4.0.4). R Foundation for Statistical Computing. https://www.r-project.org/

Schmitz, D. (2024). Instances of bias: The gendered semantics of generic masculines in German revealed by instance vectors. *Zeitschrift für Sprachwissenschaft, 43*(2). https://doi.org/10.1515/zfs-2024-2010

Schmitz, D., **Baer-Henney**, D., & **Plag**, I. (2021). The duration of word-final /s/ differs across morphological categories in English: Evidence from pseudowords. *Phonetica*, *78*(5–6), 571–616. https://doi.org/10.1515/phon-2021-2013

Schmitz, D., **Schneider**, V., & **Esser**, J. (2023). No genericity in sight: An exploration of the semantics of masculine generics in German. *Glossa Psycholinguistics, 2*(1). https://doi.org/10.5070/G6011192

Seyfarth, S., **Garellek**, M., **Gillingham**, G., **Ackerman**, F., & **Malouf**, R. (2017). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, *33*(1), 32–49. https://doi.org/10.1080/23273798.2017.1359634

Misersky, J., **Majid**, A., & **Snijders**, T. M. (2019). Grammatical gender in German influences how role-nouns are interpreted: Evidence from ERPs. *Discourse Processes*, *56*(8), 643–654. https://doi.org/10.1080/0163853X.2018.1541382