

# Zur Wahrnehmung des generischen Maskulinums in Erstgliedern von Komposita und maskuliner Epizöna

**Zusammenfassung** Der Großteil der Nomen im Deutschen, die dazu verwendet werden können, um auf Personen oder Personengruppen zu referieren, haben eine maskuline (z. B. *Lehrer*) und eine movierte (also eine durch Derivation erzeugte) feminine Form (z. B. *Lehrerin*). Die grammatisch maskulinen Formen dieser Nomen sind im Deutschen einer weit verbreiteten Auffassung zufolge ambig zwischen einer rein auf männliche Referenten ausgerichteten und einer geschlechtsneutralen Lesart (das sogenannte „generische Maskulinum“). Dieser Artikel geht der Frage nach, ob diese Generalisierung auch für Erstglieder von Komposita (z. B. *Lehrer-* in *Lehrerzimmer*) und nicht-movierbare maskuline Nomen, die keine feminine Form kennen, gilt (sogenannte Epizöna wie z. B. *Gast*). Die diskutierten empirischen Daten deuten darauf hin, dass sowohl Maskulina movierbarer Nomen, die als Erstglieder von Komposita verwendet werden, als auch nicht-movierbare maskuline Nomen eher sexusindifferent interpretiert werden.

**Abstract** Most German nouns that can be used to refer to people or groups of people have a masculine (e. g. *Lehrer* ‘male teacher’) and a derived feminine (e. g., *Lehrerin* ‘female teacher’) form. The grammatically masculine nouns are typically viewed as being ambiguous between a reading which can only be used to refer to men and a reading which can be used to refer to mixed-gender groups (the so-called “generic masculine”). The main question of the present article is whether this generalization also holds for the first elements of nominal compounds (e.g., *Lehrer-* in *Lehrerzimmer* ‘teachers’ room’) and for inherently masculine nouns from which no feminine form can be derived (e.g., *Gast* which is inherently masculine). The empirical data discussed suggest that generic masculine forms used as the first elements of N-N compounds as well as inherently masculine nouns are interpreted rather generically, i. e., not as referring to male individuals only.

**Keywords:** Generisches Maskulinum, Genus, Sexus, gendersensible Sprache, Epizöna, Komposita, Experiment

# 1 Einleitung

Maskuline Personenbezeichnungen im Deutschen weisen eine Ambiguität auf, da sie entweder dazu verwendet werden können, ausschließlich männliche Personen zu bezeichnen oder als generische Formen interpretiert werden bzw. generisch intendiert sind. So kann die Form *Lehrer* entweder eine Gruppe männlicher Lehrkräfte bezeichnen oder eine gemischt-geschlechtliche Gruppe von weiblichen und männlichen Lehrkräften. Für das generische Maskulinum wurde die Frage gestellt, ob es aufgrund der Formgleichheit der sexusspezifischen und der sexusindifferenten Lesart, einen „male bias“ gäbe, Menschen also bei der Verarbeitung von Personenbezeichnungen im generischen Maskulinum eher an männliche Repräsentanten als an eine sexusausgeglichene Gruppe denken. Ein Ausdruck wie *die Lehrer* könnte also dann zwar ambig sein zwischen einer Lesart, deren Denotat (oder Extension) nur männliche Lehrer umfasst, und einer Lesart, deren Denotat eine gemischtgeschlechtliche Gruppe umfasst. Dennoch könnte, selbst wenn aus dem Kontext heraus klar wird, dass eine (ausgewogene) gemischtgeschlechtliche Gruppe gemeint ist oder die Gruppenzusammensetzung schlicht unklar ist, der Gebrauch von *die Lehrer* dazu führen, dass bei der Verarbeitung dieser Form eine (vermutlich unbewusste) mentale Repräsentation evoziert wird, die mehrheitlich männliche Lehrer miteinschließt. Eine lange Reihe psycholinguistischer Experimente, die auf unterschiedlichen Methoden basieren, hat genau einen solchen „male bias“ nachweisen können (siehe z. B. Braun u. a. 1998; Gabriel u. a. 2008; Heise 2000; Stahlberg & Sczesny 2001; Stahlberg, Sczesny & Braun 2001; Gabriel & Mellenberger 2004; Gygax u. a. 2008; Hansen, Littwitz & Sczesny 2016 und die Überblicke in Stahlberg u. a. 2007; Gabriel, Gygax & Kuhn 2018). Getestet wurde beispielsweise, wie Probandinnen und Probanden reagieren, wenn sie eine Geschichte weiterentwickeln müssen, die entweder mit einem generisch intendierten Maskulinum oder mit einer gendersensiblen Form eingeleitet wurde (Heise 2000), wie Probandinnen und Probanden antworten, wenn sie eine Reihe von Menschen nennen sollen, auf die entweder im generisch intendierten Maskulinum Bezug genommen wird oder mit verschiedenen gendersensiblen Formen (Stahlberg, Sczesny & Braun 2001), wie das Geschlechterverhältnis verschiedener Gruppen eingeschätzt wird, die unterschiedlich bezeichnet werden (Braun u. a. 1998) oder wie schnell Sätze mit unterschiedlichen Formen und unterschiedlichen Fortsetzungen gelesen werden (Irmen & Roßberg 2004).

Derlei wissenschaftliche Befunde können in die Praxis rückwirken, da man mit ihnen begründen kann, dass es sinnvoll ist, Alternativen zum generischen Maskulinum wie Paarformen (*Lehrerinnen und Lehrer*) oder den Genderstern (*Lehrer\*innen*) zu verwenden. Es stellt sich jedoch die Frage, in welchen Fällen

dies geschehen sollte bzw. in welchen Fällen dies Sinn ergibt (für einen Überblick über verschiedene Strategien und Sichtweisen siehe [Kotthoff 2020](#) oder [Müller-Spitzer 2022](#)). Wenn man eine generische Aussage über eine Personengruppe macht (*Lehrerinnen und Lehrer arbeiten an Schulen*) ist dies intuitiv sinnvoll, wenn man dagegen Aussagen über spezifische Personen macht (*Gabi und Marianne sind Lehrerinnen*), ist dies intuitiv nicht sinnvoll. Aus diesem Grund wurde immer wieder darauf hingewiesen, dass die Verwendung bzw. Nichtverwendung geschlechtersensibler Sprache auch eine Frage der Referenz ist (siehe z. B. [Doleschal 1992](#), [Kotthoff & Nübling 2018](#): 91–129 oder [Zifonun 2018](#)): „Striktes Gendern ohne Berücksichtigung von referentiell Modus[...] ist wenig sinnvoll“ ([Zifonun 2018](#): 53). Die oben angeführten psycholinguistischen Studien wurden genau hierfür kritisiert, da sie von den Probandinnen und Probanden eine eindeutige Sexuzuordnung verlangten, da sich die in den oben genannten Experimenten verwendeten Personenbezeichnungen auf (wenn auch fiktive) konkrete Personen in konkreten Kontexten bezogen ([Zifonun 2018](#): 51).

Die Rolle der Referenz sollte folgerichtig auch in der Forschung zu gendersensibler Sprache<sup>1</sup> Berücksichtigung finden. In der vorliegenden Studie spielt Referenz insofern eine Rolle, als Personenbezeichnungen einmal als Erstglieder von Komposita untersucht wurden, die selbst nicht referieren ([Kotthoff & Nübling 2018](#): 94), und diese mit denselben Personenbezeichnungen in freier Verwendung verglichen wurden. Ein Beispiel für die hier untersuchten N-N-Komposita ist *Lehrerzimmer*, das, wenn es referentiell gebraucht wird, nur verwendet werden kann, um auf eine Art von Zimmer zu referieren, aber nicht auf Personen. Für solche nicht-referentiell verwendeten Wortbestandteile stellt sich die Frage, ob diese dieselbe Ambiguität wie ihre referentiell verwendeten Pendanten (hier: *Lehrer*) aufweisen. Genau diese Frage bildet den ersten Teil des hier vorgestellten Forschungsvorhabens. Dass diese Frage auch für die Praxis von Belang ist, zeigt auch die teils emotional geführte Debatte über diese Art von Wortbildung, für die immer wieder gefordert wurde, auch sie durch gendersensible Formen zu ersetzen, wie beispielsweise *Lehrer\*innenzimmer* oder, um ein extremeres Beispiel anzuführen, *Bürger\*innenmeister\*innen*. Solche Formen erscheinen vielen Menschen offenbar Unbehagen zu bereiten, was sich auch in teils emotionaler Presseberichterstattung zu diesem Thema widerspiegelt.<sup>2</sup>

<sup>1</sup> Wir sprechen hier von „gender-sensibler“ und nicht von „gender-neutraler“ oder „gender-gerechter“ Sprache, da das Ideal einer vollständig neutralen oder gerechten Sprache wohl in realiter nicht erreichbar ist.

<sup>2</sup> Siehe zum Beispiel: <https://www.merkur.de/welt/duden-diskussion-wenn-gender-sprache-zum-schwachsinn-wird-90195836.html>, <https://www.nordbayern.de/kultur/streitthema-gendern-sechs-erkenntnisse-einer-aufgeheizten-debatte-1.11186373> oder <https://www.heise.de>

Interessanterweise ist dieser offensichtliche Problemfall in der sprachwissenschaftlichen Literatur bisher nicht Gegenstand empirischer Studien gewesen (siehe auch die kurze Diskussion in Zifonun 2018: 52–53). Dennoch wird diese Konstruktion in der präskriptiven Literatur zum Thema erwähnt. So schlagen manche Ratgeber beispielweise gendersensible Erstglieder in Komposita vor. Das Gleichstellungsbüro der Universität Heidelberg empfiehlt etwa Komposita wie *Student\*innenwohnheim* oder *Mitarbeiter\*innengespräche* (Brendel & Speck 2021: 14). Auch der Praxisleitfaden zur gendersensiblen Sprache von Hornscheidt & Sammla (2021: 99) enthält Beispiele wie *Architekt:innen-Innung* oder *Mitarbeiter:innenversammlung*. Der Duden-Ratgeber von Diewald & Steinhauer (2017: 65–68) ist dagegen vorsichtiger. Dort wird angemerkt, dass ein Kompositum wie *Lehrerzimmer* in erster Linie einen Raum bezeichnet und keine Personenreferenz zustande kommt.

Es existieren verschiedene Vorschläge, deren Ziel es ist, den beschriebenen „male bias“ zu eliminieren bzw. wenigstens zu reduzieren. Diese Vorschläge basieren im Wesentlichen auf zwei Strategien. Bei Neutralisierungs-Strategien wird versucht, einen Geschlechtsbezug komplett zu vermeiden. Dies wird etwa mit nominalisierten Partizipien im Plural versucht, wie beispielsweise *die Lehrenden* (für eine kritische Betrachtung siehe Harnisch 2016). Bei Feminisierungs-Strategien wird versucht, Frauen sprachlich miteinzubeziehen und damit deren mentalen Einbezug zu stärken (Sczesny, Moser & Wood 2015; Sczesny, Formanowicz & Moser 2016). Dazu kann beispielsweise die Paarform verwendet werden, bei der sowohl Männer als auch Frauen benannt werden. Dazu gehört aber auch die Verwendung von Binnenmajuskeln (*LehrerInnen*) oder der schon angesprochene Gender-Stern (für eine aktuelle Evaluation des Gender-Sterns siehe Friedrich u. a. 2021). Feminisierungsstrategien funktionieren jedoch nur dann, wenn es möglich ist, eine feminine Form über ein Movierungssuffix abzuleiten. Im Deutschen existieren jedoch eine Reihe maskuliner, nicht-movierbarer Nomen wie *Gast* oder *Mensch*, sogenannte maskuline Epizöna (manchmal wird auch der Begriff „lexikalische Maskulina“ verwendet, siehe z. B. Kotthoff 2022). Weibliche Formen solcher Substantive (wie etwa *Gästin* oder *Menschin*) sind nicht existent oder zumindest relativ unüblich, obwohl die Online-Redaktion des Duden für solche Fälle kürzlich feminine Pendants vorgeschlagen hat (z. B.

---

de/tp/features/Genderstern-Co-eine-festgefahrene-Debatte-6195191.html, zuletzt aufgerufen am 27.04.2022.

die *Gästin*).<sup>3</sup> Auch solche Bildungen führen immer wieder zu emotionalen Diskussionen und zu einem entsprechenden Medienecho.<sup>4</sup>

Zwar wurde auch für maskuline Epizöna gezeigt, dass ihre Verarbeitung zu einem „male bias“ führt (Irmen & Roßberg 2004), es stellt sich jedoch die Frage, ob dieser „male bias“ tatsächlich darauf zurückgeht, dass ein Nomen wie *Gast* ambig zwischen einer sexusspezifischen und einer sexusindifferenten Lesart ist, oder ob dieser Verzerrungseffekt unabhängig von einer solchen Ambiguität besteht. Diese Frage bildet den zweiten Teil des hier vorgestellten Forschungsvorhabens.

Im Folgenden präsentieren wir die Ergebnisse zweier empirischer Untersuchungen zur Wahrnehmung von Maskulina movierbarer Nomen in Erstgliedern von Komposita und von Maskulina nicht-movierbarer Nomen, deren Ziel es war, herauszufinden, ob in diesen Fällen die gleiche Ambiguität zwischen einer sexusspezifischen und einer sexusindifferenten Lesart wahrgenommen wird, wie sie bei den Maskulina movierbarer Nomen wahrgenommen wird. Im ersten Experiment wurden Personenbezeichnungen im Plural präsentiert, im zweiten Experiment Personenbezeichnungen im Singular.

## 2 Experiment 1

Ziel des ersten Experiments war es, herauszufinden, ob nicht nur Maskulina von movierbaren Personenbezeichnungen (z. B. *Benutzer*) ambig zwischen einer sexusspezifischen und einer sexusindifferenten Lesart sind, sondern ob eine ähnliche Ambiguität auch für (nicht-referentiell gebrauchbare) Erstglieder von N-N-Komposita (z. B. *Benutzeroberfläche*) und maskuline Epizöna im Plural (z. B. *Gast*) wahrgenommen wird.

### 2.1 Materialien und Methode

Untersucht wurden drei verschiedene Bedingungen: Nomen, die sowohl eine maskuline als auch eine feminine Form haben (z. B. *Lehrer*, das mit der Form *Lehrerin* movierbar ist), Nomen, die nur eine maskuline Form aufweisen (z. B.

<sup>3</sup> Der Ausdruck *Gästin* mag heute ungebräuchlich sein, findet sich allerdings schon im Wörterbuch von Jacob und Wilhelm Grimm (<https://www.woerterbuchnetz.de/DWB?lemid=G01991>).

<sup>4</sup> Zum Beispiel: <https://www.faz.net/aktuell/rhein-main/kultur/gendergerechte-sprache-im-duden-17369494.html>, [https://rp-online.de/panorama/deutschland/duden-gaestin-und-boesewichtin-wie-das-maennliche-geschlecht-in-der-sprache-veraendert-wird\\_aid-56252853](https://rp-online.de/panorama/deutschland/duden-gaestin-und-boesewichtin-wie-das-maennliche-geschlecht-in-der-sprache-veraendert-wird_aid-56252853) oder <https://www.derstandard.at/story/2000124706827/woerterbuch-herausgeberin-eine-feigenblattdiskussion>, zuletzt aufgerufen am 12.05.2022.

<b>Maskulina movierbarer Nomen</b>	<b>Maskuline Epizöna</b>	<b>Komposita</b>
Benutzer	Gäste	Benutzeroberfläche
Kunden	Profis	Kundendienst
Handwerker	Prüflinge	Handwerkergenossenschaft
Mieter	Laien	Mieterrechte
Lehrer	Hoteliers	Lehrerzimmer
Teilnehmer	Vorstände	Teilnehmerliste
Mitarbeiter	Impflinge	Mitarbeiterraum
Besucher	Fans	Besuchereingang
Bürger	Bücherwürmer	Bürgersprechstunde
Studenten	Häuptlinge	Studentenbefragung
Politiker	Stars	Politikerdebatte
Zeugen	Menschen	Zeugenbefragung

**Tabelle 1:** Im Experiment verwendete Wörter.

*Gast*) sowie N-N-Komposita mit einer movierbaren Personenbezeichnung als Erstglied, deren Zweitglied nicht-animat ist (z. B. *Lehrerzimmer*). Die erste Bedingung wird im Folgenden „Maskulinum movierbarer Nomen“ genannt, die zweite Bedingung, also die generisch intendierten Formen der nicht-movierbaren Nomen, „maskuline Epizöna“ und die dritte Bedingung „Komposita“. Die im Experiment verwendeten Wörter sind in Tabelle 1 zusammengefasst.

Die Maskulina movierbarer Nomen sowie die maskulinen Epizöna wurden im Plural präsentiert. Die Maskulina movierbarer Nomen wurden auch für die Erstglieder der Komposita gebraucht, allerdings wurde das Experiment so gestaltet, dass für die Teilnehmer/-innen keine Begriffe doppelt verwendet wurden (es kam also beispielsweise nicht vor, dass ein Teilnehmer/eine Teilnehmerin die Wörter *Benutzer* und *Benutzeroberfläche* sah).

Für das Experiment wurde ein lateinisches Quadrat gebildet und die Wörter in drei Listen eingeteilt, die jeweils vier Items jeder Bedingung enthielten. Die einzelnen Items wurden in Trägersätze eingebettet. Den Probandinnen und Probanden wurden diese Sätze mit dem Hinweis, es handele sich um einen Ausschnitt aus einem Gespräch, gezeigt. Diese Ausschnitte begannen jeweils mit *Moment mal, du hast XY gesagt*. In einem zweiten Satz wurde das hier mit XY bezeichnete Zielwort in Form einer klärenden Rückfrage wieder aufgegriffen. Geklärt werden sollte mit dieser zweiten Frage, ob eine Ambiguität vorliegt oder nicht. Das Design ist im Prinzip an den sogenannten „Moment-mal-Test“ angelehnt, der eigentlich dafür entwickelt wurde, Präsuppositionen oder allgemeiner Not-at-issue-Bedeutung aufzudecken (Shanon 1976; Fintel 2004;

Tonhauser u. a. 2013). Tatsächlich ist der Anwendungsbereich des Tests größer, da er sich auch verwenden lässt, um Ambiguitäten aufzudecken. Dies ist in (1) illustriert, wo die durch *Moment mal* eingeleitete Rückfrage dazu verwendet wird, auf die Ambiguität des Nomens *Maus* hinzuweisen und zu klären, welche Lesart intendiert war.

- (1) A: Ich habe mir eine Maus gekauft.  
 B: Moment mal. Meinst du eine Maus für den Computer oder ein Tier?

Der Test funktioniert auch, wenn das zu desambiguierende Wort das Erstglied eines Kompositums bildet:

- (2) A: Alex ist Schlossbaumeister  
 B: Moment mal. Baut Alex Schlösser, in denen man wohnen kann oder welche, die in Türen eingebaut werden?

Im Folgenden werden drei Beispielstimuli aus dem Experiment gezeigt, wobei zu bemerken ist, dass im Experiment selbst nur Gesprächsausschnitte gezeigt wurden. Während es sich bei (1) und (2) also um Gespräche mit zwei Gesprächspartnerinnen bzw. Gesprächspartnern A und B handelt, wurde im Experiment nur die Antwort der zweiten Person verwendet, um das Experiment simpel und den kognitiven Aufwand möglichst gering zu halten:

- (3) a. **Beispiel Maskulinum eines movierbaren Nomens:**  
 Moment mal, du hast „Mitarbeiter“ gesagt. Heißt das, dass da nur Männer waren?
- b. **Beispiel maskulines Epizönum:**  
 Moment mal, du hast „Stars“ gesagt. Heißt das, dass da nur Männer waren?
- c. **Beispiel Kompositum mit prinzipiell movierbarem Erstglied:**  
 Moment mal, du hast „Mitarbeiterraum“ gesagt. Heißt das, dass das ein Raum nur für Männer ist?

Sollte eine Ambiguität zwischen generischer und geschlechtsspezifischer Lesart vorliegen, sollten solche Dialoge bzw. Dialogauschnitte glücken. Im Prinzip handelt es sich bei dieser Vorgehensweise um eine abgewandelte Form einer Methode, die schon in früheren Studien verwendet wurde, bei der zunächst ein Satz (z. B. im generisch intendierten Maskulinum) präsentiert wird und dann ein fortführender Satz gezeigt wird, der entweder als passend empfunden werden kann oder nicht (siehe z. B. die Studien von Gygas u. a. 2008 oder Körner u. a.

2022). Zu beachten ist, dass diese Methode nur dazu in der Lage ist, potentielle Ambiguitäten aufzudecken, sich jedoch vermutlich weniger dazu eignet, einen möglichen „male bias“, also eine stärkere gedankliche Einbeziehung von männlichen Exemplaren in der kognitiven Verarbeitung, aufzuspüren. Die Probandinnen und Probanden wurden gebeten, einzuschätzen, wie natürlich sie die jeweiligen Gesprächsausschnitte fanden. Konkret lautete die Aufgabenstellung:

Im Folgenden sehen Sie jeweils einen Ausschnitt aus einem Gespräch. Manche dieser Gesprächsfetzen klingen sehr seltsam, andere eher realistisch. Ihre Aufgabe besteht darin, zu sagen, ob die Sätze, die Sie lesen tatsächlich einem natürlichen Gespräch entstammen könnten oder nicht.

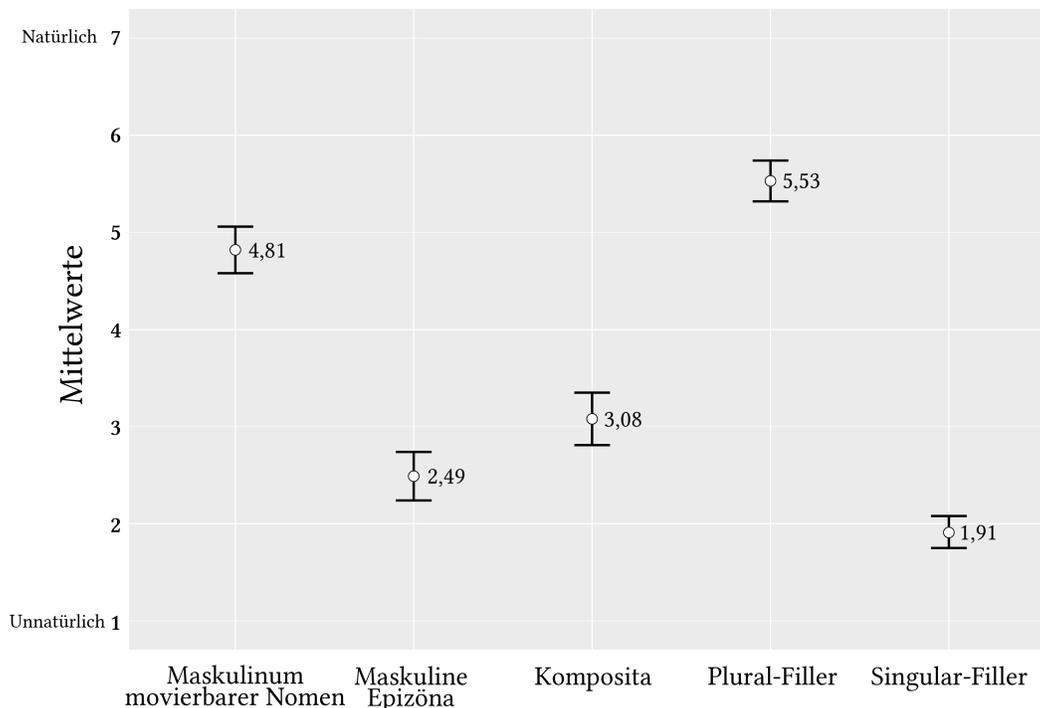
Die Bewertung erfolgte auf einer Skala von 1 („unnatürlich“) bis 7 („natürlich“). Wie erwähnt wurden jedem Teilnehmer bzw. jeder Teilnehmerin vier Items jeder Bedingung zur Bewertung vorgelegt, jeder bzw. jede hatte also insgesamt zwölf Gesprächsausschnitte zu bewerten. Zusätzlich enthielt das Experiment zwölf Filler ähnlichen Aufbaus. Sechs der Filler enthielten ein (nicht-animates) Nomen im Plural und sechs ein (nicht-animates) Nomen im Singular. Auch in diesem Fall begannen die Stimuli mit *Moment mal, du hast XY gesagt*, gefolgt wurden die Filler jedoch von einer Rückfrage, ob wirklich ein Plural gemeint war. Im Folgenden sind zwei Beispiele angeführt:

- (4) a. **Nomen im Plural:**  
Moment mal, du hast „Stifte“ gesagt. Heißt das, dass da mehrere Stifte waren?
- b. **Nomen im Singular:**  
Moment mal, du hast „Pflanze“ gesagt. Heißt das, dass da mehrere Pflanzen waren?

Die Filler-Stimuli sollten nicht nur der Ablenkung dienen, sondern auch dazu, die Teilnehmerinnen und Teilnehmer dazu zu bringen, die gesamte Skala auszunutzen, da erwartet wurde, dass die Filler mit einem Pluralnomen als komplett geglückte und die Filler mit einem Singularnomen als komplett unglückte Gesprächsausschnitte bewertet werden würden.

Das Experiment wurde als Online-Studie implementiert und die Abfolge der Stimuli wurde randomisiert.<sup>5</sup>

<sup>5</sup> Eine Liste der Stimuli findet sich unter [https://osf.io/q3p5r/?view\\_only=cfc1a326477e411684cc59c8c23a107b](https://osf.io/q3p5r/?view_only=cfc1a326477e411684cc59c8c23a107b).



**Abbildung 1:** Ergebnisse des ersten Experiments. Die Fehlerbalken zeigen die 95 %-igen Konfidenzintervalle an.

## 2.2 Teilnehmerinnen und Teilnehmer

Die Teilnehmerinnen und Teilnehmer wurden über verschiedene Mailinglisten der Universität Stuttgart akquiriert. Startpunkt des Experiments war eine eigens eingerichtete Webseite, die die Probandinnen und Probanden zufällig auf eine der Listen verteilte. Die Teilnahme war freiwillig und unbezahlt. Insgesamt nahmen 58 Studierende an dem Experiment teil, davon identifizierten sich 34 als weiblich. Das Durchschnittsalter betrug 24,02 Jahre ( $SD = 2,80$ ). Erstsprache aller Teilnehmerinnen und Teilnehmer war Deutsch.

## 2.3 Ergebnisse

Die Ergebnisse des Experiments sind in Abbildung 1 graphisch aufgearbeitet. Wie erwartet wurden die als Filler verwendeten Gesprächsausschnitte mit Nomen im Plural relativ gut und die mit Nomen im Singular relativ schlecht bewertet. Die Items mit Maskulina movierbarer Nomen wurden relativ

gut bewertet, die maskulinen Epizöna sowie die Komposita dagegen relativ schlecht.<sup>6</sup>

## 2.4 Analyse

Die statistische Analyse erfolgte mit R (R Core Team 2020). Mit dem Paket `ordinal` (Christensen 2019) wurde ein gemischtes Modell für Ordinaldaten, ein sogenanntes „cumulative link mixed model“, mit zufälligen Schnittpunkten („random intercepts“) für die Teilnehmerinnen und Teilnehmer und Items erstellt.<sup>7</sup> Das Modell sagt die Bewertungen der Teilnehmerinnen und Teilnehmer in Abhängigkeit der Bedingungen voraus:

$$(5) \quad \text{clmm}(\text{rating} \sim \text{condition} + (1|\text{participant}) + (1|\text{item}))$$

Die Ergebnisse dieses Modells sind in Tabelle 2 zusammengefasst. Die einzelnen Bedingungen sind in der Tabelle alphabetisch sortiert. Die Tabelle sollte daher eigentlich mit den Komposita starten (die mit „compound“ gelabelt waren). Das Modell hat diese Bedingung als Ausgangspunkt genommen und vergleicht sie mit derjenigen Bedingung, die als nächstes im Alphabet folgt. Dabei handelt es sich um die Maskulina movierbarer Nomen (die als „generic“ kodiert wurde). Die Items, die ein solches Maskulinum enthielten, wurden besser bewertet, als diejenigen, die ein Kompositum enthielten (und zwar um geschätzte 1,5047 Punkte auf der Skala besser). Dieser Effekt ist statistisch signifikant (mit  $p = 0.00000000000873$ ). In der nächsten Zeile werden die Plural-Nomen aus der Filler-Bedingung mit den Komposita verglichen, in der Zeile darunter werden die maskulinen Epizöna (hier mit „conditionpuremasc“ bezeichnet) mit den Komposita verglichen usw.

	Schätzung	Std.-Fehler	z-Wert	Pr(> z )	
conditiongeneric	1,5047	0,2204	6,826	0,00000000000873	***
conditionplural	2,5750	0,2200	11,707	<0,000000000000002	***
conditionpuremasc	-0,6187	0,2245	-2,756	0,00585	**
conditionsingular	-1,3410	0,2108	-6,362	0,000000000199	***

**Tabelle 2:** Resultate des gemischten Modells.

Um eine etwas übersichtlichere Statistik zu erhalten, wurden die einzelnen Bedingungen mit dem Paket `lsmeans` (Lenth 2016) verglichen. Das Ergebnis

<sup>6</sup> Alle Daten sowie eine Dokumentation der statistischen Auswertung findet sich unter [https://osf.io/q3p5r/?view\\_only=cfc1a326477e411684cc59c8c23a107b](https://osf.io/q3p5r/?view_only=cfc1a326477e411684cc59c8c23a107b).

<sup>7</sup> Auf zufällige Steigungen („random slopes“) musste aus Konvergenzgründen verzichtet werden.

dieses paarweisen Vergleichs ist in Tabelle 3 zu sehen. Die Tabelle zeigt, dass sich alle Bedingungen statistisch signifikant voneinander unterscheiden (wenn man von einem Alphalevel von 0,05 ausgeht).

Kontrast	Schätzung	Std.-Fehler	t-Statistik	p-Wert
compound - generic	-1,505	0,220	-6,826	<0,0001
compound - plural	-2,575	0,220	-11,707	<0,0001
compound - puremasc	0,619	0,224	2,756	0,0462
compound - singular	1,341	0,211	6,362	<0,0001
generic - plural	-1,070	0,207	-5,180	<0,0001
generic - puremasc	2,123	0,228	9,319	<0,0001
generic - singular	2,846	0,216	13,146	<0,0001
plural - puremasc	3,194	0,229	13,962	<0,0001
plural - singular	3,916	0,218	17,947	<0,0001
puremasc - singular	0,722	0,214	3,376	0,0066

**Tabelle 3:** Paarweiser Vergleich der verschiedenen Bedingungen mit lsmeans.

Ein weiteres Modell wurde erstellt, das das angegebene Geschlecht der Teilnehmerinnen und Teilnehmer in Betracht zieht, um zu ermitteln, ob dieser Faktor einen Einfluss auf die Bewertungen hatte:

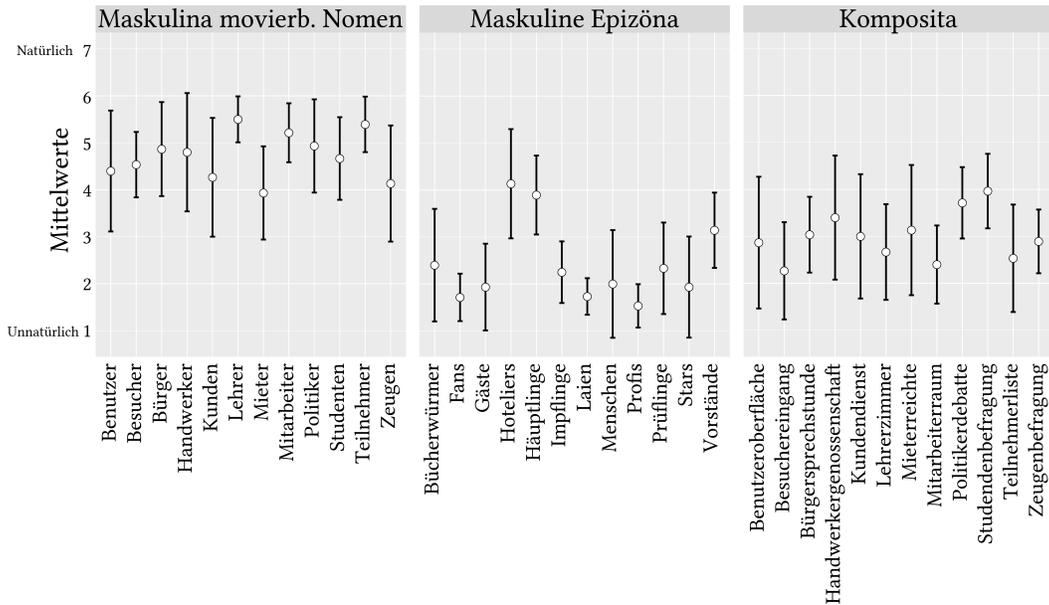
$$(6) \quad \text{clmm}(\text{rating} \sim \text{condition} + \text{gender} + (1|\text{participant}) + (1|\text{item}))$$

Das Ergebnis dieses Modell ist in Tabelle 4 dargestellt. Wie der Tabelle zu entnehmen ist, unterscheidet sich dieses nur marginal vom ursprünglichen Modell. In der letzten Zeile ist zu sehen, dass das Geschlecht der Teilnehmerinnen und Teilnehmer keine statistisch signifikanten Auswirkungen hatte ( $p = 0,59549$ ). Die Bewertungen der Männer und Frauen unterscheiden sich nur marginal (nämlich um geschätzte 0,1081 Punkte auf der Bewertungsskala).

	Schätzung	Std.-Fehler	z-Wert	Pr(> z )	
conditiongeneric	1,5053	0,2201	6,838	8,01e-12	***
conditionplural	2,5757	0,2197	11,724	<2e-16	***
conditionpuremasc	-0,6192	0,2241	-2,763	0,00573	**
conditionsingular	-1,3405	0,2105	-6,369	1,91e-10	***
genderm	0,1081	0,2036	0,531	0,59549	

**Tabelle 4:** Resultate des gemischten Modells mit Geschlecht.

Eine anonyme Gutachterin bzw. ein anonymes Gutachter hat darauf hingewiesen, dass das Item *Vorstände* insofern problematisch sei, „weil das Nomen in



**Abbildung 2:** Ergebnisse des Experiments nach Items aufgeschlüsselt. Die Fehlerbalken zeigen die 95%-igen Konfidenzintervalle an.

der Regel ein aus mehreren Personen bestehendes Gremium bezeichnen dürfte“ und es sich damit von den übrigen verwendeten Begriffen abhebt, die allesamt sowohl einen Plural als auch einen Singular ermöglichen. Sicherlich besteht der prototypische Vorstand aus mehreren Personen, auch wenn das Nomen theoretisch im Singular verwendet werden kann. Ein genauerer Blick auf die Daten zeigt in jedem Fall, dass sich das hier gezeichnete Bild nicht verändert, wenn man das Item in der Analyse nicht berücksichtigt.<sup>8</sup>

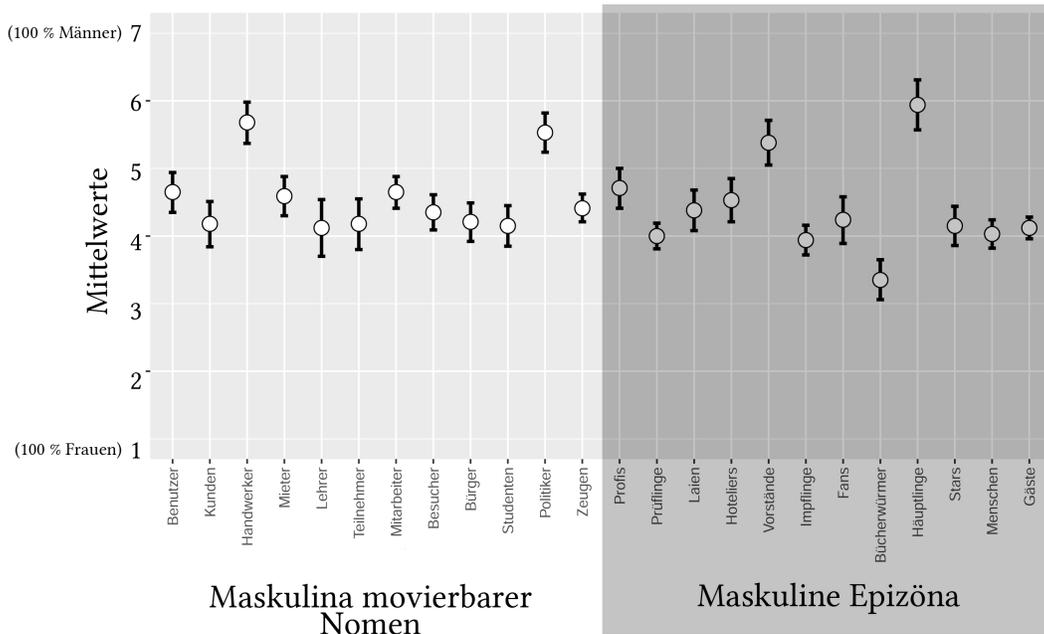
## 2.5 Weiterführende Analyse

Ein Problem mit den verwendeten Stimuli ist sicherlich, dass die Verteilung der Geschlechter innerhalb der Gruppen, die mit den verwendeten Wörtern bezeichnet werden, in der Realität nicht ausgewogen ist. So ist das Geschlechterverhältnis bei der mit *Menschen* bezeichneten Gruppe sicherlich ausgeglichen, bei den mit *Vorständen* oder *Häuptlingen* bezeichneten (potentiellen) Referentinnen und Referenten ist dies jedoch sicherlich recht unausgewogen. Mit anderen Worten: Die stereotype Geschlechterverteilung der durch die einzelnen Wörter beschriebenen Referentinnen und Referenten, die im Experiment ver-

<sup>8</sup> Eine detaillierte Analyse findet sich wiederum unter [https://osf.io/q3p5r/?view\\_only=cfc1a326477e411684cc59c8c23a107b](https://osf.io/q3p5r/?view_only=cfc1a326477e411684cc59c8c23a107b).

wendet wurden, ist relativ variabel (in den meisten Fällen war dies nicht zu vermeiden, vor allem, da etwa die Menge der maskulinen Epizöna im Deutschen begrenzt ist). Zwar wurde in der Vergangenheit argumentiert, dass die stereotype Geschlechtsverteilung bei der Bewertung sprachlichen Materials hinsichtlich der Interpretation des generischen Maskulinums im Deutschen nur eine untergeordnete Rolle spielt und stattdessen das Genus bei der Interpretation von Belang ist (Gygax u. a. 2008), dennoch zeigt ein Blick in die Bewertung der einzelnen Items (siehe Abbildung 2), dass die Stimuli, die etwa die Wörter *Hoteliers* oder *Häuptlinge* beinhalteten, anders bewertet wurden als etwa *Fans* oder *Gäste*. Hier zeigt sich, dass die stereotype Geschlechterverteilung in diesem Experiment eben doch eine Rolle bei der Bewertung der Geglücktheit der Dialogausschnitte gespielt haben dürfte, obwohl die Gesprächsausschnitte deswegen nicht als natürlich, sondern nur als natürlicher empfunden wurden und sich die Konfidenzintervalle dieser Stimuli mit denen vieler anderer Stimuli in der gleichen Bedingung überlappen. Dennoch zeigen die unterschiedlichen Bewertungen der einzelnen Items, dass Personenbezeichnungen nicht einfach generell als sexusindifferent oder sexusspezifisch interpretiert werden, sondern dass auch ein Ableich mit in Frage kommenden Referentinnen und Referenten stattfindet. Ein anonymer Gutachter bzw. eine anonyme Gutachterin stellt hier die Frage, warum sich für die Items *Häuptling* und *Hotelier* ein höherer Natürlichkeitswert ergibt, obwohl in diesen Fällen (wie auch durch die gleich beschriebene Normierungsstudie belegt) klar ist, dass mit einer Bezeichnung wie *Häuptling* hauptsächlich biologisch männliche Referenten adressiert werden können. Dies hat vermutlich eben genau damit zu tun, dass der „male bias“ bei diesen Stimuli extrem hoch ist und dieser starke Verzerrungseffekt für manche Teilnehmerinnen und Teilnehmer eine entsprechende Rückfrage legitimiert hat. Die höhere Bewertung hat also nicht mit der Tatsache zu tun, dass ein Wort wie *Häuptling* ambig zwischen einer generischen und einer sexusspezifischen Lesart ist, sondern mit der Tatsache, dass dieses Nomen nicht nur grammatisch maskulin ist, sondern prototypischerweise auch Männer bezeichnet. Dennoch muss angemerkt werden, dass die Natürlichkeitswerte von *Häuptling* und *Hotelier* sich nicht statistisch signifikant von den anderen Items in dieser Bedingung unterscheiden.

In einer weiterführenden Analyse wurde versucht, die besprochenen Unausgewogenheiten, die durch die realweltlich unausgeglichene Geschlechterverteilung der durch die verwendeten Wörtern bezeichneten Gruppen statistisch zu berücksichtigen. Dazu wurde einer Gruppe von 34 Personen (davon 19 weiblich mit einem Durchschnittsalter von 22,29;  $SD = 2,33$ ), die nicht am eigentlichen Experiment teilnahmen, gebeten, die im Experiment verwendeten Begriffe hinsichtlich ihrer Geschlechterzusammensetzung zu bewerten (das Vorgehen



**Abbildung 3:** Ergebnisse der Normierung.

orientierte sich an Horvath u. a. 2016, die sich wiederum auf Gabriel u. a. 2008 stützen). Die Teilnehmerinnen und Teilnehmer dieser Normierungstudie erhielten jeweils einen Begriff (z. B. *Häuptlinge*) und waren aufgefordertet, auf einer Skala mit sieben Punkten anzugeben, aus wie vielen Frauen bzw. Männern die jeweils bezeichnete Gruppe in der echten Welt besteht. Die Skalenenden standen hier jeweils für eine Gruppe nur aus Frauen bzw. nur aus Männern (*100 % Männer* bzw. *100 % Frauen*). Die Skalenenden wurden ausgeglichen, indem die Frauen in der einen Hälfte der verwendeten Fragebögen das rechte Ende der Skala bildeten und in der anderen das linke Ende. Die Teilnehmerinnen und Teilnehmer wurden zufällig auf die Fragebögen verteilt. Eine Übersicht über die Ergebnisse findet sich in Abbildung 3. Wie zu erwarten wurde das Geschlechterverhältnis in vielen Fällen als ausgewogen oder nahezu ausgewogen eingeschätzt (z. B. bei *Kunden* oder *Menschen*), wobei in dies in anderen Fällen (z. B. bei *Häuptlinge* oder *Vorstände*) nicht der Fall war.

Die Ergebnisse dieser Befragung (hier mit „norm“ bezeichnet) wurden nun in das statistische Modell miteinbezogen:

$$(7) \quad \text{clmm}(\text{rating} \sim \text{condition} + \text{gender} + \text{norm} + (1|\text{participant}) + (1|\text{item}))$$

Tatsächlich hat dies einen statistisch signifikanten Einfluss, wie in Tabelle 5 (letzte Zeile) zu sehen ist. Allerdings ändert sich das grundlegende Bild

nicht, wie ein erneuter paarweiser Vergleich zeigt, dessen Ergebnis in Tabelle 6 zusammengefasst sind (die Filler-Bedingungen fehlen in der Tabelle, weil diese nicht normiert werden konnten). Obwohl sich die drei Bedingungen alle statistisch signifikant voneinander unterscheiden, lässt sich anmerken, dass es einen deutlichen Effektgrößenunterschied gibt. Die Stimuli mit Maskulina movierbarer Nomen wurden laut Modell um geschätzte 1,9798 Punkte auf der Skala besser bewertet als die Komposita und um geschätzte 2,793 Punkte besser als die maskulinen Epizöna. Obwohl es einen statistisch signifikanten Unterschied zwischen Komposita und maskulinen Epizöna gibt und die Gesprächsausschnitte, in welchen sich auf das Erstglied bezogen wurde, etwas besser bewertet wurden, liegen die Bewertungen nur geschätzte 0,813 Punkte auf der Skala auseinander.

	Schätzung	Std.-Fehler	z-Wert	Pr(> z )	
conditiongeneric	1,9798	0,2565	7,719	0,0000000000000117	***
conditionpuremasc	-0,8133	0,2592	-3,138	0,00170	**
genderm	0,1030	0,4181	0,246	0,80551	
norm	0,5298	0,1922	2,756	0,00585	***

**Tabelle 5:** Resultate des gemischten Modells mit Geschlecht.

Kontrast	Schätzung	Std.-Fehler	t-Statistik	p-Wert
compound - generic	-1,980	0,256	-7,719	<0,0001
compound - puremasc	0,813	0,259	3,138	0,0048
generic - puremasc	2,793	0,274	10,200	<0,0001

**Tabelle 6:** Paarweiser Vergleich der verschiedenen Bedingungen mit Ismeans.

## 2.6 Diskussion

Die Tatsache, dass die Gesprächsausschnitte, die ein Maskulinum eines movierbaren Nomens enthielten, relativ gut bewertet wurden, zeigt, dass die generelle Vorgehensweise des Experiments funktioniert, da die Items in dieser Bedingung als ambig zwischen einer generischen und einer rein auf männliche Referenten ausgelegten Lesart bewertet wurden, was sich mit früheren empirischen Befunden deckt (z. B. Braun u. a. 1998; Stahlberg, Sczesny & Braun 2001; Sczesny, Formanowicz & Moser 2016). Dass die Bewertungen dieser Bedingung etwas niedriger ausfielen als die der Plural-Filler ist nicht überraschend, da bei diesen eben keine Ambiguität vorlag.

Die Tatsache, dass die maskulinen Epizöna und die Komposita mit Maskulina movierbarer Nomen als Erstglieder relativ schlecht bewertet wurden, deutet darauf hin, dass diese eher generisch interpretiert werden und kaum eine Ambiguität wahrgenommen wurde. Zwar erhielt der Versuch einer Desambiguierung der Erstglieder etwas bessere Bewertungen als der Versuch der Desambiguierung von lexikalisch maskulinen Nomen, dennoch liegen die durchschnittlichen Bewertungen beider Bedingungen am unteren Ende der verwendeten Skala.

Ein Blick auf die einzelnen Items zeigte, dass die verwendeten Wörter nicht einheitlich bewertet wurden, sondern auch realweltliche Überlegungen mit in die Bewertungen der Teilnehmenden miteingeflossen sein müssen (siehe Abbildung 2). So wurden etwa die Gesprächsausschnitte, die die Epizöna *Hoteliere* und *Hauptlinge* enthielten, anders bewertet als die die meisten anderen verwendeten Epizöna wie *Gäste* oder *Fans*. Man kann dies als Evidenz dafür interpretieren, dass die Probandinnen und Probanden auch die Geschlechtsverteilung realweltlicher Referentinnen und Referenten mit in ihre Überlegungen miteinbezogen haben und sich nicht nur am Genus der Wörter orientierten. Daraus lässt sich ableiten, dass manche Epizöna (z. B. *Gast*) ein höheres Generizitätspotential haben als andere (z. B. *Hauptling*). Allerdings zeigte ein Abgleich mit empirischen Einschätzungen zu realweltlichen Geschlechtsverteilungen, dass der beobachtete Effekt auch unabhängig von dieser realweltlichen Geschlechterverteilung auftritt, wie in der weiterführenden Analyse gezeigt.

### 3 Experiment 2

Empirische Studien zur Wahrnehmung des generisch intendierten Maskulinums haben in der Vergangenheit gezeigt, dass der Numerus eine wichtige Rolle bei der sexusspezifischen bzw. sexusindifferenten Interpretation von Personenbezeichnungen spielt (siehe beispielsweise De Backer & De Cuyper 2012). Konkret deutet die derzeit verfügbare Evidenz darauf hin, dass sich „der maskuline Singular am wenigsten zur geschlechterübergreifenden Verwendung [eignet]“ (Kotthoff & Nübling 2018: 115). Da im ersten Experiment nur Personenbezeichnungen im Plural verwendet wurden, wurde in einem zweiten Experiment getestet, ob sich die Ergebnisse auch auf den Gebrauch von Singularen übertragen lässt. Da es keine Numerusunterscheidung in Erstgliedern von Komposita gibt (auch wenn die Oberflächenformen vieler Komposita mit Singularen bzw. Pluralen synkretisch sind), wurde diese Bedingung im zweiten Experiment nicht weiter untersucht.

### 3.1 Materialien und Methode

Während im ersten Experiment Plurale verwendet wurden und die Rückfrage entsprechend *Heißt das, dass da nur Männer waren?* lautete, wurden im zweiten Experiment wie erwähnt Singulare untersucht und die Rückfrage entsprechend zu *Heißt das, dass da nur ein Mann war?* angepasst. Für das hier besprochene Experiment wurde eine Auswahl der schon im ersten Experiment verwendeten Maskulina movierbarer Nomen sowie der maskulinen Epizöna verwendet. Die verwendeten Wörter sind in Tabelle 7 zusammengefasst.

Generische Mask. movierbarer Nomen	Maskuline Epizöna
Benutzer	Gast
Mitarbeiter	Mensch
Teilnehmer	Laie
Zeuge	Star
Besucher	Fan
Bürger	Impfling
Kunde	Profi
Mieter	Prüfling

**Tabelle 7:** Im zweiten Experiment verwendete Wörter.

Bis auf die Tatsache, dass die Stimuli und die Rückfrage im Singular präsentiert wurden, war das Vorgehen wie im ersten Experiment. Für das Experiment wurde ein lateinisches Quadrat gebildet und die Wörter in zwei Listen eingeteilt, die jeweils vier Items jeder der beiden Bedingung enthielten. Im Folgenden finden sich zwei Beispiel-Stimuli:<sup>9</sup>

- (8) a. **Beispiel generisches Maskulinum movierbarer Nomen:**  
Moment mal, du hast „Kunde“ gesagt. Heißt das, dass da nur ein Mann war?
- b. **Beispiel maskuline Epizöna:**  
Moment mal, du hast „Mensch“ gesagt. Heißt das, dass da nur ein Mann war?

Sollte eine Ambiguität zwischen generischer und geschlechtsspezifischer Lesart vorliegen, sollten solche Dialoge bzw. Dialogauschnitte glücken. Die Bewertung erfolgte wieder auf einer Skala von 1 („unnatürlich“) bis 7 („natürlich“).

<sup>9</sup> Im Laufe des Reviewverfahrens wurde angemerkt, dass es für zukünftige Experiment Sinn ergeben würde, die Rückfrage in *Heißt das, dass da nur ein Mann war und keine Frau?* unzuformulieren, um eine Missinterpretation im Sinne von ‚Heißt das, dass da nur ein einzelner Mann war und nicht mehrere‘ zu vermeiden.

Zusätzlich enthielt das Experiment dieselben zwölf Filler wie im ersten Experiment. Nachdem die Teilnehmenden die Bewertungen der Gesprächsausschnitte vorgenommen hatten, wurden sie in einem zweiten Teil des Experiments gebeten, dieselben Wörter hinsichtlich der Frage zu bewerten, aus wie vielen Männern bzw. Frauen die bezeichneten Gruppen in der echten Welt bestünden (*100 % Männer* bzw. *100 % Frauen*). Hierfür wurden die Maskulina movierbarer Nomen beispielsweise als *Mitarbeiter und Mitarbeiterinnen* und die maskulinen Epizöna beispielsweise mit *männliche Gäste und weibliche Gäste* bezeichnet. Das Experiment wurde als Online-Studie implementiert und die Abfolge der Stimuli wurde randomisiert.<sup>10</sup>

### 3.2 Teilnehmerinnen und Teilnehmer

Auch die Teilnehmerinnen und Teilnehmer des zweiten Experiments wurden über Mailinglisten rekrutiert, allerdings über andere Listen als die im ersten Experiment verwendeten, um sicherzustellen, dass es sich nicht um Personen handelt, die schon am ersten Experiment teilgenommen hatten. Insgesamt nahmen 45 Personen an diesem Experiment teil, von welchen sich 29 als weiblich und eine Person als divers identifizierten. Das Durchschnittsalter der Teilnehmenden betrug 23,60 Jahre ( $SD = 4,05$ ). Die Erstsprache aller Teilnehmenden war Deutsch.

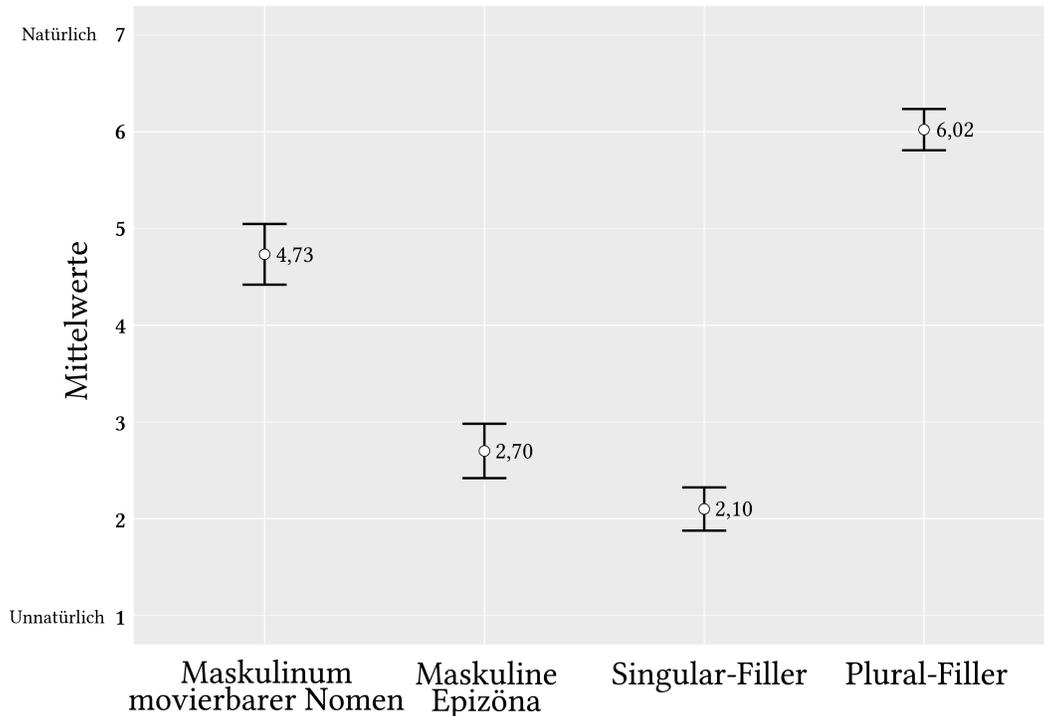
### 3.3 Ergebnisse

Die Ergebnisse des Experiments sind in Abbildung 4 graphisch aufgearbeitet. Wie der Abbildung zu entnehmen entsprechen die Ergebnisse des zweiten Experiments im Wesentlichen denen des ersten Experiments. Wie erwartet wurden die Gesprächsausschnitte mit Nomen im Plural relativ gut und die mit Nomen im Singular wieder relativ schlecht bewertet. Die Items mit generischen Maskulina movierbarer Nomen wurden relativ gut bewertet, die maskulinen Epizöna dagegen relativ schlecht.<sup>11</sup>

Abbildung 5 zeigt die die Bewertungen der einzelnen Stimuli. Wie der Abbildung zu entnehmen gab es etwas Variation bei der Bewertung der verschiedenen Stimuli, wobei das grobe Bild keine größeren Abweichungen ergibt: Die generischen Maskulina movierbarer Nomen wurden insgesamt besser bewertet als die

<sup>10</sup> Eine Liste der Stimuli findet sich unter [https://osf.io/q3p5r/?view\\_only=cfc1a326477e411684cc59c8c23a107b](https://osf.io/q3p5r/?view_only=cfc1a326477e411684cc59c8c23a107b).

<sup>11</sup> Alle Daten sowie eine Dokumentation der statistischen Auswertung findet sich unter [https://osf.io/q3p5r/?view\\_only=cfc1a326477e411684cc59c8c23a107b](https://osf.io/q3p5r/?view_only=cfc1a326477e411684cc59c8c23a107b).

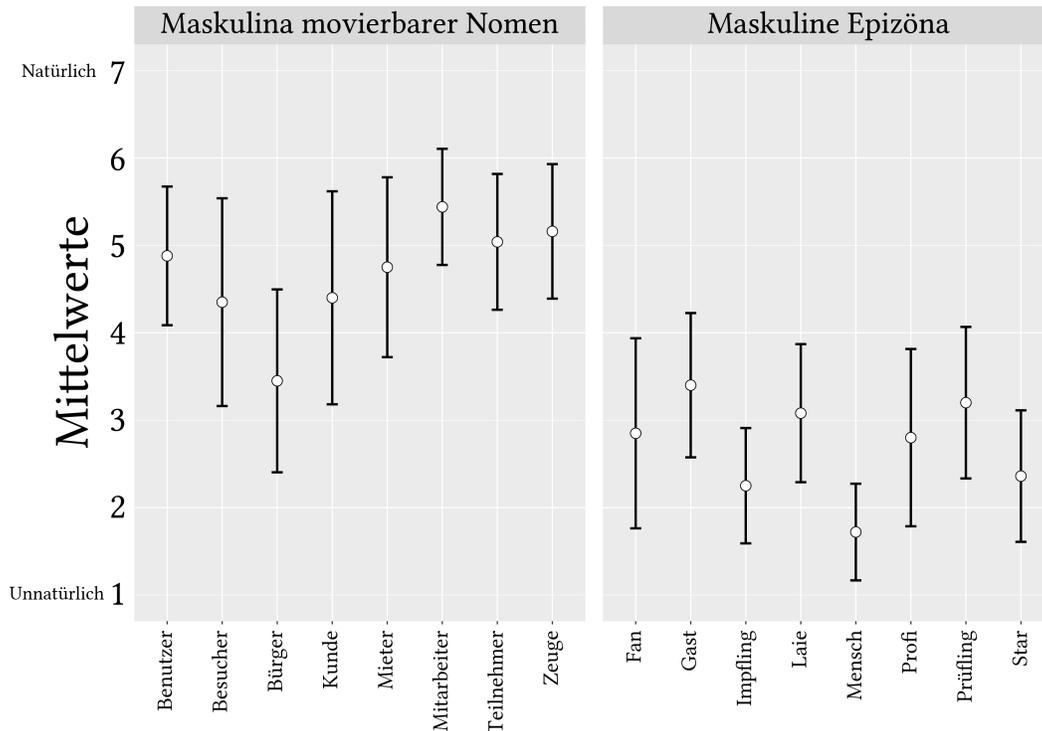


**Abbildung 4:** Ergebnisse des zweiten Experiments. Die Fehlerbalken zeigen die 95 %-igen Konfidenzintervalle an.

maskulinen Epizöna.<sup>12</sup> Einziger Ausreißer in der ersten Gruppe ist das Nomen *Bürger*, wobei das Konfidenzintervall mit den meisten anderen Konfidenzintervallen überlappt. In der zweiten Gruppe wurden *Imfpling* und *Mensch* relativ schlecht bewertet, wobei auch hier Konfidenzintervallüberlappungen existieren, man also nicht von echten Ausreißern sprechen kann.

Die im zweiten Teil des Experiments erhobenen Daten zur Frage, aus wie vielen Männern bzw. Frauen die einzelnen Gruppen in der echten Welt bestehen, sind in Abbildung 6 zu sehen. Wie der Abbildung zu entnehmen ist, wurde das Geschlechterverhältnis als relativ ausgewogen eingeschätzt.

<sup>12</sup> Ein Blick in die Abbildung zeigt noch etwas anderes: Manche der movierbaren Nomen sind im Singular und im Plural synkretisch (z. B. *ein Besucher* versus *mehrere Besucher*), andere dagegen nicht (z. B. *ein Kunde* versus *mehrere Kunden*). Dies scheint jedoch keinen Einfluss auf die Ergebnisse gehabt zu haben, wie aus den sich überlappenden Konfidenzintervallen hervorgeht (außerdem wurde durch die Rückfrage klar, dass eine singularische Interpretation intendiert war).



**Abbildung 5:** Ergebnisse des zweiten Experiments nach einzelnen Stimuli aufgeschlüsselt. Die Fehlerbalken zeigen die 95 %-igen Konfidenzintervalle an.

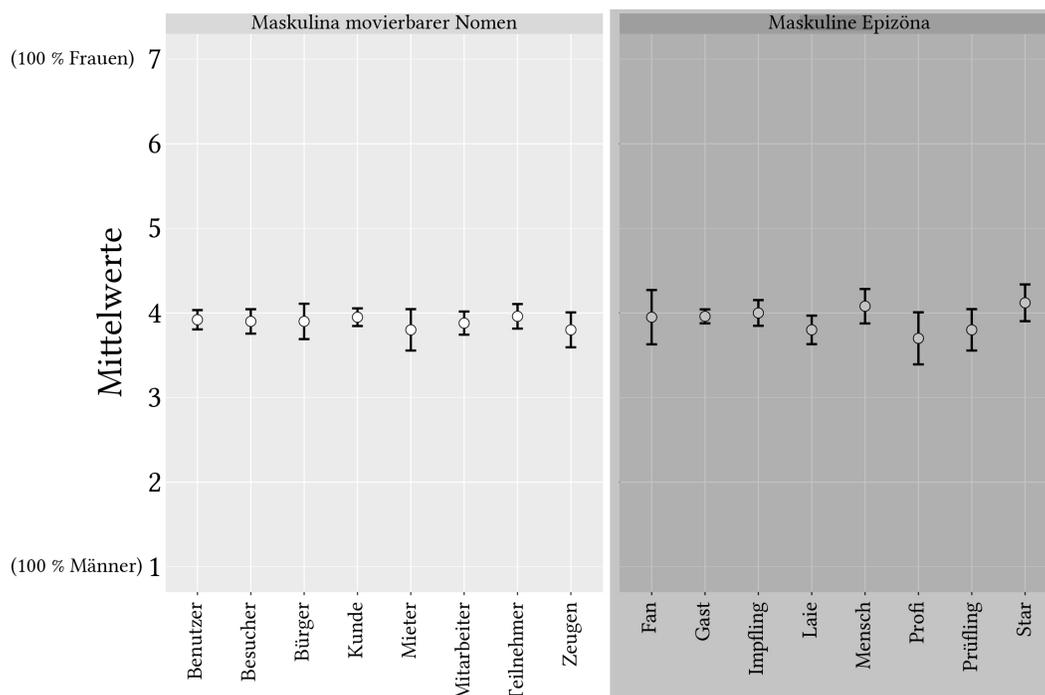
### 3.4 Analyse

Die statistische Analyse erfolgte wiederum mit R (R Core Team 2020). Mit dem Paket `ordinal` (Christensen 2019) wurde wieder ein gemischtes Modell für Ordinaldaten mit zufälligen Schnittpunkten („random intercepts“) für die Teilnehmerinnen und Teilnehmer und Items erstellt.<sup>13</sup> Das Modell sagt die Bewertungen der Teilnehmerinnen und Teilnehmer in Abhängigkeit der Bedingungen voraus, wobei das von den Teilnehmenden angegebene Geschlecht sowie die Ergebnisse der Frage nach den realweltlichen Geschlechterverhältnissen berücksichtigt wurde:

$$(9) \quad \text{clmm}(\text{rating} \sim \text{condition} + \text{gender} + \text{norm} + (1|\text{participant}) + (1|\text{item}))$$

Die Ergebnisse dieses Modells sind in Tabelle 8 zusammengefasst. Die einzelnen Bedingungen sind in der Tabelle wieder alphabetisch sortiert. Die Tabelle sollte daher eigentlich mit generisch intendierten Maskulina movierter Nomen

<sup>13</sup> Auch hier wurde aus Konvergenzgründen auf zufällige Steigungen („random slopes“) verzichtet.



**Abbildung 6:** Ergebnisse des zweiten Teils des zweiten Experiments. Die Fehlerbalken zeigen die 95 %-igen Konfidenzintervalle an.

starten (die mit „genmask“ gelabelt waren). Das Modell hat diese Bedingung als Ausgangspunkt genommen und vergleicht sie mit derjenigen Bedingung, die als nächstes im Alphabet folgt. Daher startet die Tabelle mit den maskulinen Epizöna („puremasc“). Diese Bedingung unterscheidet sich statistisch signifikant von den generischen Maskulina movierbarer Nomen ( $p = 0,00000000475$ ), wie aus der ersten Zeile der Tabelle hervorgeht. In derselben Zeile kann man ablesen, dass das Modell schätzt, dass die generischen Maskulina movierbarer Nomen 2,44 Punkte auf der Skala besser bewertet wurden als die maskulinen Epizöna. Die nächsten beiden Zeilen zeigen, dass das von den Teilnehmenden angegebene Geschlecht keine statistisch signifikante Rolle bei der Bewertung gespielt hat und die letzte Zeile, dass die Einschätzung der Teilnehmenden, aus wie vielen Männern oder Frauen die bezeichneten Gruppen in der echten Welt bestehen, ebenfalls keinen statistisch signifikanten Einfluss hatte.

### 3.5 Diskussion

Wie schon die bei den Plural-Stimuli im ersten Experiment, so zeigt sich auch im zweiten Experiment, dass maskuline Formen movierbarer Nomen eher als

	Schätzung	Std.-Fehler	z-Wert	Pr(> z )	
conditionpuremasc	2,4384	0,3916	6,227	0,000000000475	***
gendermale	0,2565	0,6298	0,407	0,684	
gendernonbinary	-1,0138	1,9619	-0,517	0,605	
norm	0,1234	0,2633	0,469	0,639	

**Tabelle 8:** Resultate des gemischten Modells.

ambig zwischen einer generischen und einer sexusspezifischen Lesart empfunden werden als maskuline Epizöna – auch wenn sie im Singular stehen.

## 4 Allgemeine Diskussion und Einschränkungen der Studie

Bisherige Studien zur Wahrnehmung des generisch intendierten Maskulinums und gendersensiblen Formen haben sich hauptsächlich darauf konzentriert, Probandinnen und Probanden eindeutige Sexuzuordnungen in konkreten (wenn auch fiktiven) Situationen abzuverlangen. Ziel der vorliegenden Experimente war es dagegen, potentielle Ambiguitäten verschiedener Formen in einer Offline-Situation zu elizitieren bzw. herauszufiltern, ob solche Ambiguitäten überhaupt vorliegen. Hierfür wurden die Probandinnen und Probanden gebeten, Rückfragen hinsichtlich einer solchen Ambiguität zu bewerten. Insgesamt zeigen die Ergebnisse, dass maskuline Formen movierbarer Nomen, die Personen- oder Personengruppen bezeichnen, als ambig zwischen einer sexusspezifischen und einer sexusindifferenten Lesart interpretiert werden. Dies gilt jedoch weder für nicht-referierende Erstglieder von Komposita noch für maskuline Epizöna.

Die Ergebnisse der vorliegenden Experimente sind aus verschiedenen Gründen mit Vorsicht zu genießen. Obwohl für die Bedingungen generisches Maskulinum movierbarer Nomen und Komposita im ersten Experiment die gleichen Substantive (einmal als selbstständige Nomen und einmal als Erstglieder) verwendet wurden, die dieselbe Oberflächenform haben, gibt es eine gewisse Einschränkung bei deren Vergleichbarkeit: Während es sich beispielsweise bei der generisch intendierten Form *Kunden* um einen Plural handelt, liegt beim Erstglied *Kunden-* in *Kundendienst* trotz Formgleichheit kein echter Plural vor. Stattdessen handelt es sich um eine aus dem Genitiv abgeleitete Form, die als Fugen-Konstruktion analysiert wird (Nübling & Szczepaniak 2008).

Generell eignen sich Komposita dennoch gut dazu, zu illustrieren, dass es insgesamt weniger relevante und relevantere Kontexte für das Gendern gibt, die mit der in jüngerer Zeit immer stärker in den Fokus der Forschung rückenden

Frage nach der Referenz sprachlicher Ausdrücke zusammenhängen (Zifonun 2018). Konkret wurde im ersten Experiment gezeigt, dass Personenbezeichnungen als Erstglieder von Komposita kein relevanter Kontext für das Gendern zu sein scheinen. Der Grund hierfür wurde in der Literatur bereits angemerkt: Erstglieder von Komposita führen nicht zu Referenz (Diewald & Steinhauer 2017: 65–68), nur das ganze Kompositum als solches referiert. Dass bei Erstgliedern von Komposita kaum eine Ambiguität zwischen sexusspezifischer und sexusindifferenter Interpretation vorzuliegen scheint, bedeutet aber natürlich nicht, dass hier kein „male bias“ evoziert wird. Ein wichtiger Punkt bezieht sich also auf die Frage, was im Experiment eigentlich genau gemessen wurde. Obwohl die Daten nahelegen, dass generische Maskulina, die als Erstglieder von Komposita dienen, nur generisch interpretiert werden können, sagt dies vermutlich nichts über die gedankliche Miteinbeziehung von Frauen und nicht-binären Personen bei der Verarbeitung solcher Komposita aus. Es bleibt – etwas überspitzt formuliert – zu testen, ob bei der Verarbeitung eines Kompositums wie *Germanistenverband* nicht doch die Vorstellung eines Verbandes erzeugt wird, der hauptsächlich aus Männern besteht, auch wenn das Erstglied hier rein sprachlich betrachtet generisch interpretiert wird. Mit anderen Worten: In der vorliegenden Studie wurde untersucht, ob die Erstglieder von Komposita sprachlich generisch behandelt werden, was aber nicht bedeutet, dass kein männlicher „bias“ vorliegt. Insgesamt bleibt festzuhalten, dass das Vorliegen bzw. die Abwesenheit einer Ambiguität kein Argument für bzw. gegen den Gebrauch gendersensibler Sprache ist. Ein weiteres Problem der vorliegenden Studie hinsichtlich der untersuchten Komposita ist, dass es zahlreiche verschiedene semantische Relationen zwischen den Gliedern von Komposita gibt, die sicherlich eigentlich im Einzelnen beleuchtet werden sollten.

Auch für die maskulinen Epizöna gilt natürlich, dass die Möglichkeit besteht, dass diese zu einer verzerrten mentalen Repräsentation der Geschlechter der (potentiellen) Referentinnen und Referenten führt, auch wenn diese rein sprachlich eher generisch interpretiert werden, und in genau diese Richtung deutet beispielsweise ein Experiment von Irmen & Roßberg (2004). In dieser Lesezeit-Studie lasen die Probandinnen und Probanden Sätze, die unter anderem maskuline Epizöna enthielten (z. B. das Nomen *Laien*) und denen ein Satz folgte, der entweder stereotyp männlich, weiblich oder in dieser Hinsicht neutral war. Folgten den maskulinen Epizöna ein stereotypisch männlicher oder neutraler Satz, waren die Lesezeiten schneller im Vergleich zu der Bedingung, in der ein stereotypisch weiblicher Satz folgte. Insgesamt zeigt sich also, dass man einen Unterschied machen muss zwischen dem Vorliegen bzw. der Abwesenheit von Ambiguitäten auf der einen und einem „male bias“ auf der anderen Seite. Des weiteren sollte in weiteren Studien die Wahrnehmung bzw. das Vorliegen

potentieller Ambiguitäten neutraler (z. B. *das Kind*) und femininer Epizöna (z. B. *die Person*) untersucht werden.

Wie von einer anonymen Gutachterin bzw. einem anonymen Gutachter richtig angemerkt, sollte in zukünftigen Studien auch der Unterschied zwischen Rollenreferenz (z. B. der Bezug auf eine Berufsgruppe; *Führungskraft zu sein, bedeutet Verantwortung zu tragen*) und Sexusreferenz (der konkrete Bezug auf eine Person oder Gruppe; *Da war ein Bäcker im Raum*) stärker berücksichtigt werden, wobei die verwendeten Stimuli klar sexusreferentiell zu interpretieren sind (*Moment mal, heißt das, dass da nur Männer waren?*). Um diese Unterscheidung zu untersuchen, könnte eine abgewandelte Form des *Moment-Mal*-Tests Verwendung finden (z. B. *Moment mal, du hast „Führungskraft“ gesagt. Heißt das, dass dieser Beruf nur für Männer geeignet ist?*).

## 5 Fazit

Deutsche Nomen, die Personen oder Personengruppen bezeichnen, erlauben in vielen Fällen eine maskuline und eine feminine Form (*Wissenschaftler* bzw. *Wissenschaftlerin*), sie sind also movierbar. Die maskulinen Formen sind ambig zwischen einer Lesart, die nur Männer bezeichnet und einer (zumindest intendiert) generischen Lesart, die sowohl Männer als auch Frauen bezeichnet. In der vorliegenden Studie wurde untersucht, ob diese Ambiguität auch bei maskulinen Epizöna auftritt, die keine feminine Form haben und ob diese Ambiguität auch bei generischen Maskulina auftritt, die als Erstglieder von Komposita fungieren. Die Ergebnisse des vorgestellten Experiments deuten darauf hin, dass sowohl maskuline Epizöna (im Singular wie im Plural) als auch generische Maskulina in Komposita-Erstglieder eher generisch interpretiert werden. Allerdings bleibt zu untersuchen, ob es sich hierbei um einen rein sprachlichen Effekt handelt oder ob maskuline Epizöna und generische Maskulina in Komposita nicht doch zu geschlechtsverzerrten mentalen Repräsentationen führen. Insgesamt zeigt sich, dass die Erforschung von gendersensibler Sprache eine Vielzahl an sich gegenseitig beeinflussenden Faktoren berücksichtigen muss, zu denen die Frage des referentiellen Modus gehört, die Frage ob potentielle Ambiguitäten vorliegen oder nicht und ob es eine mentale Verzerrung bei der evozierten Repräsentation gibt.

Deutsche Personenbezeichnungen movierbarer Nomen sind, wie die vorliegende Studie erneut empirisch gezeigt hat, ambig zwischen einer sexusindifferenten und einer sexusspezifischen Lesart. Diese Ambiguität wird – neben dem Vorliegen eines „male bias“  
– immer wieder als Argument für die Verwendung gendersensibler Sprache

verwendet. Die hier präsentierten Daten legen nahe, dass die Verwendung gendersensibler Sprache jedoch nicht in allen Fällen notwendig ist. So lassen die vorgestellten Daten darauf schließen, dass die besprochene Ambiguität bei Erstgliedern von Komposita nicht wahrgenommen wird, da diese nicht referentiell verwendet werden. Daher macht es – zumindest aus der Perspektive der Ambiguitätswahrnehmung – wenig Sinn, solche Erstglieder zu gendern (wie z. B. in *Lehrer\*innenzimmer*). Des Weiteren konnte keine Ambiguität zwischen einer sexusspezifischen und einer sexusindifferenten Lesart bei maskulinen Epizöne nachgewiesen werden. Daraus könnte man ableiten, dass es nicht notwendig ist, movierte Formen solcher Personenbezeichnungen abzuleiten (wie z. B. *Vorständin*). Diese Vorschläge gelten jedoch wie angesprochen nur aus der Perspektive der Ambiguitätswahrnehmung und sagen noch nichts über einen potentiellen „male bias“ aus.

## Literatur

- Braun, Friederike, Anja Gottburgsen, Sabine Sczesny & Dagmar Stahlberg. 1998. Können *Geophysiker* Frauen sein? Generische Personenbezeichnungen im Deutschen. *Zeitschrift für germanistische Linguistik* 26. 265–283. <https://doi.org/10.1515/zfgl.1998.26.3.265>.
- Brendel, Caren & Agnes Speck. 2021. *Mitgemacht statt mitgedacht – Grundlagen und praktische Anwendungen einer „Sprache für alle“*. Begleittext zum Online Tutorial „Eine Sprache für alle“ für einen inklusiven Sprachgebrauch. Online: [https://www.uni-heidelberg.de/md/gsb/begleittext\\_zum\\_online-training\\_\\_gendergerechte\\_sprache\\_.pdf](https://www.uni-heidelberg.de/md/gsb/begleittext_zum_online-training__gendergerechte_sprache_.pdf), zuletzt aufgerufen am 28.04.2022.
- Christensen, Rune Haubo Bojesen. 2019. *ordinal—Regression models for ordinal data*. R package version 2019.12-10. <https://CRAN.R-project.org/package=ordinal>.
- De Backer, Maarten & Ludovic De Cuyper. 2012. The interpretation of masculine personal nouns in German and Dutch: A comparative experimental study. *Language Sciences* 34(3). 253–268. <https://doi.org/10.1016/j.langsci.2011.10.001>.
- Diewald, Gabriele & Anja Steinhauer. 2017. *Richtig gendern: Wie Sie angemessen und verständlich schreiben*. Berlin: Dudenverlag.
- Doleschal, Ursula. 1992. *Movierung im Deutschen: eine Darstellung der Bildung und Verwendung weiblicher Personenbezeichnungen*. Unterschleissheim/München: Lincom.
- Fintel, Kai von. 2004. Would you believe it? The King of France is back! (Presuppositions and truth-value intuitions). In Marga Reimer & Anne

- Bezuidenhout (Hrsg.), *Descriptions and Beyond*, 315–341. Oxford: Oxford University Press.
- Friedrich, Marcus CG, Veronika Drößler, Nicole Oberlehberg & Elke Heise. 2021. The Influence of the Gender Asterisk („Gendersternchen“) on Comprehensibility and Interest. *Frontiers in Psychology* 12. <https://doi.org/10.3389/fpsyg.2021.760062>.
- 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>.
- Gabriel, Ute, Pascal M Gygax & Elisabeth A Kuhn. 2018. Neutralising linguistic sexism: Promising but cumbersome? *Group Processes & Intergroup Relations* 21(5). 844–858. <https://doi.org/10.1177/1368430218771742>.
- Gabriel, Ute & Franziska Mellenberger. 2004. Exchanging the Generic Masculine for Gender-Balanced Forms – The Impact of Context Valence. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie* 63(4). 273. <https://doi.org/10.1024/1421-0185.63.4.273>.
- 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>.
- Hansen, Karolina, Cindy Littwitz & Sabine Sczesny. 2016. The social perception of heroes and murderers: Effects of gender-inclusive language in media reports. *Frontiers in psychology* 7. 369. <https://doi.org/10.3389/fpsyg.2016.00369>.
- Harnisch, Rüdiger. 2016. Das generische Maskulinum schleicht zurück. In Andreas Bittner & Constanze Spieß (Hrsg.), *Formen und Funktionen*, 159–174. Berlin & Boston: de Gruyter. <https://doi.org/10.1515/9783110478976-010>.
- Heise, Elke. 2000. Sind Frauen mitgemeint? Eine empirische Untersuchung zum Verständnis des generischen Maskulinums und seiner Alternativen. *Sprache & Kognition* 19. <https://doi.org/10.1024/0253-4533.19.12.3>.
- Hornscheidt, Lann & Ja'n Sammla. 2021. *Wie schreibe ich divers? Wie schreibe ich gendergerecht? Ein Praxis-Handbuch zu Gender und Sprache*. Hiddensee: w\_orte und meer.
- Horvath, Lisa K, Elisa F Merkel, Anne Maass & Sabine Sczesny. 2016. Does gender-fair language pay off? The social perception of professions from a cross-linguistic perspective. *Frontiers in Psychology* 6. 2018. <https://doi.org/10.3389/fpsyg.2015.02018>.

- Irmen, Lisa & Nadja Roßberg. 2004. Gender markedness of language: The impact of grammatical and nonlinguistic information on the mental representation of person information. *Journal of Language and Social Psychology* 23(3). 272–307. <https://doi.org/10.1177/0261927X04266810>.
- Körner, Anita, Bleen Abraham, Ralf Rummer & Fritz Strack. 2022. Gender Representations Elicited by the Gender Star Form. *Journal of Language and Social Psychology*. 0261927X221080181. <https://doi.org/10.1177/0261927X221080181>.
- Kotthoff, Helga. 2020. Gender-Sternchen, Binnen-I oder generisches Maskulinum, ... (Akademische) Textstile der Personenreferenz als Registrierungen? *Linguistik Online* 103(3). 105–127. <https://doi.org/10.13092/lo.103.7181>.
- Kotthoff, Helga. 2022. Zwischen berechtigtem Anliegen und bedenklicher Symbolpolitik. *Aus Politik und Zeitgeschichte* 72. 12–13.
- Kotthoff, Helga & Damaris Nübling. 2018. *Genderlinguistik: Eine Einführung in Sprache, Gespräch und Geschlecht*. Tübingen: Narr Francke Attempto.
- Lenth, Russell V. 2016. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software* 69(1). 1–33. <https://doi.org/10.18637/jss.v069.i01>.
- Müller-Spitzer, Carolin. 2022. Zumutung, Herausforderung, Notwendigkeit? Zum Stand der Forschung zu geschlechtergerechter Sprache. *Aus Politik und Zeitgeschichte* 72. 23–29.
- Nübling, Damaris & Renata Szczepaniak. 2008. On the way from morphology to phonology: German linking elements and the role of the phonological word. *Morphology* 18(1). 1–25. <https://doi.org/10.1007/s11525-008-9120-7>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Sczesny, Sabine, Magda Formanowicz & Franziska Moser. 2016. Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology* 7. 25. <https://doi.org/10.3389/fpsyg.2016.00025>.
- Sczesny, Sabine, Franziska Moser & Wendy Wood. 2015. Beyond sexist beliefs: How do people decide to use gender-inclusive language? *Personality and Social Psychology Bulletin* 41(7). 943–954. <https://doi.org/10.1177/0146167215585727>.
- Shanon, Benny. 1976. On the two kinds of presuppositions in natural language. *Foundations of Language* 14(2). 247–249.
- Stahlberg, Dagmar, Friederike Braun, Lisa Irmen & Sabine Sczesny. 2007. Representation of the sexes in language. In Klaus Fiedler (Hrsg.), *Social communication*, 163–187. New York & Hove: Psychology Press.

- Stahlberg, Dagmar & Sabine Sczesny. 2001. Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen. *Psychologische Rundschau* 52. 131–140.
- Stahlberg, Dagmar, Sabine Sczesny & Friederike Braun. 2001. Name your favorite musician: Effects of masculine generics and of their alternatives in German. *Journal of Language and Social Psychology* 20(4). 464–469. <https://doi.org/10.1177/0261927X01020004004>.
- Tonhauser, Judith, David Beaver, Craige Roberts & Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89(1). 66–109.
- Zifonun, Gisela. 2018. Die demokratische Pflicht und das Sprachsystem: Erneute Diskussion um einen geschlechtergerechten Sprachgebrauch. *Sprachreport* 34(4). 44–56.