HOMOPHONOUS SEMANTIC MINIMAL PAIRS DIFFER IN THEIR SUBPHONEMIC ACOUSTIC DURATIONS: THE CASE OF GENERIC AND SPECIFIC MASCULINES IN GERMAN

Dominic Schmitz
Heinrich-Heine-Universität Düsseldorf
dominic.schmitz@uni-duesseldorf.de

Abstract: Previous research showed that subphonemic durational variation is modulated by lexical and morphological differences. The present study takes research on subphonemic differences one step further and asks: are there subphonemic durational differences in phonologically, morphologically, and lexically identical members of semantic minimal pairs? To answer this question, a sentence reading task on German was conducted. Target words were generic masculines and specific masculines ending in *-er*. Items were preceded by a context and embedded in a sentence, with similar contexts and sentences for the forms of the same item. 40 participants produced 30 targets each. The duration of the *-er* suffix, i.e. /ɐ/, was analysed in linear mixed-effects regression models. The results showed that generic masculines come with longer /ɐ/ durations than specific masculines. These findings add to the body on subphonemic durational differences unaccounted for by established theories of speech production.

1 Subphonemic differences and the generic masculine in German

Previous research identified subphonemic durational differences conditioned by either morphological or lexical differences. Most prominently, word-final /s/ in English shows different durations depending on its morphological nature. Non-morphemic /s/ appears to be longest, followed by suffix /s/, which in turn is followed by auxiliary clitic /s/ [1]. Further durational differences have been found, for example, in prefixes [2], in homophonous free and bound (pseudo-)stems [3], and in homophonous words [4].

Such subphonemic durational differences are unexpected from a traditional theoretical view, as morphological and lexical information is assumed to be unavailable to speech production [5]. Psycholinguistic models of speech production similarly assume that morphological information is no longer available once the phonological stage is reached [6]. Exemplar-based models may account for subphonemic durational differences, suggesting that the experienced durational differences are stored and with that also produced [7], but lack an explanation as to how exactly these durational differences come into existence in the first place. Thus far, it is the Discriminative Lexicon Model and its computational implementations [8] that are able to offer explanations for these durational differences based on meaning- and form-related features of the elements under investigation [9, 10].

As evidence on subphonemic durational differences caused by morphological and lexical differences amasses, the present study takes research on such fine-phonetic differences one step further and asks whether are there subphonemic durational differences in phonologically, morphologically, and lexically identical members of semantic minimal pairs.

With specific and generic masculines, German–among other languages–offers a good testcase to gain first insights into whether fine-semantic differences also cause fine-phonetic differences. In German, grammatically masculine role nouns with feminine counterparts may not only be used to refer to male individuals but also to individuals of any other gender [11]. If such a masculine is used to refer to a male individual, e.g. *Max ist Lehrer* 'Max is a teacher', this usage is called *specific*. If, however, the same role noun is used to refer to an individual of another gender, e.g. *Tina ist Lehrer* 'Tina is a teacher', or to a referent of undeclared gender, e.g. *mein Kind ist Lehrer* 'my child is a teacher', this usage is called *generic*. Henceforth, I will call the generic usage referring to gender-specified referents *specified generic* and the generic usage referring to referents with undeclared gender *unspecified generic*.

Lehrer, just as many other role nouns in German, ends in the suffix *-er*, produced as /ɐ/, as this suffix is attached to the stem of a verb to derive a role noun.[1] All role nouns derived this way are grammatically masculine and come with a feminine counterpart.[2]

Making use of such specific masculines, specified generic masculines, and unspecified generic masculines, the present study aims to find answers to the following two research questions:

RQ1 Does the fine-semantic difference between specific masculines and generic masculines lead to fine-phonetic durational differences in word-final *-er*?

RQ2 Does the difference in specification between specified generic masculines and unspecified generic masculines lead to fine-phonetic durational differences in word-final *-er*?

2 Methodology

2.1 Materials

For the present study, a sentence reading task was conducted. Items were embedded in simple target sentences, all of which were preceded by a context phrase or sentence setting the scene. Both phrases/sentences in combination made clear whether a masculine role noun was used specifically or generically, and context sentences did not differ between different role noun forms of the same target. That is, except for the person who was referred to, as this was required to obtain the different types of role noun readings. An illustration is given in Example 1, with changing referents and related pronouns in italics and the target in bold.[3]

Ex.1 *Matteos Vater/Mein Kind/Marlenes Mutter* kann richtig gut nähen. *Er/Es/Sie ist* **Schneider** von Beruf.
Matteo's dad/My child/Marlene's mum is really good at sewing. *He/It/She* is a **tailor** by profession.

Items were 20 role nouns ending in *-er*, 10 of which were stereotypically female, while the other 10 were stereotypically male. All were taken from the list provided by [12], which includes information on stereotypicality. Table 1 lists all items.

2.2 Procedure

During the reading task, participants were presented with one set of context and target phrase/sentence at a time. They were instructed to first read both quietly, before then producing them aloud. The experiment progressed in a self-paced fashion, and participants were told that progressing quickly was not the main objective of their task. Each participant produced 30 sentences with target items, i.e. 10 sentences per type of masculine, and additionally 10 sentences in which the role noun was grammatically feminine. The latter functioned as filler items.

[1] Henceforth I will use *-er* and /ɐ/ interchangeably.
[2] The feminine counterpart is formed by attaching the feminine suffix *-in*, e.g. *Lehrerin* 'female teacher'.
[3] Find a full list at https://osf.io/dvrq4/.

Table 1 – Overview of all target words, including English translations.

stereotypically female				
Balletttänzer	*Eiskunstläufer*	*Flugbegleiter*	*Geburtshelfer*	*Haushälter*
(ballet dancer)	*(ice skater)*	*(flight attendant)*	*(obstetrician)*	*(housekeeper)*
Hellseher	*Kosmetiker*	*Pfleger*	*Schneider*	*Verkäufer*
(clairvoyant)	*(beautician)*	*(caregiver)*	*(tailor)*	*(vendor)*
stereotypically male				
Bauarbeiter	*Elektriker*	*Fußballspieler*	*Kranführer*	*Maurer*
(construction worker)	*(electrician)*	*(footballer)*	*(crane operator)*	*(mason)*
Programmierer	*Rennfahrer*	*Reporter*	*Schreiner*	*Wahrsager*
(programmer)	*(race driver)*	*(reporter)*	*(carpenter)*	*(fortune-teller)*

After the reading task, participants were asked to provide information on their gender, age, and attitude towards the use of generic masculines, pair forms, and gender star forms. Pair forms are phrases consisting of both the masculine and feminine form, e.g. *Lehrerin und Lehrer*; gender star forms constitute a novel role noun form making use of a new allegedly gender-neutral suffix, e.g. *Lehrer*in*. Attitudes towards the three forms were given using a 5-point Likert scale.

2.3 Participants

40 participants with German as L1 took part in the experiment. The mean age of the participants was 29.1 years, with an age-range from 20 to 64. Half of the participants were recruited at the author's institution, whereas the other half were recruited online.

2.4 Acoustic and statistical analyses

The acoustic analyses of all recordings took place in Praat [13]. For all targets, the duration of *-er*, i.e. /ɐ/, was determined using both the spectrogram and the waveform, following segmentation criteria based on the phonetic literature [14]. All boundaries were placed at the nearest zero crossing. Recordings with production errors, stutter, and laughter were excluded. Overall, 1113 of 1200 recordings were retained for further analysis. Based on the annotations in Praat, durations were extracted using the rPraat package [15] in R [16].

The duration of word-final *-er* was statistically analysed following the analysis of word-final *-s* in English as conducted by Schmitz et al. [17].[4] That is, the duration of *-er* was predicted in linear mixed-effects regression models using the lme4 package [18] in R. As fixed effects, the following variables were included:

TYPEOFER. This is the variable of interest. It encodes whether a word-final *-er* is that of an unspecified generic masculine (`GMu`), a specified generic masculine (`GMs`), or a specific masculine (`SM`).

DURBASE. Indicating a local speaking rate [e.g. 1], base duration was measured. Base duration here is equal to the summed duration of all word-internal segments preceding the *-er*.

PRETYPE, FOLTYPE. Different types of preceding and following segments may result in different co-articulatory effects. Hence, the type of the preceding and following speech sound was included in the analysis. These variables take the values `fricative`, `liquid`, `nasal`, `plosive`, and `vowel`.

SYLNUMBER. The number of syllables was included to account for effects of syllable shortening [19]. The number was assessed by counting the prototypical number of syllables of

[4]Find the relevant data and R script at https://osf.io/dvrq4/.

the target words. As participants were instructed to read clearly and in an average pace, this prototypical number of syllables is also the number that was produced.

NUMBER. As it is unclear whether the number of the target word makes a difference regarding the duration of word-final *-er*, this information was included in the analysis as a binary variable with the values `singular` and `plural`.

STEREOTYPICALITY. Stereotypicality was included in the analysis, as it might make a difference whether a given masculine noun is stereotypically female or male. Stereotypicality data are adopted from [12].

SPEECHRATE. As speech rate typically shows a direct influence on the duration of segments, where a higher speech rate comes with generally shorter segments, this variable was incorporated into the analysis. Speech rate was measured using a Praat script by [20].

TRIALNUMBER. The number of the trial was included to account for potential effects of priming, training, and fatigue. The order of trials was fully randomised.

GENDER. As the usage of generic masculines is directly connected to the notion of gender, speaker gender was part of the model formula. This variable takes the values `female`, `male`, and `nonbinary`.

AGE. Speaker age was included for two reasons. First, it may show an effect on phonetic realisation. Second, age may act as a proxy for the attitude towards more progressive topics, such as gender and language.

ATTGM, ATTPF, ATTST. To get a more direct idea of a speaker's attitude towards gender and language and its potential influence on phonetic realisation, speakers' ratings on the usage of generic masculines, pair forms, and gender star forms were included. Each of the three variables takes integer values of the range $[1,5]$, where 1 represents the opinion that the usage of a given form is *very bad* and 5 represents the opinion that the usage of a given form is *very good*.

Further, the following variables were introduced as random intercepts:

SPEAKER. To account for inter-speaker differences not accounted for by the already introduced speaker-dependent variables, the ID of each speaker was specified as a random intercept term.

WORD. Similarly, to account for inter-word differences not reflected in the already introduced variables, the word was specified as a random intercept term.

To avoid issues of collinearity that might lead to unreliable model estimates [9], all predictor variables were checked for problematic levels of correlation (i.e., $|\rho| \geq 0.5$). Strong correlations were found for DURBASE and SYLNUMBER ($\rho = 0.75$), and ATTST and ATTGM ($\rho = -0.67$). To proceed, SYLNUMBER and ATTGM were disregarded.

Then, continuous variables were checked for their distributions. It was found that neither `durEr` nor `durBase` nor `speechRate` followed a normal distribution. To achieve a more normal distribution and, with that, a better model fit, `durEr` was log-transformed. The other two variables were still not normally distributed after log-transformations, and thus remained in their original forms.

With the retained and transformed variables, two initial linear mixed-effects regression models were fitted to log-transformed dependent variables. One was fitted to the absolute duration of word-final *-er*, `durEr`, and one was fitted to the relative duration of word-final *-er*, `durErRel`. The latter model was included following previous studies on the duration of different types of word-final /s/ [1, 17, 21] and is motivated by the finding that differences in duration may play out differently in absolute and relative segment duration [22]. For both dependent variables, the initial model was specified with the aforementioned fixed effects and random intercepts.

The initial model was then reduced stepwise using the `step` function provided by the lmerTest package [23]. This stepwise reduction process fits models repeatedly, performing a combination of forward selection and backward elimination, adding and removing predictors iteratively. The fitted models are compared via their AIC values. The result of this process is the model formula resulting in the lowest AIC, i.e. best model fit.

Finally, the reduced model was trimmed based on its residuals to obtain an even better model fit [24]. Data points with residuals further than 2.5 standard deviations from the mean were removed. The result of this last step were the final models.

3 Results

The final model fitted to the absolute duration of *-er* contained only the predictor of interest, TYPEOFER, and the two random intercepts, SPEAKER and WORD. The model fitted to the relative duration of *-er* additionally contained DURBASE as fixed effect. During the residual trimming process of both models, $n = 19$ data points (1.7%) were removed. Table 2 provides summaries for both models.

Table 2 – Summary of the final model fitted to the absolute *-er* durations.

Absolute duration model	**Estimate**	**Std. Error**	**df**	***t*-value**	***p*-value**
Intercept	-2.483	0.038	46.52	-65.580	<0.001
TYPEOFERGMu	<0.001	0.018	1036.00	-0.052	0.959
TYPEOFERSM	-0.250	0.018	1036.00	-13.998	<0.001
Relative duration model	**Estimate**	**Std. Error**	**df**	***t*-value**	***p*-value**
Intercept	-2.494	0.055	106.0	-45.658	<0.001
TYPEOFERGMu	0.001	0.018	1035.0	0.056	0.955
TYPEOFERSM	-0.249	0.018	1037.0	-13.924	<0.001
DURBASELOG	-1.015	0.052	319.8	-19.336	<2e-16

Bonferroni-corrected pairwise comparisons for the levels of TYPEOFER in both models showed that the durational differences between the specific masculine and either of the generic masculines was highly significant ($p < 0.001$), while the difference between the specified and unspecified generic masculine was not ($p = 1$). The effect size of TYPEOFER is large with $\eta^2 = 0.20$ with 95% $CI = [0.17, 1.00]$ for the absolute duration and with $\eta^2 = 0.20$ with 95% $CI = [0.48, 1.00]$ for the relative duration of *-er*.

The effects of TYPEOFER and DURBASE are illustrated in Figure 1. In both models, specific masculines come with significantly shorter *-er* durations than both types of generic masculines. For DURBASE it is found that shorter base durations come with shorter *-er* durations.

4 Discussion

The present study set out to find a first answer to whether fine-semantic differences in words with identical lexical and morphological makeup lead to fine-phonetic differences. Analysing the productions of specific masculines, specified generic masculines, and unspecified generic masculines embedded within disambiguating sentences, evidence was found in favour of such fine-phonetic differences caused by fine-semantic differences.

RQ1, which asked whether there are durational differences between specific masculines and generic masculines, may be answered positively. In both absolute and relative durations,

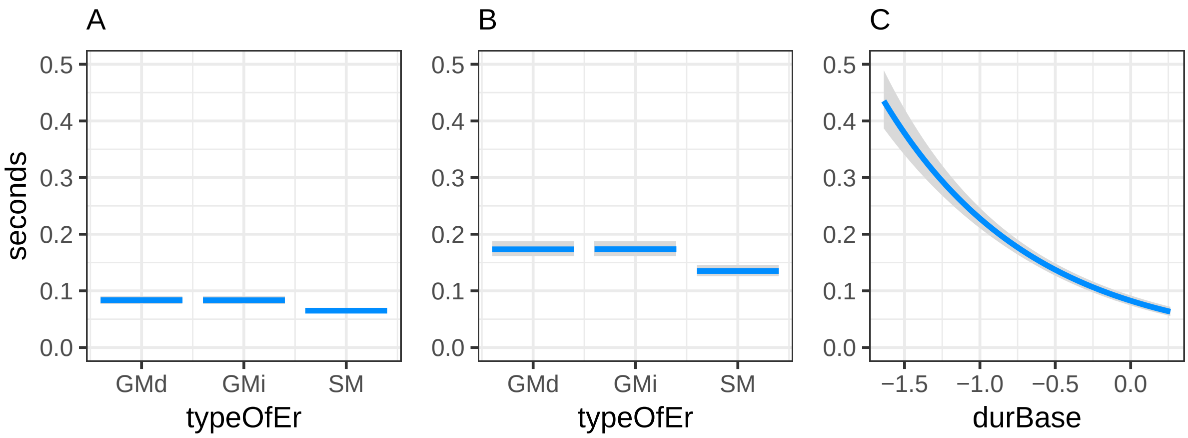

Figure 1 – Partial effects of TYPEOFER as predicted by the model fitted to absolute *-er* duration (Panel A) and fitted to relative *-er* duration (Panel B), as well as the partial effect of DURBASE in the model fitted to relative *-er* duration (Panel C).

-er is produced significantly longer when it is part of a generic masculine than when it is part of a specific masculine.

Regarding **RQ2**, which asked whether durational differences are also to be found between specified and unspecified generic masculines, no such difference was found. Both specified and unspecified generic masculines are produced with highly similar durations.

Taking the results together, it appears that the fine-semantic difference between specific and generic usage leads to a durational difference in a role noun's word-final suffix, i.e. in *-er*. Whether the generic usage is referring to a gender-specified individual or to an individual whose gender is not specified seems to not make a difference.

These novel findings raise the question of why these differences exist. One potential explanation may lie in the cognitive load required to parse specific and generic masculines. While specific masculines are relatively easily parsed–a masculine form is used to refer to a male individual–generic masculines are not. Here, a grammatically masculine form that may be used to refer to male individuals is used to refer to non-male individuals, even though a feminine counterpart for referring to female individuals exists. Comprehension efforts are assumably higher [‘cognitive markedness’, cf. 25], which in turn may lead to longer durations in production. One potential framework to test this idea in is that of the Discriminative Lexicon [8].

Finally, the present findings add to the results of previous research on subphonemic durational differences caused by morphological or lexical differences. That is, not only lexical or morphological differences in phonologically identical words or segments lead to fine-phonetic durational differences, but so do fine-semantic differences in lexically, morphologically, and phonologically identical words. This, then, adds to the evidence calling into question assumptions made by traditional feed-forwards models and psycholinguistic models of speech production [5, 6], as these do not account for such differences.

5 Conclusion

This study was the first to investigate whether fine-semantic differences cause fine-phonetic differences in lexically, morphologically, and phonologically identical words. Making use of specific and generic masculines in German in a reading task, it was found that such fine-semantic differences come with subphonemic durational differences. These findings add to the body of evidence that calls into question established models of speech production. The results call not only for a revision of established models of speech production, but also for further research in this area, exploring why these differences exist.

References

[1] PLAG, I., J. HOMANN, and G. KUNTER: *Homophony and morphology: The acoustics of word-final S in English*. *Journal of Linguistics*, 53(1), pp. 181–216, 2017.

[2] HEDIA, S. B. and I. PLAG: *Gemination and degemination in english prefixation: Phonetic evidence for morphological organization*. *Journal of Phonetics*, 62, pp. 34–49, 2017.

[3] BLAZEJ, L. J. and A. M. COHEN-GOLDBERG: *Can we hear morphological complexity before words are complex?* *Journal of Experimental Psychology: Human Perception and Performance*, 41(1), p. 50, 2015.

[4] LOHMANN, A.: *Time and thyme are not homophones: A closer look at gahl's work on the lemma-frequency effect, including a reanalysis*. *Language*, 94, pp. e180–e190, 2018.

[5] KIPARSKY, P.: *Lexical morphology and phonology*. In I.-S. YANG (ed.), *Linguistics in the morning calm: Selected papers from SICOL1*, pp. 3–91. 1982.

[6] LEVELT, W. J., A. ROELOFS, and A. S. MEYER: *A theory of lexical access in speech production*. *Behavioral and brain sciences*, 22(1), pp. 1–38, 1999.

[7] PIERREHUMBERT, J. B.: *Exemplar dynamics: Word frequency, lenition and contrast*. *Typological studies in language*, 45, pp. 137–158, 2001.

[8] CHUANG, Y.-Y. and R. H. BAAYEN: *Discriminative learning and the lexicon: NDL and LDL*. *Oxford Research Encyclopedia of Linguistics*, 2021.

[9] TOMASCHEK, F., P. HENDRIX, and R. H. BAAYEN: *Strategies for addressing collinearity in multivariate linguistic data*. *Journal of Phonetics*, 71, pp. 249–267, 2018.

[10] SCHMITZ, D., I. PLAG, D. BAER-HENNEY, and S. D. STEIN: *Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning*. *Frontiers in Psychology*, 12, 2021.

[11] DIEWALD, G.: *Zur Diskussion: Geschlechtergerechte Sprache als Thema der germanistischen Linguistik – exemplarisch exerziert am Streit um das sogenannte generische Maskulinum*. *Zeitschrift für germanistische Linguistik*, 46, pp. 283–299, 2018.

[12] MISERSKY, J., P. M. GYGAX, P. CANAL, U. GABRIEL, A. GARNHAM, F. BRAUN, T. CHIARINI, K. ENGLUND, A. HANULIKOVA, A. ÖTTL, J. VALDROVA, L. V. STOCKHAUSEN, and S. SCZESNY: *Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak*. *Behavior Research Methods*, 46, pp. 841–871, 2014.

[13] BOERSMA, P. and D. WEENINK: *Praat: Doing phonetics by computer*. 2019. URL `http://www.praat.org/`.

[14] LADEFOGED, P.: *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Wiley, 2003.

[15] BORIL, T. and R. SKARNITZL: *Tools rPraat and mPraat*. In P. SOJKA, A. HORÁK, I. KOPECEK, and K. PALA (eds.), *Text, Speech, and Dialogue: 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016, Proceedings*, pp. 367–374. Springer International Publishing, Cham, 2016.

[16] TEAM, R. C.: *R: A language and environment for statistical computing*. 2021. URL `https://www.r-project.org/`.

[17] SCHMITZ, D., D. BAER-HENNEY, and I. PLAG: *The duration of word-final /s/ differs across morphological categories in English: Evidence from pseudowords*. *Phonetica*, 78(5-6), pp. 571–616, 2021.

[18] BATES, D., M. MÄCHLER, B. M. BOLKER, and S. C. WALKER: *Fitting linear mixed-effects models using lme4*. *Journal of Statistical Software*, 67, 2015.

[19] SCHMITZ, D., H.-E. CHO, and H. NIEMANN: *Vowel shortening in German as a function of syllable structure*. In *Proceedings 13. Phonetik und Phonologie Tagung (P&P13)*, pp. 181–184. 2018.

[20] DE JONG, N. and T. WEMPE: *Praat script syllable nuclei [Praat script]*. 2008. URL `https://sites.google.com/site/speechrate/Home/praat-script-syllable-nuclei-v2`.

[21] SCHMITZ, D. and D. BAER-HENNEY: *Morphology renders homophonous segments phonetically different: Word-final /s/ in German*. In *Proceedings of Speech Prosody 2024 (SP2024)*. Universiteit Leiden, Netherlands, 2024.

[22] BAAYEN, R. H., Y.-Y. CHUANG, E. SHAFAEI-BAJESTAN, and J. P. BLEVINS: *The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning*. *Complexity*, 2019, p. 4895891, 2019.

[23] KUZNETSOVA, A., P. B. BROCKHOFF, and R. H. B. CHRISTENSEN: *lmertest package: Tests in linear mixed effects models*. *Journal of Statistical Software*, 82, 2017.

[24] BAAYEN, R. H. and P. MILIN: *Analyzing reaction times*. *International Journal of Psychological Research*, 3, pp. 12–28, 2010.

[25] HASPELMATH, M.: *Against markedness (and what to replace it with)*. *Journal of Linguistics*, 42(1), p. 25–70, 2006. doi:10.1017/S0022226705003683.