2025. Proc Ling Soc Amer 10(1). 5905. https://doi.org/10.3765/plsa.v10i1.5905.

Semantic change of female-denoting nouns in diachronic German corpora

Carla Sökefeld & Patrícia Amaral*

Abstract. Computational studies of language change using word embedding models have focused on lexical semantic shifts. Current work in Natural Language Processing has revealed gender bias in word embeddings trained on internet data. However, little is known about what distributional methods can reveal about possible gender bias in diachronic corpora. This paper addresses this gap by examining gender-specific nouns denoting humans, e.g. *Magd* 'maidservant', *Knecht* 'farmhand' in a historical corpus of German (1350-1899). We show that word embedding models provide evidence for gender bias and negative evaluations associated with female-denoting terms in diachronic German corpora.

Keywords. semantic change; gender bias; pejoration; word embeddings; distributional semantics; corpus linguistics

1. Introduction. Recent work in Natural Language Processing has revealed gender bias in word embeddings trained on internet data (e.g., Caliskan et al. 2017, 2022). For instance, Caliskan et al. (2022) show that static English word embeddings trained on Internet data provide ample evidence for biases regarding men and women. Specifically, they find that there are significant (i) differences in frequencies of words associated with men vs. women, (ii) differences in semantic categories (e.g. power vs. family) associated with men vs. women, and (iii) differences in parts-of-speech, with verbs being generally associated with men, and adjectives and adverbs being associated with women. At the same time, from a historical perspective, it is well-established that across languages, female-denoting nouns undergo semantic shifts characterized as pejoration, i.e. they tend to acquire derogatory meanings (e.g., Meillet 1926; Campbell 2013; Hock & Joseph 1996). As Ullmann notes, pejoration is the result of "prejudice against groups, specifically social classes and occupations" (Ullmann 1979:232). Meillet exemplifies this tendency with the change affecting both French nouns *garce* and *fille*, which "ont été appelés successivement à fournir le nom de la fille publique" (Meillet 1926:242), 'which were recruited successively for the name of the prostitute' (our translation). Despite these observations and mentions in the literature, there is a lack of empirically-based studies of semantic change of female-denoting nouns cross-linguistically. Likewise, little is known about what distributional methods can reveal about possible gender bias in diachronic corpora. This paper fills this gap by adopting a distributional approach to examine semantic shifts in gender-specific nouns in the history of German. While we focus on one language, we expect that our methods and findings can provide insight into tendencies of lexical semantic change as well as the affordances of distributional methods for this type of research in other languages.

The results of the current study are based on research conducted as part of the PhD project of the first author on the semantic change and pejoration of gender-specific personal nouns with a basic 'female person' or 'male person' meaning in German. Personal nouns are defined as nouns that are used to refer to a human referent (Elmiger 2018), such as occupational terms (*Ingenieur* 'engineer') or kinship terms (*Geschwister* 'sibling'). The personal nouns that are the subject of

* Acknowledgments: This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute. Authors: Carla Sökefeld, Universität Hamburg (carla.soekefeld@uni-hamburg.de) & Patrícia Amaral, Indiana University (pamaral@iu.edu).

this study were chosen because they constitute basic personal nouns in the sense that, at least at some point in language history, they foremost mark gender, referring to a wide range of (male or female) referents without marking more specialized meanings like e.g. occupation (though some develop such a meaning over time). It is precisely these basic (female) personal nouns that most famously underwent pejoration in German.

In this paper, we first review previous work on the gender-specific semantic change of personal nouns in German as well as on computational studies of semantic change. In section 3, we introduce the data and method used to train distributional word embedding models, before presenting the results of the study in section 4. We discuss the way dimensions of meaning can be traced with the help of word embeddings models in section 5, which is followed by a brief conclusion in section 6.

2. Previous studies.

2.1. GENDER-SPECIFIC SEMANTIC CHANGE IN THE HISTORY OF GERMAN. Nübling (2011) describes three distinct paths of pejoration along which basic female personal nouns underwent semantic change: Social degradation, functionalization, and sexualization. Social degradation constitutes the loss of a 'high status' meaning or gain of negative expressivity, as in the case of *Weib* 'woman' > 'woman (derogatory)' and *Frau* 'woman of high social status' > 'woman'. Functionalization pertains to the gain of low-status 'occupation' meanings such as 'servant' or 'menial worker', e.g. *Magd* 'girl, virgin' > 'maidservant' and *Mamsell* 'young woman of high social status' > 'servant, worker'. The latter example demonstrates that personal nouns can be affected by multiple paths of pejoration. Finally, sexualization describes the gain of sexualized meanings, either in the form of a 'virginity' or 'promiscuity' meaning, e.g. *Jungfrau* 'young woman of high social status' > 'virgin', *Metze* 'girl, servant' > 'promiscuous girl; sex worker'.

Whereas earlier explanations (Keller 1994) see the inflationary use of "high-status" personal nouns for lower-status referents as the reason for these pejorations, Nübling (2011) identifies bias and the societal derogation of women as the cause. Research conducted on the depiction of women in historical dictionaries (Blankenberger 2003; Warnke 1993) corroborates the prevalent linguistic derogation of women. In identifying and describing gender biases present in historical corpora, our work sheds light on the probable causes of the widespread pejoration of female personal nouns in German.

2.2. COMPUTATIONAL STUDIES OF SEMANTIC CHANGE USING THE DISTRIBUTIONAL APPROACH. According to the distributional approach to meaning, we can investigate and capture word meaning through the contexts in which the word occurs (Harris 1954; Firth 1957). Hence, similarity in word meaning is measured as occurrence in the same linguistic environments. Linguistic environment is defined over the set of all words of a language, as represented in the corpus being used. Distributional models provide insight into word meaning by identifying a network of word associations that may include synonyms, antonyms, (co-)hyponyms, hyperonyms, as well as words that are syntagmatically related to the target word (e.g. due to the argument structure of a predicate). Such nearest neighbors of a target word in the embedding space may also include words broadly associated with the target word and depend heavily on the corpus used.

While context dependence has been seen as a limitation of these methods (Lenci 2018), from the perspective of semantic change and detection of pejoration, this may be an asset, as distributional methods provide a window to examine language usage as represented by textual genres

and discourse conventions. Such methods have proved to be not just powerful tools to represent meaning, but also to provide insight into changes in meaning over time (see e.g., Tahmasebi et al. 2021; Geeraerts et al. 2024).

Word embeddings, i.e. vector representations of meaning based on co-occurrence statistics in a corpus, are usually obtained from training language models on simple tasks like predicting a word from its surrounding context (Jurafsky & Martin 2025:118). During this training step, representations for each vocabulary item in a corpus in the form of multi-dimensional vectors are learned, which reflect the usage contexts in which the words appear. The closer two vectors are in the multi-dimensional embedding space, the more similar the usage contexts of the two corresponding vocabulary items are and therefore, the more similar their meaning can be considered. The closeness (or distance) of two vectors is measured by way of cosine similarity (Jurafsky & Martin 2025:110).

Using this method, we can study changes in the meaning of a word by tracing changes in its nearest neighbors, i.e. words that are close to it in semantic space, measured through cosine similarity. Commonly, instead of considering the absolute distance between vectors, neighbors are ranked by their closeness, so that e.g. the 30 nearest neighbors of a target word are analyzed (see e.g., Hu et al. 2022; Amaral et al. 2023). In a seminal paper that established the potential of these methods, the semantic change of *gay*, *broadcast*, and *awful* can be shown by a change in their nearest neighbors over the decades, as can be seen in Figure 1, from Hamilton et al. (2016).

Figure 1. Two-dimensional visualization of the changes in the semantic space of *gay*, *broadcast*, and *awful* over time, from (Hamilton et al. 2016:Figure 1).

Building on previous research that showed good results for studies of semantic change (Hu et al. 2022:a.o.), the algorithm we use is Skip-Gram with Negative Sampling (SGNS) implemented in word2vec (Mikolov et al. 2013), which produces static embeddings, i.e., one embedding per vocabulary item, and has been shown to perform well on lexical semantic change detection tasks (Schlechtweg et al. 2020). We explore the information provided by these embeddings to study pejoration and gender bias in the history of German.

3. Investigating the semantic change of gender-specific nouns.

3.1. CORPORA USED AND MODEL TRAINING. To create a diachronic corpus spanning the investigation period from 1350 to 1899, we used data from two different corpus projects. First of all, the *Referenzkorpus Frühneuhochdeutsch*[1] (ReF) was used, a reference corpus of Early New High German (1350-1650). It contains roughly 3.4 million tokens from six different text types

[1] https://www.ruhr-uni-bochum.de/wegera/ref/, last accessed March 11, 2025.

(chronicles and reports, uplifting texts, clerical and theological texts, practical texts, law and business texts, and recreational texts). Secondly, the *Deutsches Textarchiv*[2] (DTA) as well as its extension DTA-e was used, which contains texts from 1472 to 1987 belonging to four different text types (fiction, functional literature, science, and newspaper). The vast majority of texts was written by male authors, as female literacy and participation in literality was restricted throughout much of history.

All texts from the two sources dated between 1350 and 1899 were incorporated, arriving at a corpus of roughly 221 million tokens. The corpus was divided into time slice sub corpora spanning 50 years each (1350-1399, 1400-1449, etc.) so that semantic changes of the target lexemes could be traced. Because of different availability of data, sub corpus sizes differ substantially. The smallest time slice (15_1) contains roughly 360.000 tokens, whereas the biggest time slice (19_2) contains more than 64 million tokens.

In order to obtain more robust results for the earliest, data-sparse time slices, we pre-trained a base model on all 18 million tokens of ENHG data (1350-1649). This base model was subsequently used to initialize the training on the time slice sub corpora (Kim et al. 2014), which yields more stable results while still keeping the data basis appropriate for the language stage (Kaiser et al. 2021). The model training process is shown in Figure 2.

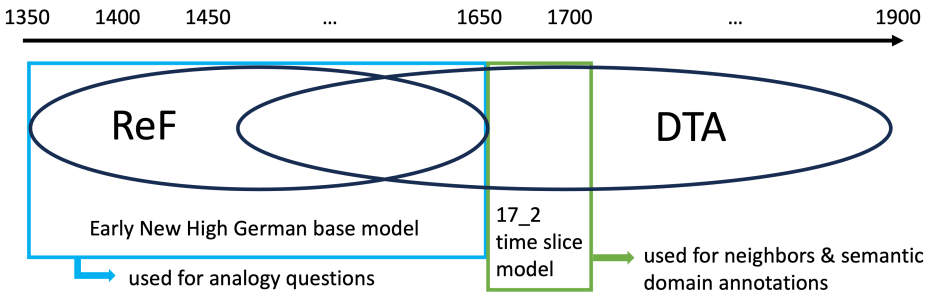

Figure 2. Corpora used for model training

The hyperparameter settings used for model training are context window = 5, vector dimensions = 100, frequency threshold = 5. These settings were determined empirically as yielding the best results on analogy evaluation tests (cf. Mikolov et al. 2013), see Section 4.2. The neighbors were obtained by averaging the words' ranks over 5 runs of training in order to stabilize results (cf. Hu et al. 2022).

Orthographic variation is a notorious problem for studies on historical language stages (cf. Hu et al. 2022; Amaral et al. 2023). In order to mitigate this problem, the embedding models were trained on the level of lemmas, all tokens were converted to lower case, and persisting spelling variants of the target personal nouns on the lemma level were normalized to one form beforehand (e. g., *knechtt* and *chnehte* were corrected to *knecht*).

[2] https://www.deutschestextarchiv.de/, last accessed March 11, 2025.

4. Results of this study. We took two different approaches to the analysis of gender bias. First of all, we analyzed the nearest neighbors of personal nouns across time slices (Section 4.1). Secondly, we conducted an analysis of the gender analogy results concerning 'power' and 'family' personal nouns (Section 4.2).

4.1. ANALYSIS OF GENDER-SPECIFIC PERSONAL NOUNS OVER TIME. We manually annotated semantic domains for nearest neighbors of 15 female and 12 male personal nouns, as exemplified in Table 1.

Semantic domain	Examples of neighbors
Age	*steinalt* 'ancient', *zwölfjährig* 'twelve years old'
Appearance	*rotbackig* 'red-cheeked, *bärtig* 'bearded'
Family	*familienvater* 'father of a family', *schwester* 'sister'
Sexuality	*reyne* 'pure', *jungfraw* 'virgin'
Negative evaluation	*verbrecherisch* 'criminal', *schurke* 'villain'
Positive evaluation	*hochverehrt* 'esteemed', *biederman* 'man of integrity'
Work	*kochin* 'cook', *einhirtesein* 'to be a shepherd'
Romance/marriage	*wittwe* 'widow', *heiratsfähig* 'marriageable'

Table 1. Examples of semantic domains according to our annotation scheme

Taking all the nearest neighbors of the personal nouns for the 17_2 and 19_2 time slices, we can observe some general tendencies in our results.[3] Overall, we find more marriage- and sexuality-related neighbors for female than for male personal nouns, whereas work-related neighbors make up the biggest semantic domain for male personal noun.

We present *Magd* and *Knecht* as example personal nouns, which in contemporary German have the same dictionary meaning in terms of descriptive content 'servant/worker (working predominantly in the agricultural domain)' and only differ with respect to gender. Such a comparison can provide insight into gender-specific distinctions. Table 2 and Table 3 show the 15 nearest neighbors of *Magd* and *Knecht* in the two time slices 16_1 and 17_2, selected to illustrate the semantic change process of the two personal nouns. As for the general tendencies described above, we find sexuality-related neighbors for *Magd*, whereas work-related neighbors make up most of the neighbors of *Knecht*.

Magd used to have the meaning 'girl, virgin' into the 16th century, which is still observable in 16_1 through the two neighbors *reyne* 'pure' and *jungfraw*, which took on a 'virgin' from the late 15th century onwards. There are only two work-related neighbors indicating 'servant' usages in 16_1, *färber* 'dyer' and *weingartner* 'wine grower'. In contrast, the two nearest neighbors in 17_2 as well as other lower-ranking neighbors (e.g. *dienst-bott* 'servant') point to work-related usages and a functionalization of *Magd*. Moreover, there are several proper names among the neighbors. Two of these are the names of literary characters and biblical characters naming maid servants (*Rizarize*, *Hagar*), which correspond to the 'servant' meaning of *Magd*. We return to the topic of proper names as nearest neighbors below. Furthermore, there are still neighbors related to sexuality in 17_2, but in contrast to 16_1, they no longer relate to virginity. Instead, the third-nearest neighbor *dirne* 'servant, prostitute' as well as lower-ranking neighbors *hure* 'whore' and

[3] Time slice 17_2 was chosen because it is the first sub corpus in which enough data is available for the majority of the personal nouns. Time slice 19_2 was chosen to allow for a maximal diachronic comparison.

Neighbors of Magd 16-1		**Neighbors of Magd** 17-2	
word	Rank	word	Rank
reyne ‘pure’	1	knecht ‘servant, menial’	1
gebärmutter ‘uterus’	2	gesinde ‘menial staff’	2
latona (female first name)	3	dirne ‘servant, prostitute’	3
hertsens ‘of the heart’	4	weib ‘woman (low status)’	4
jungfraw ‘virgin’	5	rizarize (female first name)	5
färber ‘dyer’	6	frau ‘woman (high status)’	6
zehendt ‘tenth’	7	dienst-bott ‘servant’	7
viertagig ‘four-day’	8	magdelein ‘maidservant, girl (dim.)’	8
gboren ‘born’	9	sarä (female first name)	9
weingartner ‘wine grower’	10	junker ‘squire’	10
wagn ‘to dare’	11	raguel (female first name)	11
lustich ‘funny’	12	hure ‘whore’	12
gelbsucht ‘jaundice’	13	hagar (female first name)	13
spielend ‘playing’	14	kochin ‘cook’	14
gewasche ‘washed’	15	concubine ‘concubine’	15

Table 2. Top 15 nearest neighbors of *Magd* in the 1st half of the 16th c. (16-1) and the 2nd half of the 17th c. (17-2)

concubine ‘concubine’ point to sexualizing usages of *Magd* in the context of promiscuity and sex work.

Knecht, on the other hand, used to have a broad ‘male person, man’ meaning but took on a ‘servant’ meaning over time. This can already be observed in 16_1 with e.g. the nearest neighbor *diener* ‘servant’. In 17_2, all of the 15 nearest neighbor relate to menial work or subordination (e.g. *leibeigen* ‘in bondage’, *sklave* ‘slave’), showing that male personal nouns are not exempt from functionalization. Nevertheless, it is overwhelmingly female personal nouns that exhibit sexualization as shown in the neighbors of *Magd*.

In terms of semantic change, we see a change in the semantic domains of the neighbors of our target words over time, with an increase in work-related neighbors and an increase in generic negatively-evaluated neighbors for female personal nouns like *Magd*. Here we identify a difference when we compare the neighbors of *Knecht* in the two time periods. While male personal nouns like *Knecht* also see an increase of work-related neighbors, neighbors related to sexuality - which are often negative - predominantly appear for female personal nouns.

4.2. ANALYSIS OF ANALOGY TESTS. As mentioned above, we conducted analogy tests to evaluate different hyperparameter settings and determine the best settings for model training. Analogy tests are based on the notion that the strength of an embedding model in representing semantic similarity can be examined “performing simple algebraic operations with the vector representation of words” (Mikolov et al. 2013:5). For example, a good model should be able to represent the relation between gendered word pairs such as *king* - *queen* and *man* - *woman*: If the embeddings accurately represent word meaning, subtracting the vector of *king* from the vector of *queen* and then adding the vector of *man* should result in a point in the vector space close to the embedding of *woman*. These analogies can likewise be tested for other semantic or grammatical relations, like the singular and plural of nouns as shown in Figure 3.

One section of tests used to evaluate the model hyperparameters in our study concerns such gender analogies of personal nouns. Three types of gendered word pairs were used (see Table 4):

Neighbors of Knecht 16-1		**Neighbors of Knecht** 17-2	
word	Rank	word	Rank
diener ‘servant’	1	diener ‘servant’	1
haußuater ‘housefather’	2	magd ‘menial servant’	2
hauszuatter ‘housefather’	3	leibeigen ‘in bondage’	3
gsegn ‘to bless’	4	leibeigene ‘in bondage’	4
oberster ‘highest’	5	knechten ‘to subjugate’	5
knechten ‘to subjugate’	6	geselle ‘journeyman’	6
pförtner ‘gatekeeper’	7	leibeigner ‘bondsman’	7
einhirtesein ‘to be a shepherd’	8	mit-knecht ‘fellow servant’	8
kriegesmann ‘warrior’	9	gesinde ‘menial staff’	9
fürend ‘leading’	10	leib-eigen ‘in bondage’	10
frömder ‘stranger’	11	dienerin ‘servant (f.)’	11
scheit ‘log’	12	mitknecht ‘fellow servant’	12
bezal ‘to pay’	13	sklave ‘slave’	13
frumer ‘pious, devout’	14	burgmann ‘soldier’	14
waltherus (name)	15	stalljunge ‘stableboy’	15

Table 3. Top 15 nearest neighbors of *Knecht* in the 1st half of the 16th c. (16-1) and the 2nd half of the 17th c. (17-2)

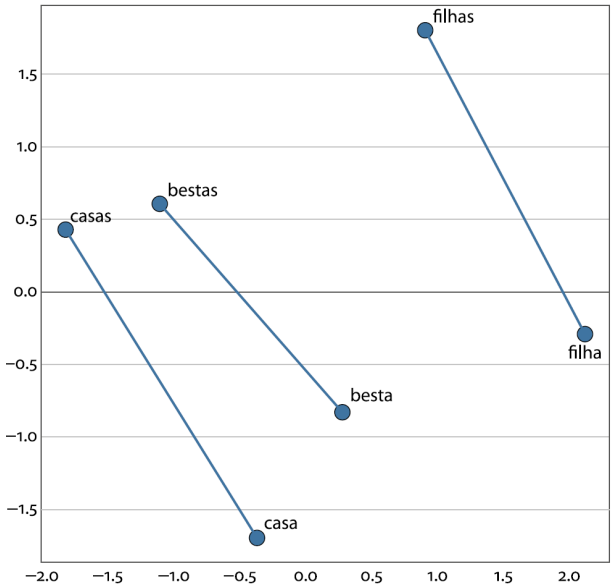

Figure 3. Visual representation of analogy for singular-plural nouns in Medieval Portuguese, from Tian et al. (2021)

“power” terms that denote a person in a position of ruling power, e. g. *herzog* ‘duke’, *herzogin* ‘duchess’; “family” terms denoting kinship, e. g. *bruder* ‘brother’, *schwester* ‘sister’; and general “gender” words, e. g. *männlich* ‘male’ and *weiblich* ‘female’. These categories were mainly chosen because they consist of gendered word pairs and do not interfere with the target category of basic ‘female person’ and ‘male person’ lexemes.

The evaluation results reveal that the embedding models perform poorly at resolving an analogy question such as *fürstin* is to *fürst* as *gemahlin* is to *gemahl*, in which the model is supposed

Domain	Word pairs
general gender	*er* 'he' - *sie* 'she', *männlich* 'male' - *weiblich* 'female'
family	*Bruder* 'brother' - *Schwester* 'sister', *Gemahl* 'spouse (m.)' - *Gemahlin* 'spouse (f.)
family	*Sohn* 'son' - *Tochter* 'daughter', *Vater* 'father' - *Mutter* 'mother'
power	*Erzherzog* 'archduke' - *Erzherzogin* 'archduchess',
	Freiherr 'baron' - *Freiherrin* 'baroness'
power	*Fürst* 'prince' - *Fürstin* 'princess', *Graf* 'count' - *Gräfin* 'countess'
power	*Herzog* 'duke' - *Herzogin* 'duchess', *Kaiser* 'emperor' - *Kaiserin* 'empress'
power	*Kurfürst* 'elector' - *Kurfürstin* 'electress', *König* 'king' - *Königin* 'queen'

Table 4. Gendered word pairs in the three domains "power", "family" and "general" used for the analogy questions

to arrive close to the vector of *gemahl* by subtracting and adding the vectors of the three given words from each other. But rather than simply treating such failed gender analogies as a computational problem to be solved, a closer investigation shows that they in fact reveal gender biases present in the data.

First of all, the semantic category of the analogy question influences the gender of the resulting personal noun. In questions with kinship as a target category, the analogy results more often in female words, even if a male word would be expected. For example, in the "family" question *gemahlin* 'spouse (f.)' - *gemahl* 'spouse (m.)'; *schwester* 'sister', the male *bruder* 'brother' (or another male kinship term) would be expected. Instead, the result is the female kinship term *tochter* 'daughter'. The opposite is true for questions with "power" as the target category. For example, in the question *fürst* 'prince' - *fürstin* 'princess'; *erzherzog* 'archduke', the result is the male first name *leopold* instead of the intended female "power" term *erzherzogin* 'archduchess'.

Analogy questions can be asked in two directions, with either a female or a male personal noun as the target. Figure 4 shows the results by target category for questions where the target is expected to be male. Although in both categories, more than half of the results are indeed male, the proportion of female results increases in the "family" category. Figure 5 shows the results for the questions where the the target is expected to be female, and indeed more than half of the results are female in the "family" questions. But among the "power" questions, only 40 % of the results are female, and there are almost as many male as female resulting words. These differences in proportion indicate that the semantic category overrides the gender information of the personal nouns. Although the number of analogy questions, especially for the "family" category, is rather low, a tendency is clearly observable. The difference in proportions of male and female results by expected target gender for the "power" category is statistically significant (χ^2=10.69, p $<$0.05). Future work should verify the influence of semantic target category on a larger sample of analogy questions.

An inspection of the semantic neighborhood of these "power" terms reveals that there are systematic differences between female and male nouns. Among the 20 nearest neighbors of the seven different male and female "power" personal nouns, there are five discernible semantic groups: place names (e. g. *fürstenberg*), other 'position of power' personal nouns (e. g. *regentin* 'regent (f.)', *markgraf* 'margrave'), personal names (e. g. *leonora*, *maximilian*), words denoting marriage or sexuality (e. g. *vermählt* 'married', *vestalischen* 'vestal; virginal'), and positive

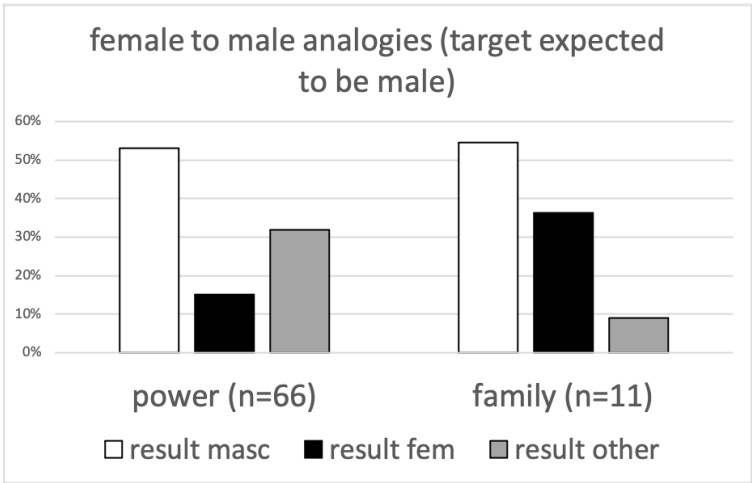

Figure 4. Proportions of female and male results by semantic category in "female to male" analogy questions

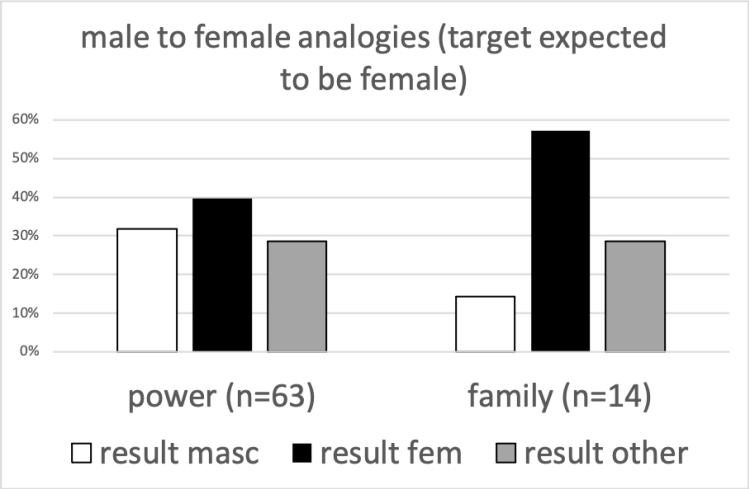

Figure 5. Proportions of female and male results by semantic category in "male to female" analogy questions

adjectives (e. g. *hochgebore* 'high-born', *vielehrentugendreich* 'much honorably virtuous').

Male "power" terms have almost twice as many place names as nearest neighbors, indicating associations with landownership. While proper names and other "power" terms appear as neighbors regardless of gender, only female personal nouns have neighbors pertaining to marriage or sexuality. Although not very frequent for female "power" terms, either, neighbors like *gemahl* 'spouse (m.)' or *vermählt* 'married' do not appear at all for their male counterparts. Table 5 presents the ten nearest neighbors of the "power" personal nouns *Gräfin* 'countess' and *Graf* 'count' in the ENHG base model, exemplifying the differences in the semantic neighborhoods between male and female "power" terms.

Our results point to considerable differences in the distributions of gendered word pairs such as *Gräfin* and *Graf*. Considering that neither "power" nor "family" lexemes of the gender-specific

Neighbors of Gräfin 'countess'		**Neighbors of Graf 'count'**	
Word	Rank	Word	Rank
kurfürstin 'electress'	1	ortenburg (place name)	1
herzogin 'duchess'	2	hardeck (place name)	2
vermählt 'married'	3	buchheim (place name)	3
elisabethen (first name, f.)	4	altheim (place name)	4
markgrafin 'margravine'	5	colalto (place name)	5
sibylle (first name, f.)	6	solms (place name)	6
leonora (first name, f.)	7	nassau (place name)	7
sabinae (first name, f.)	8	schwarzburg (place name)	8
flansen (place name)	9	frstenberg (place name)	9
freiin 'baroness'	10	freiherr 'baron'	10

Table 5. Top 10 nearest neighbors of *Gräfin* and *Graf* in the ENHG base model.

personal nouns seem to be used analogously in the historical data, it is only logical that gender analogies fail.

5. Discussion: Tracing dimensions of meaning over time. Looking at the neighbors obtained through our models in both time periods for *Magd* vs. *Knecht*, we see some commonalities and differences. Both nouns have similar work-related words. However, *Magd* has neighbors denoting organs of the body or bodily functions (e.g., *gebärmutter, gboren*) and neighbors associated with sexuality (e.g. *jungfraw, dirne*). For *Knecht*, there are no neighbors associated with the semantic domains of bodily functions or sexuality.

The comparison of the semantic domains represented in the nearest neighbors of the female-denoting personal noun *Magd* in the first half of the sixteenth century and the second half of the seventeenth century shows several patterns over time. First, in the second time period there is an increase of terms associated with professional occupation, especially with professions that are societally low-ranked, e.g. *Gesinde* 'menial servant' or 'menial staff'. This increase corresponds to the process of functionalization described above (Nübling 2011). Second, the nearest neighbors in the seventeenth century include three words associated with sexuality that are negatively evaluated, e.g. *Hure* 'prostitute'. This change instantiates the process of sexualization. Both processes can be seen as indicating a semantic shift of pejoration. On the other hand, the change observed in the nearest neighbors of *Knecht* over both time periods demonstrates an increase in work-related terms. That is, the patterns of change are different for female- and male-denoting personal nouns.

These findings show that nearest neighbors provide insight into gender-specific usage contexts for female-denoting and male-denoting nouns. With respect to semantic change, the neighbors provide evidence for the presence of evaluative meaning components of words at different time points, allowing us to study pejoration (and amelioration) over time. For example, the neighbors show that over time *Magd* comes to be associated with evaluative terms in the domain of sexuality, i.e., it acquires a derogatory meaning over time.

Importantly, we find that gender bias and the tendency for pejoration affects both nouns denoting professions and what we call"power nouns", i.e. nouns for nobility titles. While one might think that pairs of terms are equivalent with the exception of gender (*Graf* vs. *Gräfin* 'count/countess', *Magd* vs. *Knecht* 'female servant/male servant'), their descriptive meaning (Graf [+fem]) does not tell the whole story. Our findings reveal stark gender associations in language use that dif-

ferentiate the female from the male counterpart, like sexuality or marital status for the female-denoting term. This is an important component of the meaning of the word that our analysis of nearest neighbors casts light on. Word embeddings thus provide a method to gain insight into the "conceptual baggage words acquire in their discursive world travels" (McConnell-Ginet 2008).

An additional source of information about meaning components over time is the presence of proper names among the nearest neighbors of our target words. Traditionally, the philosophical literature on proper names has focused on their referential character and debates whether one should posit the existence of any denotational meaning for these expressions (see overviews by Cumming (2023); Nyström (2016)). A different view considers names to be meaningful (Jespersen 1924) and highlights the information they can convey about the possible categories of their referents or emotional associations (Nyström 2016). Among the latter approach, research has highlighted cultural and societal associations of proper names that follow from the histories and properties of their human referents. We propose that the association of names with specific individuals and, consequently, with the properties and/or stereotypes associated with such individuals or the social classes they belong to, should be considered meaning.[4] For example, there are associations of proper names stemming from their use in significant texts (e.g., the Bible) or as names of powerful individuals like royalty (e.g. Maximilian as a king's name), see Debus (2012). In our findings, the proper names that appear among nearest neighbors provide a rich source of information regarding the social networks of individuals and their social standing: for instance, among the neighbors of *Magd* in the seventeenth century, there are two biblical names of servants. An in-depth analysis of this topic is left for further work.

5.1. METHODOLOGICAL ISSUES IN THE STUDY OF LEXICAL SEMANTIC CHANGE. Studies of lexical meaning have identified shifts in conventional content that involve acquiring an emotive or evaluative dimension. Such content does not contribute to the truth conditions of a sentence; rather, it affects the appropriateness or felicity of the sentence in a particular context. Many terms are found in the literature for this type of meaning, including *connotation, emotive* (Jakobson 1960), *non-descriptive, expressive* (Cruse 1986), *expressive* (Potts 2007), *use-conditional* (Gutzmann 2013). The study of (near-)synonymy provides multiple examples of word pairs with the same descriptive meaning but different expressive content, e.g. *father* and *daddy*, or *neonate* and *baby* (Cruse 1986).

From the point of view of historical linguistics, semantic change affecting lexical items often consists of adding an expressive meaning component, as in amelioration and pejoration. However, as noted above, multiple factors may be involved here: a term may gain an evaluative component and/or an association with a profession that is positively or negatively evaluated. Hence, identifying expressive meaning requires examining word usage. This can be particularly challenging in the case of language change. Analyzing the linguistic environments of a word is essential to this endeavor. Crucially, word embeddings, as an unsupervised method, provide a helpful tool in investigating such semantic shifts. The change in the neighbors of *Magd* instantiates such addition of an expressive (in this case, negative) meaning component through the association with

[4] The notion of meaning we are adopting here is similar to the notion of "evoked" content (Cruse 1986), although Cruse does not mention proper names in this respect. According to Cruse, evoked content stems from three interacting dimensions: (i) field (defined as the field of discourse, as in legal discourse, political speeches, etc.), (ii) mode (pertaining to how a message is conveyed, e.g. written vs. spoken language), and (iii) style (subsuming language characteristics that stem from areas of experience of the participants in a linguistic exchange, including power relations, experience death, sex, religion, money).

the societally negatively evaluated domain of menial and service work, which was added to the descriptive meaning component 'woman'.

6. Conclusion. Our findings with German historical data corroborate the tendency for pejoration described in historical linguistics works regarding the semantic change undergone by female-denoting nouns. Distributional methods are helpful tools to explore pejoration and amelioration, and in general, for the study of how words acquire expressive meaning. Both nearest neighbors and analogy tests (that are "side products" of the models) provide a way to trace changes in meaning components of lexical items. By providing insight into actual language usage as represented in diachronic corpora, word embeddings allow us to investigate multiple dimensions of meaning in an unsupervised way.

In addition, we have shown that word embedding models can provide empirical evidence for gender bias over different time periods. This work complements the existing research on contemporary data and provides some historical depth on this phenomenon. Future work should deepen the understanding of historical gender biases by widening the scope of investigated lexemes, including e.g. occupational titles (*Sekretär* 'secretary, undersecretary, state secretary' vs. *Sekretärin* 'secretary, office assistant'), as well as conducting analyses for other languages.

References

Amaral, Patrícia, Hai Hu & Sandra Kübler. 2023. Tracing semantic change with distributional methods: The contexts of algo. *Diachronica* 40(2). 153–194.

Blankenberger, Stefan. 2003. Das Bild von Mann und Frau in historischen Wörterbchern des 15.-19. Jahrhunderts. Mainz (Magisterarbeit).

Caliskan, Aylin, Pimparkar Parth Ajay, Tessa Charlesworth, Robert Wolfe & Mahzarin R. Banaji. 2022. Gender bias in word embeddings: A comprehensive analysis of frequency, syntax, and semantics. In *Proceedings of the 2022 aaai/acm conference on ai, ethics, and society* AIES 22, 156170. ACM. https://doi.org/10.1145/3514094.3534162. http://dx.doi.org/10.1145/3514094.3534162.

Caliskan, Aylin, Joanna J. Bryson & Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334). 183186. https://doi.org/10.1126/science.aal4230. http://dx.doi.org/10.1126/science.aal4230.

Campbell, L. 2013. *Historical linguistics*. MIT Press.

Cruse, D.A. 1986. *Lexical semantics* Cambridge Textbooks in Linguistics. Cambridge University Press.

Cumming, Sam. 2023. Names. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University Winter 2023 edn.

Debus, Friedhelm. 2012. *Namenkunde und namengeschichte : eine einfhrung*. Berlin: Erich Schmidt Verlag.

Elmiger, Daniel. 2018. French anthroponyms as a heterogeneous category: Is there such a thing as personal nouns? *International Journal of Language and Culture* 5(2). 184–202. https://doi.org/10.1075/ijolc.00006.elm. http://www.jbe-platform.com/content/journals/10.1075/ijolc.00006.elm.

Firth, John R. 1957. *Papers in Linguistics*. London: Oxford University Press.

Geeraerts, Dirk, Dirk Speelman, Kris Heylen, Mariana Montes, Stefano De Pascale, Karlien Franco & Michael Lang. 2024. *Lexical Variation and Change: A Distributional Semantic Approach*. Oxford University Press. https://doi.org/10.1093/oso/9780198890676.001.0001. https://academic.oup.com/book/55109.

Gutzmann, Daniel. 2013. Beyond expressives: Explorations in use-conditional meaning. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 1–58. Boston: Brill.

Hamilton, William L., Jure Leskovec & Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th annual meeting of the association for computational linguistics*, 1489–1501. Berlin, Germany. https://doi.org/10.18653/v1/P16-1141. https://www.aclweb.org/anthology/P16-1141.

Harris, Zellig. 1954. Distributional structure. *Word* 10(2-3). 146–162.

Hock, H.H. & B.D. Joseph. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Mouton de Gruyter. https://books.google.com/books?id=oGH-RCW1fzsC.

Hu, Hai, Patrícia Amaral & Sandra Kübler. 2022. Word embeddings and semantic shifts in historical spanish: Methodological considerations. *Digital Scholarship in the Humanities* 37(2). 441–461.

Jakobson, Roman. 1960. Linguistics and poetics. In Thomas Sebeok (ed.), *Style in language*, 350–377. Cambridge: Massachusetts Institute of Technology Press.

Jespersen, Otto. 1924. *The philosophy of grammar*. London: H. Holt and Company.

Jurafsky, Daniel & James H. Martin. 2025. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models. https://web.stanford.edu/~jurafsky/slp3/.

Kaiser, Jens, Sinan Kurtyigit, Serge Kotchourko & Dominik Schlechtweg. 2021. Effects of Pre- and Post-Processing on type-based Embeddings in Lexical Semantic Change Detection. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics* 125–137.

Keller, Rudi. 1994. *On language change: the invisible hand in language*. London ; New York: Routledge.

Kim, Yoon, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde & Slav Petrov. 2014. Temporal Analysis of Language through Neural Language Models 61–65. https://doi.org/10.48550/ARXIV.1405.3515. https://arxiv.org/abs/1405.3515. Publisher: arXiv Version Number: 1.

Lenci, Alessandro. 2018. Distributional models of word meaning. *Annual Review of Linguistics* 4. 151–171.

McConnell-Ginet, Sally. 2008. Words in the World: How and Why Meanings Can Matter. *Language* 84(3). 497–527.

Meillet, Antoine. 1926. Comment les mots changent de sens. In *Linguistique historique et linguistique générale*, 230–271. Paris: Honoré Champion.

Mikolov, Thomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *Proceedings of international conference on learning representations (iclr)*, .

Nübling, Damaris. 2011. Von der Jungfrau zur Magd, vom Mädchen zur Prostituierten: Die Pejorisierung der Frauenbezeichnungen als Zerrspiegel der Kultur und als Effekt mnnlicher Galanterie? *Jahrbuch für Germanistische Sprachgeschichte* 2. 344–359. https://doi.org/10.1515/9783110236620.344. https://www.degruyter.com/view/j/jbgsg.2011.2.issue-1/9783110236620.344/9783110236620.344.xml.

Nyström, Staffan. 2016. Names and meaning. In *The oxford handbook of names and naming*, chap. 39. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199656431.013.26. https://doi.org/10.1093/oxfordhb/9780199656431.013.26.

Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(3). 165–197.

Schlechtweg, Dominik, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky & Nina Tahmasebi. 2020. SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 1–23. Barcelona (online): International Committee for Computational Linguistics. https://doi.org/10.18653/v1/2020.semeval-1.1. https://aclanthology.org/2020.semeval-1.1.

Tahmasebi, Nina, Lars Borin, Adam Jatowt, Yang Xu & Simon Hengchen. 2021. *Computational approaches to semantic change*. Zenodo. https://doi.org/10.5281/ZENODO.5040241. https://zenodo.org/record/5040241.

Tian, Zuoyu, Dylan Jarrett, Juan Manuel Escalona Torres & Patrícia Amaral. 2021. BAHP: Benchmark of Assessing Word Embeddings in Historical Portuguese. In *Proceedings of LaTeCH-CLfL*, 113–119.

Ullmann, S. 1979. *Semantics: An introduction to the science of meaning*. Basil Blackwell.

Warnke, Ingo. 1993. Zur Belegung von 'Frau' und 'Weib' in historischen deutschen Wörterbüchern des 16. und 17. Jahrhunderts. In Britta Hufeisen (ed.), *"Das Weib soll schweigen ..." (1.Kor.14,34): Beiträge zur linguistischen Frauenforschung* (Kasseler Arbeiten zur Sprache und Literatur Bd. 19), 127–152. Frankfurt am Main ; New York: P. Lang.