In German, all professors are male

Dominic Schmitz

Abstract Findings of previous behavioural studies suggest that the semantic nature of what is called the "masculine generic" in Modern Standard German is not generic but biased towards a masculine reading. Such findings run counter the traditional assumption of masculine generic forms to be gender-neutral and are the cause of debates within and outside the linguistic community. The present paper aims to explore the semantics of masculine generics by implementing ideas of discriminative learning; an approach that thus far has not yet been used in this matter. If the present results account for a male bias of the masculine generic, findings by previous studies, which relied on very different methodological approaches, are confirmed. Indeed, it is found that masculine generics are semantically much more similar to masculine explicits than to feminine explicits. This supports the notion of a male bias in masculine generics.

Keywords naive discriminative learning, semantic vectors, language bias, gender

1. Introduction

Modern Standard German knows three grammatical genders: the feminine, the masculine, and the neuter. In contexts in which the sex and/or gender of the referent is a) unknown, b) not of importance, or c) mixed (i.e. there are referents of different sexes/genders), speakers of German regularly make use of the so-called *generisches Maskulinum* 'masculine generic'. The generic nature of this form refers to the notion of it being

gender-neutral, independently of its grammatical gender (Doleschal 2002).[1] Masculine generics are used in the singular and plural as illustrated by (1) and (2), respectively. In both examples, referents can be of any sex or gender.

(1) Wird heute ein
be.PRS.3SG today.ADV DET.INDF.M.SG

Professor an eine
professor.M.NOM.SG to DET.INDF.F.SG

Universität berufen, kommt
university.F.ACC.SG appoint.PTCP.PRS come.PRS.3SG

dieser oft mit einem
DET.DEF.M.SG often with DET.INDF.M.SG

ganzen Forschungsteam.
whole.ADJ research group.N.DAT.SG

(2) Die Professor-en der
DET.DEF.PL professor-M.NOM.PL DET.DEF.PL

[1] In this paper, I will use the term "gender-neutral" as an umbrella term concerning both sex, i.e. the simplified biological perspective, and gender, i.e. the social and cultural perspective. I acknowledge that both terms - sex and gender - are neither identical and nor are forms of sex and gender clearly correlated or matched up. For the present case, however, it is of negligible importance whether we specifically refer to sex or gender.

regulären	Schweizer	Uni-s.
regular.ADJ	Swiss.ADJ	uni-F.GEN.PL

Importantly, explicit masculine forms, that is forms which unambiguously refer to masculine referents, are identical in form. This is shown in (3) and (4).

(3)

Michael	Rosenberger	ist	
Michael	Rosenberger	be.PRS.3SG	
Professor		für	Moraltheologie.
professor.M.ACC.SG		for	moral theology.F.ACC.SG

(4)

Hans-Peter	und	Volker Stenzl	[…]	als
Hans-Peter	and	Volker Stenzl		as
Professor-en.				
professor-M.NOM.PL				

To create a corresponding feminine explicit form, an additional suffix is required. Accordingly, the explicitly feminine counterpart of *Professor* is *Professorin*, where the *-in* constitutes the feminine gender suffix. Similarly, the plural counterpart of *Professoren* is *Professorinnen*. Here, the feminine suffix is added in-between the masculine form *Professor* and the plural suffix *-en* with a reduplication of *n* as to indicate vowel quality.

As has been illustrated, in German both generic and explicit masculine forms of role nouns share the same form. The feminine counterparts, however, are different in their segmental makeup as an additional inflectional suffix is

required. With two allegedly semantically distinct members of such a role noun paradigm sharing their form, one question naturally suggests itself: How semantically different are such members?

This question is explored in the remainder of this paper. In the following section, I give a brief overview of existing research on the matter. In Section 3, the method of the present paper is detailed and in Section 4, the statistical analysis is outlined. In Section 5, I present the results of the analysis. Finally, Section 6 provides a discussion of the present results.

2. Background

The nature of the masculine generic has been a topic of linguistic research at least since the late 1990s. Braun et al. (1998), having participants guess the percentage of female attendants to ecotrophology and geophysics conferences after reading a text on either conference, found that participants provided lower percentages if the pertinent text used masculine generics. Rothermund (1998) gave brief descriptions of situations to participants. If the descriptions contained masculine generics, participants took longer to decide that a stereotypically male distractor shown afterwards was not part of the original description. Heise (2000) found that when naming protagonists of a story, participants more often used male names when confronted by masculine generics in the story itself. Stahlberg & Sczesny (2001) and Stahlberg et al. (2001) asked participants to provide names of their favourite painters, athletes, and musicians, among other categories. When the pertinent prompts were given in the masculine generic, participants more often provided male representatives

as answers. In a cloze task, Rothmund & Scheele (2004) found that clozes are more often resolved with male referents if the given context contained masculine generics. Gygax et al. (2008) asked participants whether a sentence was a meaningful continuation of a first sentence containing a masculine generic. The authors found that the proportion of positive judgements was higher for male continuations. While this is but a brief cursory overview of the existing research on the nature of masculine generics in German, one can clearly conclude that previous studies on the matter point towards the same finding: The masculine generic seems to be not gender-neutral but biased towards male referents.

However, the aforementioned studies do not come without issues, of which I mention the two most crucial ones here. First, about half of all studies investigating the nature of the masculine generic make use of students as participants (e.g. Heise 2000; Stahlberg & Sczesny 2001; Gygax et al. 2008). Students are particularly prone to progressive change (Bailey & Williams 2016). Thus, including only students as participants might influence results to an unknown extent. Second, with a few exceptions, studies tend to ignore world-knowledge which might manifest itself in stereotypes, which in turn might influence the nature of pertinent masculine generic forms.

The present paper attempts to tackle the aforementioned issues. To factor in stereotypicality of role nouns, stereotypicality ratings as elicited by Gabriel et al. (2008) are incorporated into the analysis. Taking into account the potential influence of participant groups, the present investigation does not make use of participants at all. Thus, it is independent of social groups and their pertinent characteristics. This is achieved by implementing a naive

discriminative learning network (NDL; e.g. Baayen & Ramscar 2015). The following section introduces the method, that is the choice of target items and the NDL implementation, in more detail.

3. Method

3.1. Target items

The set of 120 target words for the present study was adopted from Gabriel et al. (2008). In their study, the authors elicited stereotypicality ratings for role nouns. Thus, Gabriel et al.'s set of items and their ratings represent the perfect selection of target words for the present paper: If all role nouns, independently of their stereotypical associations, show similarities in terms of their underlying semantics, this makes potential findings more robust.

To investigate the semantic similarities between masculine generics, and masculine and feminine explicits, paradigms for all target words were taken into account. This is what I call "target word paradigms". Each such paradigm consists of a target word's singular and plural generic and explicit forms, as is illustrated in Table 1.

Table 1. Target word paradigm of the target word *Professor* 'professor'.

	masculine generic	masculine explicit	feminine explicit
singular	*Professor*	*Professor*	*Professorin*
plural	*Professoren*	*Professoren*	*Professorinnen*

Each cell of Table 1 represents what is called a TYPE. The six TYPEs are: singular masculine generic, singular masculine explicit, singular feminine explicit, plural masculine generic, plural masculine explicit, and plural feminine explicit.

3.2 From sentences to semantic vectors

To compute semantic vectors with NDL, one needs cues and outcomes to begin with. For the present implementation, it was decided to use bases of content words, function words, and inflectional functions (e.g. number, case) as cues and outcomes. Such an implementation is similar to previous ones (cf. Baayen et al. 2019). As a toy example, take the following sentence: *Worms live in apples*. Parsing this sentence into its cues/outcomes, one obtains: WORM PLURAL LIVE PRESENT IN APPLE PLURAL. Importantly, the genericity of a target was considered an inflectional function on its own. Thus, *Professor* in (1) is parsed as PROFESSOR SINGULAR MASCULINE GENERIC, while *Professor* in (3) is parsed as PROFESSOR SINGULAR MASCULINE EXPLICIT. I will return to the notions of cues and outcomes later in this section.

To obtain a sufficient number of cues and outcomes, a text corpus in which target words are disambiguated for their generic or explicit meaning was created. This corpus is based on sentences from the Leipzig Corpora Collection's (Goldhahn, Eckart & Quasthoff 2012) "news" sub-corpus. Sampling one million sentences for each year from 2010 to 2019, a total of ten million sentences were extracted to create an initial collection of sentences. Considering a range of ten years, it was avoided that one specific topic was overrepresented, rendering the corpus potentially unrepresentative of general language use.

From this collection, sentences containing target word paradigm members were sampled. During this process, issues with several target items became apparent, which led to the

exclusion of seven target word paradigms.[2] The number of sentences per target word paradigm was proportional to the overall frequency of the target word paradigm within the initial collection of sentences (see the Supplementary Materials for further information). This resulted in the sampling of 30,000 sentences containing target words.

Additionally, 800,000 sentences without target words were sampled from the initial collection. While the target word paradigms were the focus of the present investigation, a semantic space such as created by NDL is meant to represent the semantic knowledge of a speaker. Thus, language material beyond the target word paradigms is required. While working with a larger corpus is generally preferable in this regard, a huge number of sentences comes with extensive computation times during the implementation of NDL. To keep the carbon footprint of this study reasonably low, I aimed at an overall number of sentences close to similar implementations (e.g. Baayen et al. 2019).

The text corpus was then annotated in two ways. First, all sentences were annotated automatically using the RNNTagger software (Schmid 1999). Using the RNNTagger, inflectional features such as case, number, and tense were annotated. As the present paper is not concerned with derivational processes, no annotation based on derivation was conducted. Second, the 30,000 sentences containing target words were manually annotated for genericity by the author and three assistants, as to the author's knowledge, there is no automatic annotation software available to carry out this task.

[2] For example, several items included in Gabriel et al. (2008) did not represent masculine generics but gender-neutral forms (e.g. *Hilfskraft* 'aide').

All annotators were native speakers of German with an educational level comparable to A levels or higher. Taking into account the context of each target word, the following three features were annotated: gender (masculine vs. feminine); number (singular vs. plural); and genericity (explicit vs. generic). If for a sentence it was unclear whether a target word was intended in an explicit or generic manner, this sentence was discarded and a new sentence for the pertinent target word was sampled if available. Lastly, for the sentences containing target words, the automatic and the manual annotations were brought together. A list of all target items is given in the Supplementary Materials.

Finally, an NDL implementation was trained using the Python plugin pyndl (Sering et al. 2017). For each sentence of the corpus, each individual base, function word, and inflectional function within the sentence (outcomes) was predicted by the other bases, function words, and inflectional functions (cues) of the same sentence. This approach relies on the so-called Rescorla-Wagner rules (Wagner & Rescorla 1972; Rescorla & Wagner 1972) and is hence grounded in theory on cognitive mechanisms that has been shown to successfully model important effects observed in learning. Crucially, cues and outcomes may be absent or present. Thus, cues may or may not occur with certain outcomes. The association weight of an outcome to a cue increases every time the pertinent outcome and cue co-occur, while it decreases every time the pertinent outcome occurs, but the cue does not. Hence, each encounter of an outcome with cues leads to a recalibration of all association weights of the relevant outcome. At the end of this process, then, a stable state is reached. Coming back to the aforementioned toy example, *worms live in apples*, this then

means that for each outcome (WORM PLURAL LIVE PRESENT IN APPLE) each cue (WORM PLURAL LIVE PRESENT IN APPLE) is predictive within the toy example sentence. Thus, NDL increases their association weights, while it decreases the association weights of the given outcomes with all other cues of the corpus during the processing of this particular sentence. This procedure is repeated for each of the 830,000 sentences for all of their 49,044,960 cue/outcome tokens. Taking the association weights of a pertinent outcome and all cues once the stable state at the end of this process is reached, one obtains the outcome's semantic vector. Computing vectors not only for words but also for inflectional functions allows for the analysis of explicit versus generic forms, and follows the reasoning of Baayen et al. (2019). The resulting semantic vectors of 7,510 dimensions (after reduction by approx. 50 %, cf. Baayen et al. 2019) are analysed in the following section.

4. Analysis

To analyse the semantics of masculine generic, masculine explicit, and feminine explicit role nouns, one must first assemble their semantics by means of their individual semantic vectors. Constructing the semantics of a complex word by the addition of the vectors of its pertinent parts follows the general idea of linear discriminative learning (cf. Baayen et al. 2019). Take, as an example, the paradigm of *Professor*. For all of its members, the base meaning of "being a professor" is required. This semantics is entailed in the vector of $\overrightarrow{Professor}$, i.e. what was called "base" before. To create the singular masculine generic form, one must add $\overrightarrow{singular}$, $\overrightarrow{masculine}$, and $\overrightarrow{generic}$. Hence, the semantics are

$$\overrightarrow{Professor_{s,m,g}} = \overrightarrow{Professor} + \overrightarrow{singular} + \overrightarrow{masculine} + \overrightarrow{generic}. \quad (1)$$

To create the singular explicit masculine form, one must add $\overrightarrow{explicit}$ instead of $\overrightarrow{generic}$. To create the singular explicit feminine form, one must additionally add $\overrightarrow{feminine}$ instead of $\overrightarrow{masculine}$. For plural forms, the same reasoning applies. This was done for all members of the 113 target word paradigms.

The resulting semantic vectors were then compared by their cosine similarity. That is, each TYPE was compared to each other TYPE of the same number. The resulting cosine similarities between the TYPEs are given in Table 2.

Table 2. TYPE combinations for which cosine similarities were computed.

Singular			Plural		
masculine generic	vs.	masculine explicit	masculine generic	vs.	masculine explicit
masculine generic	vs.	feminine explicit	masculine generic	vs.	feminine explicit
masculine explicit	vs.	feminine explicit	masculine explicit	vs.	feminine explicit

Using the gdsm package (Schmitz & Schneider 2022), cosine similarities were computed across all target words. For example, the cosine similarity between all singular masculine generic and all masculine explicit target words was computed. The analysis of cosine similarities within individual target word paradigms is beyond the scope of the present investigation. In the present case, cosine similarity can take values within the interval of [0,1]. Higher values indicate a higher similarity of two vectors, while lower values indicate a lower similarity. As vectors reflect words' semantics, a higher

similarity of vectors corresponds to a higher semantic similarity of word forms.

To check for an influence of stereotypicality on the semantic similarity of pertinent types, cosine similarities entered a beta regression analysis as dependent variable. Stereotypicality ratings were introduced as the sole predictor. Thus, for example, the cosine similarity values between singular masculine generic and masculine explicit target word forms were predicted in such a model by the stereotypicality ratings of the pertinent target words. Beta regression models were fitted using the mgcv package (Wood 2017). The analyses as well as the pertinent data and vectors are part of the Supplementary Material.

5. Results

The mean cosine similarities between types across all target word paradigms are given in Table 3 and are illustrated in Figure 1.

Wilcoxon-Mann-Whitney-Tests were used to check whether the cosine similarity between two pertinent TYPEs was significantly different than the cosine similarity between to other TYPEs. For example, it was compared whether the cosine similarity between singular masculine generics and explicits was significantly different than the cosine similarity between singular masculine generics and feminine explicits. The results are very clear: All resulting p-values are lower than $2.2e^{-16}$.

Taking into account the mean cosine similarities given in Table 3, one can conclude the following: For both singular and plural, masculine generics and explicits are semantically most similar; they are significantly more similar than

masculine generics and feminine explicits and significantly more similar than masculine and feminine explicits.

Table 3. Mean cosine similarity of type combinations for which cosine similarities were computed.

Singular			Mean
masculine generic	vs.	masculine explicit	0.9966
masculine generic	vs.	feminine explicit	0.9343
masculine explicit	vs.	feminine explicit	0.9388
Plural			Mean
masculine generic	vs.	masculine explicit	0.9911
masculine generic	vs.	feminine explicit	0.8221
masculine explicit	vs.	feminine explicit	0.8351

Importantly, the aforementioned significant differences are true for all TYPE comparisons. The implemented beta regression models showed that no matter which cosine similarity values entered as dependent variable, stereotypicality ratings did not reach significance as predictor ($p > 0.8$ for stereotypicality as predictor in all models). That is, no matter which TYPE combination the cosine similarity values belonged to, stereotypicality of target words did not show any effect. Thus, the cosine similarities as given in Table 3 are apparently independent of stereotypicality.

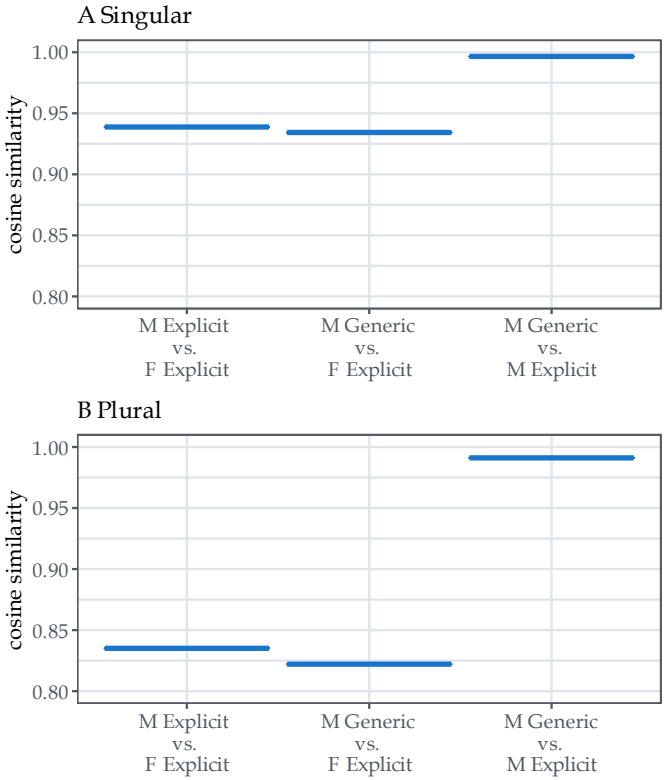

Figure 1. Differences in cosine similarity between two pertinent types for the singular (panel A) and the plural (panel B). Abbreviations: M = masculine; F = feminine.

6. Discussion

This paper set out to explore the question of how semantically different members of identical form in a role noun paradigm in German are. Using semantic vectors computed by an NDL implementation, the issue of potential influences of participant bias was avoided. Taking into account the stereotypicality of

pertinent target items, information outside the language system itself was factored in. The present analyses found results which are in line with findings of previous studies on the nature of the masculine generic. That is, the masculine generic is semantically most and highly similar to the masculine explicit. The feminine explicit, on the other hand, is least similar to the generic masculine. This finding can potentially account for the male bias of the masculine generic reported on in previous research. It thus appears that the masculine generic is not generic after all but semantically highly similar to the masculine explicit. Hence, in German all professors are male, unless the female inflected word form is used. Stereotypicality did not show an effect on the semantic similarity of different paradigm member types.

However, the present implementation also comes with limitations. The semantic vectors are based on sentences from news websites. Semantic vectors based on different genres might yield different results. Also, a set of 113 target word paradigms only scratches the surface of the overall number of role nouns in German. Additionally, an implementation of linear discriminative learning (cf. Baayen et al. 2019; Schmitz et al. 2021) might provide more insight into the underlying nature of role nouns and masculine generics in particular than a rather simple analysis of cosine similarities. Such matters are subject for future research.

Supplementary Material

Find the Supplementary Material as well as the R script containing the presented analyses and the required data and vector sets here: https://osf.io/wf43s/

Acknowledgements

The author is grateful to Janina Esser, Dennis Dahlhausen, and Laureen Schmitz for assisting in the manual annotation of sentences as well as to Viktoria Schneider and Janina Esser for valuable theoretical input, and to the audience of the 28th Germanic Linguistics Annual Conference for helpful feedback.

References

Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan & James P. Blevins. 2019. 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. 1–39. https://doi.org/10.1155/2019/4895891.

Baayen, R. Harald & Michael Ramscar. 2015. Abstraction, storage and naive discriminative learning. *Handbook of Cognitive Linguistics* 39. 100–120. https://doi.org/10.1515/9783110292022-006.

Bailey, Mandi & Lee R. Williams. 2016. Are college students really liberal? An exploration of student political ideology and attitudes toward policies impacting minorities. *The Social Science Journal*. No longer published by Elsevier 53(3). 309–317. https://doi.org/10.1016/J.SOSCIJ.2016.04.002.

Braun, Friederike, Anja Gottburgsen, Sabine Sczesny & Dagmar Stahlberg. 1998. Können Geophysiker Frauen sein? Generische Personenbezeichnungen im Deutschen. *Zeitschrift fur Germanistische Linguistik* 26(3). 265–283. https://doi.org/10.1515/zfgl.1998.26.3.265.

Doleschal, Ursula. 2002. Das generische Maskulinum im Deutschen. Ein historischer Spaziergang durch die deutsche Grammatikschreibung von der Renaissance bis zur Postmoderne. *Linguistik Online*. University of Bern 11(2). https://doi.org/10.13092/lo.11.915.

Gabriel, Ute, Pascal Gygax, Oriane Sarrasin, Alan Garnham & Jane Oakhill. 2008. Au pairs are rarely male: Norms on the gender

perception of role names across English, French, and German. *Behavior Research Methods* 40(1). 206–212. https://doi.org/10.3758/BRM.40.1.206.

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

Gygax, Pascal, Ute Gabriel, Oriane Sarrasin, Jane Oakhill & Alan Garnham. 2008. Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes* 23(3). 464–485. https://doi.org/10.1080/01690960701702035.

Heise, Elke. 2000. Sind Frauen mitgemeint? Eine empirische Untersuchung zum Verständnis des generischen Maskulinums und seiner Alternativen. *Sprache & Kognition* 19(1/2). 3–13. https://doi.org/10.1024//0253-4533.19.12.3.

Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcementy. In A. H. Black & W. F. Prokasy (eds.), *Classical conditioning II: Current research and theory*, 64–99. Appleton-Century-Crofts.

Rothermund, K. 1998. Automatische geschlechtsspezifische Assoziationen beim Lesen von Texten mit geschlechtseindeutigen und generisch maskulinen Text-Subjekten. *Sprache & Kognition* 17. 183–198.

Rothmund, Jutta & Brigitte Scheele. 2004. Personenbezeichnungsmodelle auf dem Prüfstand. *Zeitschrift für Psychologie / Journal of Psychology* 212(1). 40–54. https://doi.org/10.1026/0044-3409.212.1.40.

Schmid, H. 1999. Improvements in part-of-speech tagging with an application to German. In S. Armstrong, K. Church, P. Isabelle, S. Manzi, E. Tzoukermann & D. Yarowsky (eds.), *Natural language processing using very large corpora*, 13–25. Springer. https://doi.org/10.1007/978-94-017-2390-9_2.

Schmitz, Dominic, Ingo Plag, Dinah Baer-Henney & Simon David Stein. 2021. Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology*. Frontiers Media S.A. 12. https://doi.org/10.3389/fpsyg.2021.680889.

Schmitz, Dominic & Viktoria Schneider. 2022. gdsm: General functions for Distributional SeMantics. https://github.com/dosc91/gdsm.

Sering, Konstantin, Marc Weitz, David-Elias Künstle & Lennart Schneider. 2017. Pyndl: Naive discriminative learning in python. https://doi.org/http://doi.org/10.5281/zenodo.597964.

Stahlberg, Dagmar & Sabine Sczesny. 2001. Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen. *Psychologische Rundschau* 52(3). 131–140. https://doi.org/10.1026//0033-3042.52.3.131.

Stahlberg, Dagmar, Sabine Sczesny & Friederike Braun. 2001. Name Your Favorite Musician. *Journal of Language and Social Psychology*. Sage PublicationsSage CA: Thousand Oaks, CA 20(4). 464–469. https://doi.org/10.1177/0261927X01020004004.

Wagner, Allan R. & Robert A. Rescorla. 1972. Inhibition in pavlovian conditioning: Application of a theory. In R. A. Boakes & M. S. Halliday (eds.), *Inhibition and learning*, 301–334. Academic Press Inc.

Wood, Simon N. 2017. *Generalized additive models: An introduction with R*. CRC Press. https://doi.org/10.1201/9781315370279.