Language and Cognition (2025), 17, e73, 1–26
doi:10.1017/langcog.2025.10032

ARTICLE

Gender conflicts in German possessives: comparing inanimate to human reference reveals asymmetries

Christin Schütze[1], Jürg Fleischer[2] and Ulrike Domahs[1]

[1]Institut für Germanistische Sprachwissenschaft, Philipps-Universität Marburg, Marburg, Germany and [2]Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin, Berlin, Germany
Corresponding author: Christin Schütze; Email: christin.schuetze@staff.uni-marburg.de

(Received 23 June 2025; Revised 22 August 2025; Accepted 22 August 2025)

Abstract

Although German, as a grammatical gender language, requires noun–pronoun agreement in anaphora, exceptions to the rule occur, e.g., in possessive constructions when the gender-incongruent possessive pronoun *sein* (masculine/neuter, his/its) refers to feminine antecedents instead of congruent *ihr* (feminine, her). While this violation is merely grammatical for inanimate referents, it can provoke semantic mismatches for human possessors (especially gender-specific female nouns like *die Hexe*$_i$ – *sein*$_i$ (the witch – his/its), but less so with gender-indifferent human nouns, such as *die Kontaktperson* (the contact person). A self-paced reading (SPR) experiment tested the acceptability and processing of sentences in which incongruent *sein* referred to feminine possessors, which differed in animacy status (inanimate versus human). Introducing this agreement violation reduced acceptability and elevated reading and reaction times (RTs), but effects varied by antecedent animacy. These results suggest an animacy restriction in possessive reference and substantiate the impact of meaning-based gender cues on pronominalization.

Keywords: acceptability judgments; anaphoric reference processing; grammatical gender; possessive pronouns; referential gender; self-paced reading

1. When gender (dis)agrees

'For every rule, there is an exception' – and rules of grammar are no exception to this common idiom; otherwise, one cannot explain the plentitude of linguistic cases of doubt (Duden, 2016) or the phenomenon of interest here: the existence and possible acceptability of possessive pronouns that fail to agree in gender in German, a

The research was conducted at Philipps-Universität Marburg (now Universität Marburg), Neurolinguistics lab. All authors consent for the manuscript to be submitted in its current form.

language with anaphoric pronominal agreement based on a noun's gender class. In languages with gender distinctions, pronouns fulfill the task of referent tracking, a crucial requirement to obtain discourse coherence and disambiguation in communication, which makes anaphora interpretation a vital component of language comprehension (Garnham, 2001, 36). Hence, the German possessive pronouns *ihr* (feminine) and *sein* (masculine/neuter) must conform to the gender features of their referent to indicate possession of an identified antecedent and maintain syntactic and semantic cohesion and, thus, follow morphological and semantic or pragmatic rules – a functional duality that can cause vast variation, as this paper will explore.

1.1. Possessive anaphoric reference in German

The German gender system assigns one of three values to a noun (feminine, masculine, neuter) and overtly and obligatorily marks this gender category on co-referring expressions like pronouns and other satellite elements (Hellinger & Bußmann, 2015, 7, 13). German requires syntactic gender agreement based on the three grammatical gender categories of all nouns, be they inanimate or animate, but also permits semantic gender agreement based on lexical, referential and/or stereotypical gender inferred for person nouns, as will be exemplified below (Section 1.2). As a consequence, different types of information can control the form of agreement in the human domain: There are two potentially competing routes (sometimes referred to as two available cues, cf. Cacciari et al., 1997, or rules, cf. Oelkers, 1996, 4) to establish a feature match with human antecedents based on morpho-syntactic grounds or else picking up the underlying lexical or conceptual information to referent gender (Caffarra et al., 2014). Generally, feature properties (such as gender or number) are 'inherited' (Osterhout & Mobley, 1995, 740) from the controller (i.e., possessor noun) to the controllee or target (i.e., possessive pronoun) in Corbett's (2006) terminology.

What follows is that the gender features of an anaphoric pronoun are usually predictable from the gender class the noun belongs to, thus imposing syntactic agreement constraints, which we call grammatical or formal gender, e.g., *das Kind – seine Eltern* (the child (neut.) – its (neut.) parents). In anaphoric reference, successful referent tracking and resolution in discourse is steered by this feature match between the anaphor and the antecedent. While agreement in anaphors to inanimate nouns like *die Region – ihr* (the region [fem.] – its [fem.]) is very straightforward as it is purely grammatical – here, the formal feature match will be referred to as *congruency* – anaphoric agreement to animate, human nouns can carry socio-semantic information of referent gender on top of the grammatical feature. The latter surfaces when referring to a child (neut. *Kind*) of known female or male gender using *sie* (she [fem.]) or *er* (he [masc.]) instead of *es* (it [neut.]) as a pronoun, cf. Birkenes and Fleischer (2022), albeit on Middle High German *kint*). This type of gender-based agreement will be referred to as *convergence* with or *divergence* from referent gender cues of the noun denotation, recurring to Thurmair's (2006) terminology to avoid irritation. If the source of gender agreement is determined on the basis of conceptual properties of the referent rather than formal characteristics of the noun, as in *der Sprintstar – ihre Freundinnen* (the sprint star [masc.] – her [fem.] female friends, cf. Oelkers, 1996), we refer to it as *referential gender* (following Dahl, 2000, 106). Note that grammatical categories are given in abbreviated form in brackets

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

(fem., masc., neut.), while the gender associated with a referent is spelled in full as in *gender-specific female*, or *gender-nonspecific*.

1.2. Gender variability in reference processing

Control of (the gender of) a target, like a pronominal element, implies a dependency relation of coreference: in language processing, features are not only compared but potential candidates are checked for coreference based on the matching or mismatching features imposed by the noun reflected on the pronoun (Ackerman, 2019, 12). Complexities arise because when the comprehension system disentangles this relation, the function of the agreement target, that is a possessive pronoun, can be conceptual pragmatic (Panther, 2009, 68) in cases like the grammatically neuter, conceptually female (*hybrid noun*) *das Mädchen – ihre Eltern* (the girl [neut.] – her [fem.] parents), or morphosyntactic as in *das Mädchen – seine Eltern* (the girl [neut.] – its [neut.] parents) (see Köpcke et al., 2010, 178, and Hübner, 2021a, 2021b). Yet, whether semantic convergence or grammatical consistency governs human reference agreement is inconclusive with dozens of studies assessing pronominalization of human anaphors in German: Oelkers (1996), Thurmair (2006), Panther (2009), Köpcke et al. (2010), Birkenes et al., 2014, Binanzer et al. (2022), and Hübner (2021a, 2021b), among others, investigated under which circumstances and to what extent human antecedents are referred to with (formally) congruent or (conceptually) convergent pronouns, each of them reporting that a gender-based (semantic or pragmatic) pronoun choice often outweighs grammatical cues.

Previous research on how animacy affects grammatical structures like gender agreement has illustrated how this property complicates anaphoric processing by comparing inanimate to animate references. In the context of this paper, we draw on a relatively simple, tripartite version of the *Animacy Hierarchy*: *human* > *animate* > *inanimate* (Siewierska, 2004, 46) and center the human–inanimate contrast. Animacy status of an antecedent has repeatedly been reported to modulate the effect of anaphora resolution and to determine the interpretation of different types of pronouns in German (Hammer et al., 2008; Fleischer, 2022, 277–280; Bader & Portele, 2025), manifesting that human possessors play a more distinct role for gender agreement phenomena. Most evidently, Köpcke et al. (2010), 177) illustrated how an animacy switch from an inanimate object to a metonymically denoted (male) human referent determined pronominal reference in *die 1. Geige – sein* (the first fiddle [fem.] – his [masc.]) / its [neut.]), whereas the instrument is feminine and grammatically only compatible with *ihr*. Dahl (2000), 99) therefore stipulated that '[grammatical] gender is one of the most obvious places where animacy shows up', rendering it the most fundamental value distinction of the grammatical gender categorization (Köpcke & Zubin, 1997, 112). Apparently, however, formal agreement in inanimate antecedents is not as irrefutable as one might expect under morphosyntactic rules: Fleischer (2022) registered corpus evidence of numerous incongruent possessive references to inanimate nouns conceived as 'gender-insensitivity' of the masculine and neutral pronoun *sein*, e.g., *Qualität hat seinen Preis* 'Quality (fem.) has its (masc./neut.) price'. This observation is backed up by the renowned German grammar Duden (2022), in which the incongruent possessive pronoun *sein* referring to feminine nouns is listed as a case of doubt, the usage of which is technically inaccurate yet common among (certain) speakers (see regional examples in Fleischer,

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

2022, 261). To explain the occasional ignorance of agreement constraints, possessives referring to inanimate possessors are attested as an instance of underspecification regarding gender features (Fleischer, 2022, 280–283) that ultimately license reference to feminine antecedents with *sein* where *ihr* would be prescriptively mandated.

Moreover, the agreement rule applied – grammatical or gender based - may be moderated by the distance between a pronoun and its antecedent noun, as identified in multiple studies (Binanzer et al., 2022; Czech, 2014, albeit on relative and personal pronouns; Hammer et al., 2008; Birkenes et al., 2014, 19; Hübner, 2021a, 13, on personal pronouns; Köpcke et al., 2010, 182–183, on relative and possessive pronouns; Panther, 2009, 78 on personal, relative and possessive pronouns; Birkenes & Fleischer, 2022, 254–257). These studies found a relationship between choice of the agreement pattern and linear syntactic distance such that the probability of exhibiting conceptual gender agreement grew proportionally with the distance between controller and target, meaning that the more constituents separated a pronoun from the noun, the less likely grammatically governed and the more likely conceptually motivated anaphoric pronominalization became. The way this correlation affects different kinds of targets is captured by the *Agreement Hierarchy*: *attributive* > *predicate* > *relative pronoun* > *personal pronoun* (Corbett, 2006, 206–237). The implicative hierarchy reliably predicts the (rising) likelihood by which a co-referring expression like a pronoun will be controlled by semantic (meaning based) rather than syntactic (formal-rule-based) agreement. While in this renowned *Agreement Hierarchy*, possessive pronouns did not explicitly find space, later work places them alongside personal pronouns (Köpcke et al., 2010, 182), demonstrating that they, too, have a strong tendency to follow the controller's referential gender in agreement relations. In fact, possessive pronouns have been shown to be particularly open for meaning-driven forms of agreement, that is, conceptual gender convergence (Birkenes et al., 2014, 12–13; Fleischer, 2022, 280; Hübner, 2021b, 42; Köpcke et al., 2010, 175, 182; Oelkers, 1996, 10; Panther, 2009, 76; Thurmair, 2006, 191, 199–200). The Duden grammar (2022, 434) notes gender-based pronoun choice particularly with somewhat longer distances, yet Panther (2009, 79–80) observed semantic being preferred over grammatical agreement for possessive pronouns in particular, even when both elements were in immediate adjacency.

1.3. The present research

The study focuses on grammatically feminine nouns to assess mismatch effects with the grammatically incongruent and potentially gender-divergent masculine/neuter pronoun *sein* instead of feminine *ihr* by comparing these anaphors and contrasting characteristics of nouns to which they refer. Our goals were (i) to expand the research on anaphoric reference to possessive pronouns and (ii) to experimentally test the observations of the aforementioned (Section 1.2) corpus data of inanimate agreement deviations (Fleischer, 2022) in an experimental setting.

For this purpose, we constructed sentences containing a grammatically feminine noun phrase (NP) referred to by a possessive pronoun that was either congruent (fem. *ihr*) or incongruent (masc./neut. *sein*) with the antecedent's gender (fem.). The sentences differed in noun types and the distance between the possessor and possessive pronoun. Specifically, we differentiated animacy by contrasting inanimate (*die Jahreszeit* [the season]) with human (*die Zugbegleitung* [the train attendant])

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

possessors. Human antecedents consisted of both gender-specific, typically female referents (*die Witwe* [the widow]) and gender-nonspecific epicene nouns (*die Küchenhilfe* [the kitchen help]). This division pertains to social gender information transmitted in human reference, which is not part of the study presented here but a subgroup investigation particularized in Schütze (accepted).

Focusing on the role of congruency, animacy and distance in anaphoric relations, the experiment aimed to address

a) how the grammatical features of the antecedent NP would affect sentence acceptability and agreement processing,
b) whether when judging the sentence and in the course of reading, agreement processing is sensitive to the conceptual properties (human/inanimate) of the antecedent and if so,
c) whether this depends on the syntactic position of the antecedent within the sentence (long/short distance).

From Ackerman's (2019, 13) proposal for strict matching in grammatical gender languages, we derive that incongruent sentences fail successful coreference dependency and should mostly be rejected, since φ-features such as person, number, gender of a pronoun and a candidate antecedent are not identical. The evidence on German possessive deviation cited above (Section 1.2), however, casts considerable doubt on the overall rejection expectation. Despite a possibly somewhat diffuse distinction between feminine and masculine/neuter pronouns in inanimate possessive anaphors (cf. Fleischer, 2022), the use of the incongruent form is a grammatical violation expected to be registered as such: We anticipate an agreement restriction reflected in acceptability declines and processing difficulties in the incongruent condition as German has mandatory agreement in gender markings of noun–pronoun anaphors. When the given gender cues are inconsistent with referent gender, we assume that such gender-incongruent possessive pronouns will produce greater processing difficulty compared to congruent ones (Irmen & Kurovskaja, 2010, 372). Above all, we expect deviating patterns for inanimate versus human nouns given the additional socio-semantic information that a gendered pronoun carries when referring to humans. In line with the *Animacy Hierarchy* (Section 1.2), violations in inanimate reference should have a weaker impact, that is higher acceptability and smaller increase in reading time (rt) compared to human referents, such that the availability of both syntactic and semantic agreement cues in human nouns to be checked for agreement may impact processing of the possessives manifested in increased processing costs as reflected in slowed down reading (Esaulova & von Stockhausen, 2022: 56) and reduced acceptability (Irmen & Kurovskaja, 2010; Osterhout & Mobley, 1995) for human possessors referred to with the incongruent pronoun, particularly for gender-specific antecedents (when associated referent gender was part of the denotation), as identified in recent work by Schütze (accepted).

According to the outlined previous findings, we hypothesize that a lack of syntactic and/or semantic fit between grammatical and referential gender, along with increased possessor–possessee distance, will lead to enhanced processing costs, causing rts to vary depending on sentence configuration. Conversely, an agreement violation effect may be less pronounced in inanimate than human anaphors when distance increases due to information decay over time (Frank, 2019, 92), such that the backward search for antecedents of anaphoric expressions (suggested by Garnham,

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

2001, 90) would be less elaborate in close pronoun–antecedent proximity, i.e., when less time elapses between activating and re-accessing a referent (Frank, 2019, 92), reducing retrieval effort and yielding facilitation (Niu & Liu, 2022).

We conducted a self-paced reading (SPR) study to assess whether these predictions are borne out in incremental language processing. Participants read sentences online on a computer screen and judged them as grammatically acceptable or not (Schütze, 2016). Following Schütze and Sprouse's (2014, 27–29) terminological notions, we reserve 'grammaticality' including the distinction between grammatical and ungrammatical, as an inherent property of prescriptive grammar, and use 'acceptability' in connection with a task-dependent, subjective judgment of perceived sentence well formedness. In this dual task paradigm, the offline measure of sentence judgment was combined with the online measure of sentence reading: Acceptability rates reflect a conscious and final decision; response times may reflect sentence-final processing difficulties, but rts are sensitive to interpretations during early comprehension in real time (Garnham, 2001, 61).

Despite extensive research on German gender agreement and animacy interactions (outlined in Section 1.2), there are still gaps in understanding if and how deviant pronouns are processed differently for human versus inanimate referents. To date, the comprehension of possessive pronouns referring to inanimate or human role nouns has not been investigated experimentally in German. Since there is variation of or choice between genders – as attested for personal, demonstrative and relative pronouns (in German, see Section 1.2; cross-linguistically, see Audring, 2008) – there is ample reason to expect variation for possessive pronouns, too (as observed by Köpcke et al., 2010; Oelkers, 1996; Thurmair, 2006).

2. Experimental study

In the experiment, we tested whether gender-incongruent possessive pronouns impact sentence acceptability, response time and rt, and if so, whether this depends on the possessor noun's animacy status and its distance from the anaphoric pronoun. The complete dataset, including stimuli, analysis script and model output, is available at https://osf.io/g9kbt/.

2.1. Method

The experiment used a combined *SPR* and sentence acceptability judgment (*AJ*) method similar to Irmen and Kurovskaja's (2010) procedure, in which sentences were presented phrase by phrase at participants' own pace (rather than word by word, improving the task by limiting spillover to later regions, cf. Mitchell, 1984: 74–76). In non-cumulative *SPR* as used in this study, a sentence starts with an initial word or phrase, and participants press a key to reveal each successive sentence fragment upon which the previous phrase disappears, until the entire sentence has been displayed. The time between two key presses indicates the rt of a phrase (e.g., Jegerski, 2014; Mitchell, 1984). Two of the major advantages of *SPR* are that (i) it provides a time-sensitive, precise measure of sentence reading and that (ii) it can be easily run online with many participants using personal devices. Sentence judgments were binary classifications instead of scalar or gradient responses (unlike Irmen and Kurovskaja's (2010) implementation, but see Bader and Häussler (2010) on how these measures relate).

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

The study was created with *Experiment Builder* in *GORILLA* and hosted on their platform (Anwyl-Irvine et al., 2020; Cauldron Science, 2022), a licensed program to combine questionnaires and experiments with reliable precision (Bridges et al., 2020) to be run online and distributed over the crowdsourcing web platform *Prolific* (Prolific, 2014; Peer et al., 2017) in August 2023. Data collection with these services was liable to charges.

2.1.1. Participants

We recruited 96 native speakers of German through the *Prolific* participant pool in 2022, who were not necessarily currently living in a German-speaking country. Thirty five women, 55 men, one participant identifying as diverse, and five participants with no indication of gender identity between 19 and 62 years (M_{Age} = 32.38, SD_{Age} = 10.0), representing various economic backgrounds, took part in the experiment and were compensated for participation.

2.1.2. Design

The experiment followed a balanced 2 (CONGRUENCY: congruent versus incongruent) × 2 (ANIMACY: inanimate versus human) × 2 (DISTANCE: short versus long) design. Participants were presented with each condition: By varying CONGRUENCY and DISTANCE configurations, we generated four different versions of each sentence per ANIMACY noun types, which each participant read in one condition. Four item lists resulted from this, to which participants were randomly and equally assigned.

Acceptability was manipulated via sentence grammaticality tied to the pronoun's gender inflection, which either agreed (fem. *Ihr* [her/its]) or disagreed (masc./neut. *Sein* [his/its]) formally with the preceding feminine controller noun. In the congruent versions (Table 1, 1–2 with *ihr*), the gender of the possessor and pronoun were in concord (feminine), leading to a felicitous sentence (Table 1). The incongruent reference with *sein* to inanimate antecedents (1) violated formal gender agreement, while the incongruent pronominalization of a human female referent (Section 2.2) with *sein* constituted both a grammatical and a semantic violation. Only the incongruent (*sein* pronoun) version of human epicene referents (Section 2.1) could possibly receive a conceptually plausible interpretation as converging with the referent gender.

To test the influence of the distance between antecedent noun and pronoun, an adverbial was inserted between possessor and possessee, which yielded (a) short (two-word) or (b) long (five-word) distance between these units. Therefore, the position of the critical phrase containing the congruent or incongruent pronoun varied in sentences with short (phrase 5) and long distance (phrase 6), cf. 1–2a versus 1–2b in Table 1. Sentence content remained unchanged, and sentence structure was held constant across all critical items. We further controlled sentence length, fitting all sentences into the scheme of seven phrases consisting of 10–15 words (see Table 1, phrases separated by vertical bars) and controlling the number of words per phrase across items (though we were less restrictive in the sentence-initial and -final fragments).

2.2. Materials

All selected nouns were definite, feminine and singular, which surfaced in the preceding feminine article *die* so that anaphoric (masc./neut) *sein* would lead to a feature mismatch in (at least) grammatical gender. Note that possessive reference to

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

Table 1. Example sentences for inanimate (1) and human (2) possessors in short (a) and long (b) distances between possessive pronoun and noun antecedent referred to with the congruent or incongruent pronoun

Conditions		Example sentence (with translation) under congruency conditions
Animacy status	Distance	
1. Inanimate	a. Short (2 words)	In diesem Jahr | will | **die Alpenregion** | im Juni | **ihr/sein** Jubiläum | mit großem Programm | nachfeiern.
	b. Long (5 words)	In diesem Jahr | will | **die Alpenregion** | im Juni | mit großem Programm | **ihr/sein** Jubiläum | nachfeiern.
		'This year, the$_{FM}$ alpine region$_{FEM}$ will celebrate its$_{FEM}$/its$_{MASC/NEUT}$ anniversary in June with a big program'.
2. Human	a. Short	Sicherheitshalber | hat | **die Kontaktperson** | für Notfälle | **ihre**/**seine** Telefonnummer | auf dem Formular | angegeben.
2.1 Gender nonspecific, epicene	b. Long	Sicherheitshalber | hat | **die Kontaktperson** | für Notfälle | auf dem Formular | **ihre**/**seine** Telefonnummer | angegeben.
		'As a precaution, the$_{FM}$ emergency contact person$_{FEM}$ has indicated her$_{FEM}$/his/its$_{MASC/NEUT}$ name on the form'.
2.2 Gender specific, female	a. Short	Am liebsten | möchte | **die Patentante** | im Winter | **ihre**/**seine** Urlaubstage | in den Bergen | verbringen.
	b. Long	Am liebsten | möchte | **die Patentante** | im Winter | in den Bergen | **ihre**/**seine** Urlaubstage | verbringen.
		'Preferably, the$_{FM}$ godmother$_{FEM}$ would like to spend her$_{FEM}$/his/its$_{MASC/NEUT}$ holidays in winter in the mountains'.

plural nouns in German, also preceded by *die*, invariably occurs as *ihre*, overlapping with the singular feminine form, which is why we were cautious to refrain from conceptually plural possessors (see Schütze [accepted] on this matter).

Inanimate nouns were mainly drawn from the various corpus examples found by Fleischer (2022). Human nouns were added in two subsets of equal size: The gender-specific, typically female-referring feminine nouns were kinship terms (*Patentante* (godmother)), fabulous or mythical creatures (*Hexe* [witch]) as well as occupations (*Hebamme* [midwife]), all of which carry the lexical gender feature [+female] without being overtly marked female-inflected denotations (i.e., through *–in*-suffixation as in *Lehrerin* (teacher, female) – except for the less transparent *Cousine* (cousin, female). The gender-nonspecific epicene nouns were mainly occupations (*Teamleitung* [team leader, gender-indifferent]), positions (*Majestät* [majesty]) and a few role nouns (*Geisel* [hostage]); for a similar distinction, see Kreiner et al. (2008: 247). Other constraints imposed on the possessive phrases – besides the manipulations outlined in Section 2.1.2 – included varied possessee noun gender (groups of fem. and masc./neut. nouns in close to parity to address potential issues of agreement attraction involving *ihr* or *sein*, respectively), number features (comparable proportions of singular and plural nouns) and word structure (both simple and compound nouns).

In total, participants evaluated 180 sentence items: 60 critical items containing possessives (30 inanimate, 30 human nominal referents, of which 15 epicenes and 15 female specific), 60 distractor items and 60 fillers. Filler sentences were well formed, whereas distractor sentences were equipped with grammatical inconsistencies of various kinds as well as 'cases of doubt' and borderline cases of ungrammaticality to obscure the study focus. The Duden dictionary of German cases of doubt (Duden, 2016) lists incongruent *sein* references to feminine antecedents as one instance of such cases, and we drew numerous examples from this rich source of

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

inspiration to create distractor sentences. See Supplemental Material (https://osf.io/g9kbt/) for further details on stimulus selection.

2.3. Procedure

At the start of the experiment, participants gave informed consent. They then received detailed instructions for the task: They would read sentences split up into phrases and could control the presentation speed themselves. Their first task was to rate the sentence by indicating whether they would accept the sentence as grammatically correct or not. Emphasis was put on the information that this judgment was not about the content of the sentence. The second task was explained as a test of participants' reading for comprehension that ensured they were engaged in attentive reading throughout, requiring them to respond to 'yes'/'no' questions about the content, with instant feedback (green tick mark or red cross). Participants were instructed to decide as quickly as possible and, if in doubt, rely on their initial intuition, as there would be a time limit. Instructions were followed by a practice block of five items to familiarize with the task procedure before proceeding to the main task. Before each new trial, a black fixation cross appeared in the middle of a gray background on the screen for 250 ms, followed by a 550 ms pause before sentence presentation of the initial to the last phrase, displayed upon the key press centered on the screen. Participants advanced through the seven sentence slides at their own pace by pressing the space bar to reveal phrases incrementally. The next key presses after sentence offset presented an acceptance prompt and, when applicable (after one third of the items), an additional comprehension question (CQ). The binary judgment of the sentence's acceptability was requested by a large '?' with the response options 'accept' or 'reject'. Within 5000 ms (small countdown icon shown on 3000 ms, prompting timely decisions before a failure to respond; for a similar duration, see De Vogelaer et al., 2020), participants had to come to a decision by pressing the defined keys. CQs were to be answered with 'yes' or 'no' using the same keys within 7000 ms, but these prompts never targeted the possessive dependency. Assignment of response keys was counterbalanced across participants to control for handedness, yielding two versions. When participants were ready for a new sentence, they pressed the space bar to begin an intertrial interval of 500 ms blank screen.

Items were presented in pseudo-randomized order and arranged in five blocks of 36 sentence items each (12 of each item type, critical items evenly split between inanimate and human referents). Short breaks were allowed in between blocks to rest or move briefly, but participants had been informed of a maximum completion time of 60 minutes. A progress bar was displayed on screen throughout. At the beginning and after half of the blocks, an attention check required pressing a nonresponse key. After completing all blocks, having read and rated all items, participants were asked for additional demographic information and debriefed for feedback on the experiment, particularly their assumptions about the study focus. To successfully finish participation and receive payment, they were redirected to *Prolific*. The entire experiment took 30–45 minutes on average.

2.4. Data analysis and preprocessing

The dependent variables of interest, i.e., the measurements to be analyzed, were (1) the proportion of participants' *AJ*(in percentages), (2) the response reaction time

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

(*RT*) needed to either accept or reject a sentence and (3) the *rt* of the critical phrase, that is, the sentence fragment containing the possessive (both in ms).

In this setting, the *rt* per phrase was measured as the duration it remained visible on screen before the key to proceed is pressed. Analyses were performed using the statistical software *R* (R Core Team, 2023) in *RStudio* (RStudio Team, 2024). Prior model comparisons relied on Akaike information criterion(AIC) values for goodness of fit (Gries, 2021, 366–370), the *step* function (*lme4* package, Bates et al., 2015) for stepwise model complexity comparisons and the *predictor strength* function of the *SfL* package (Schmitz & Esser, 2021). Numeric predictors (i.e., trial number, graphemic length and Zipf frequency of nouns, cf. Supplemental Material) were centered and z-scaled. Data manipulation applied *tidyverse*-style syntax (Wickham et al., 2019). Post-hoc comparisons of effects were conducted with *Tukey*-adjusted pairwise tests for multiple comparisons using the contrast function of the *emmeans* package (Lenth et al., 2024).

Using a fixed threshold, acceptability judgment responses faster than 200 ms and slower than 3500 ms were excluded from the analysis as these are prone to reflect inattentiveness, erroneous or unintentional key presses (Baayen & Milin, 2010, 15; Jegerski, 2014, 20). Similarly, only rts above 50 ms and below 1500 ms per phrase were included in data analyses. As a further step, outliers – reaction and rts exceeding 3 standard deviations (SDs) from individual means – were computed and discarded from the dataset (2.83% of response data and 2.43% of rt data; for data trimming, see Baayen & Milin, 2010, 15–17; Jegerski, 2014, 21; Whelan, 2008, 477). Due to right-skewed distributions of RT data, log-transformed reaction and rts were analyzed instead (Baayen & Milin, 2010). We operationalized participant attentiveness as the ratio of *CQ* accuracy, retaining only datasets from participants who scored 75% or more correct, which all participants achieved (M_{CQ} = 93.6%, SD_{CQ} = 0.24, $M_{RT.\ CQ}$ = 1688 ms, $SD_{RT.CQ}$ = 740). Besides this accuracy check for eligibility, strategic non-reflective response behavior was screened to ensure that the level of acceptability was neither 0% nor 100% throughout the experiment, verifying no participant consistently accepted or rejected sentences irrespective of stimulus type, that is grammatical inconsistencies, or inattentively. Since participants are referred to *Prolific* upon successful completion, any incomplete participations were already excluded from the final dataset on the platform.

3. Results

3.1. Acceptability judgments

First, we report results of the analyzed proportion and speed of sentence acceptability as judged by the participants, that is, responses and RTs for the critical sentences with possessives (*ihr*/*sein*).

3.1.1. Rating responses

We fitted a generalized mixed-effects model using *glmer* with a logit link (Bates et al., 2015) to predict the binary decision of sentence acceptability, i.e., response (to the *AJ* task) with response as the dependent variable and an interaction of the factors CONGRUENCY, DISTANCE and ANIMACY, allowing for random intercepts for trial to account for adapted response behavior in the course of the session, as well as for the

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

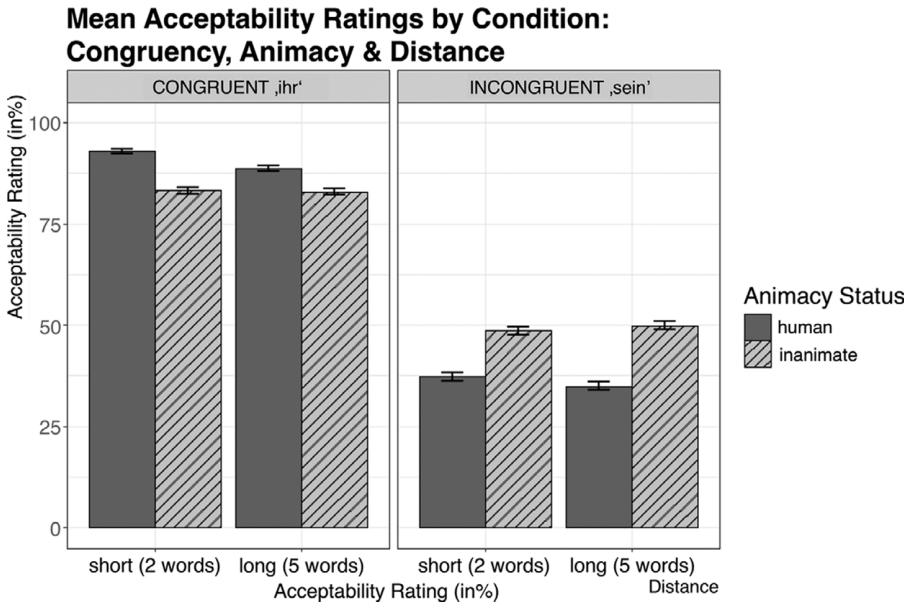

Figure 1. Mean acceptability ratings (in %) for congruent (*ihr*) and incongruent (*sein*) sentences with human (no pattern) or inanimate (striped) feminine antecedents; error bars represent SD.

different noun items and for subjects to account for any individual sensitivity to the agreement phenomenon studied here.[1]

Overall, the sentences with possessive anaphoric pronouns had a mean acceptability rating of 64.9% ($SD = 0.48$). However, sentences with congruent *ihr* anaphors were judged twice as acceptable ($M = 86.9\%$, $SD = 0.34$) as incongruent *sein* pronominalizations ($M = 42.7\%$, $SD = 0.50$) as depicted in Figure 1; this negative effect of incongruency on acceptability ratings being highly significant ($\beta = -3.49$, $SE = 0.18$, $z = -19.82$, $p < .001$). Surprisingly, even the grammatically congruent inanimate items ($M = 83.0\%$, $SD = 0.38$) were judged as 8% less acceptable than human anaphors ($M = 90.8\%$, $SD = 0.29$, depicted in the left panel of Figure 1; $\beta = -1.07$, $SE = 0.22$, $z = -4.83$, $p < .001$). Looking further into the ANIMACY distinction, a significant interaction between incongruent and inanimate conditions ($\beta = 1.63$, $SE = 0.22$, $z = 7.42$, $p < .001$) revealed a reverse pattern with incongruent possessive pronouns as there was a clear dominance of accepted *sein* references to inanimate ($M = 49.2\%$, $SD = 0.50$) over human referents ($M = 36.2\%$, $SD = 0.48$, a difference visible in the right panel of Figure 1), as confirmed in post-hoc tests ($est = 0.53$, $SE = 0.80$, $z = -4.23$, $p < .001$). Linear syntactic DISTANCE between noun antecedent and pronoun had a significant impact on sentence acceptability: A shorter distance between possessor noun and possessive pronoun was found to be slightly more acceptable ($M = 65.6\%$, $SD = 0.48$) than longer distance between these anaphoric units ($M = 64.2\%$, $SD = 0.48$; $\beta = -0.55$, SE = 0.20, $z = -2.80$, $p = .005$). Interestingly, this became more pronounced when taking the animacy status of the possessive reference into account: Resolving the significant interaction between

[1]Model term: *Response ~ Congruency * Animacy * Distance + (1|Item) + (1|Subject) + (1|Trial).*

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

ANIMACY (inanimate) and DISTANCE (β = 0.52, SE = 0.25, z = 2.13, p = 0.03), the negative impact of inanimacy on the judgment responses became less severe as distance increased, but this trend was not supported in post-hoc comparisons where the acceptability rates were comparable (n.s.) for inanimate nouns irrespective of distance. On the contrary, human antecedents were accepted significantly more often in shorter (M = 65.2%, SD = 0.48) than longer distances (M = 61.9%, SD = 0.49); the improved sentence acceptability was also reflected in post-hoc contrasts (*est* = 0.32, SE = 0.11, z = 2.84, p = 0.025).

For sentence AJs, incongruent *sein* was by far less accepted than congruent *ihr*, but asymmetries were observed between human and inanimate possessors referred to with the incongruent possessive pronoun. Taken together, agreement violations of possessive references were tolerated more for inanimate than human nouns. In contrast, incongruent possessive pronouns referring to human nouns, especially at a longer distance, reduced sentence acceptability profoundly (Figure 1).

3.1.2. Reaction times

When analyzing RTs for *AJ*, the sentence-final decision as (non-)acceptable is likely to influence the time to come to this decision. A linear mixed-effects model was fitted using *lme4* (Bates et al., 2015), with log-transformed *RT* (*logRT*) as the dependent variable to predict *RT* of the acceptability response as a function of the conditions – similar to the *AJ* analysis – modeling the interaction of CONGRUENCY, ANIMACY, DISTANCE and the RESPONSE to the *AJ* while accounting for random intercepts as before.[2]

Overall, RESPONSE to the AJ had a significant influence as decision times decreased when a sentence was accepted (M = 687 ms, SD = 408) compared to the longer *RT* when a sentence was rejected (M = 749 ms, SD = 450; β = 0.17, SE = 0.06, t = 2.70, p = .007). However, this effect was reversed in interaction with CONGRUENCY (or rather lack thereof), when references were incongruent instead of congruent: Rejecting a sentence with an incongruent *sein* pronoun led to faster response times (M = 716 ms, SD = 424) than for a congruent *ihr* pronoun (M = 891 ms, SD = 525; β = −0.21, SE = 0.07, t = −2.88 p = .004), whereas the decision to accept resulted in amplified *RTs* for incongruent (M = 706 ms, SD = 408) compared to congruent anaphors (M = 678 ms, SD = 408), an observation found to be significant in post-hoc tests (*est* = −0.05, SE = 0.02, z = −2.9, p = 0.02). Unsurprisingly, participants were faster with '"match' decisions (accepting congruency, rejecting incongruency) than in 'mismatch' cases. There were no main effects of CONGRUENCY, ANIMACY or DISTANCE on response times.

To sum up, rejection responses required longer decision times than judging sentences as acceptable, especially so for the grammatically correct, congruent pronouns (Figure 2).

3.2. Self-paced reading

Next, we report results of the analysis of reading speed as a measure of participants' progress through the anaphoric unit containing the possessive pronoun and, thereby, a proxy to referential processing. For *rt* analyses, we kept both outcomes of *AJ* responses since the decision (whether to approve or dismiss) can affect how the sentence is read,

[2]Model term: *logRRT ~ Congruency * Animacy * Distance * Response + (1|Item) + (1|Subject) + (1|Trial).*

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

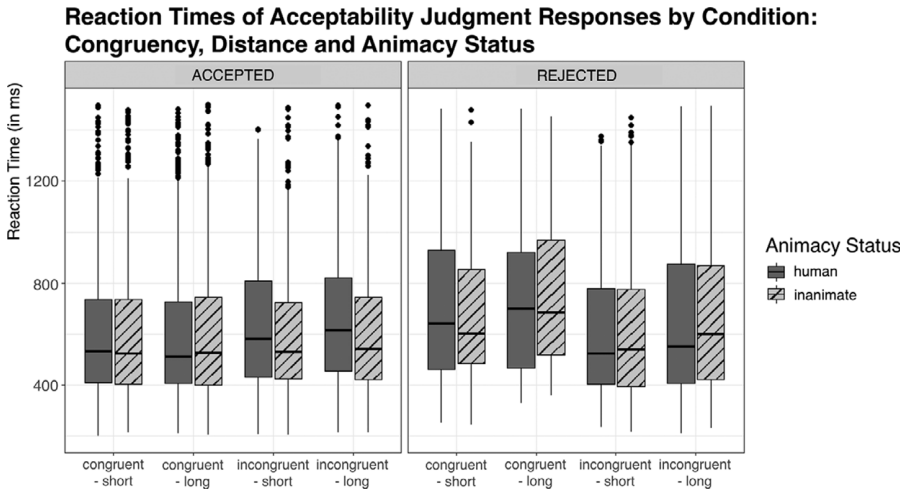

Figure 2. Mean RTs (in ms) for congruent (*ihr*) and incongruent (*sein*) sentences in short and long distances with human (no pattern) or inanimate (striped) feminine antecedents for sentence approvals (left panel) and rejections (right panel).

that is, due to early intuitions about pronoun agreement (e.g., double mismatches like in *die Nonne – sein* [the nun – his]) or because the decision might be forming while or after reading the entire sentence to resolve the reference successfully (early versus late commitment, discussed in Stewart et al., 2007 and below in Section 4.1).

3.2.1. Reading times

Mean rts for post-antecedent NP regions are shown in Figure 3; the critical phrase containing the possessive pronoun is framed in a grey shade. As is immediately evident, both human and inanimate sentences with incongruent possessives (dashed lines with triangles) generally produced large, immediate effects strongly localized to the pronoun (with some delay also spilling over one phrase further downstream in the short distance condition, cf. left panel of Figure 3), compared to congruent sentences (solid lines connected with dots), which do not. rt differences at the possessive phrase (Figure 4) illustrate how incongruent anaphors clearly prolonged continuing to the next phrase, that is, reading the subsequent phrase (next two to five words after the critical possessor noun), both in short and long distance versions.

A linear mixed-effects model was fitted to predict log-transformed rts (*logrt*) using *lme4* (Bates et al., 2015), with an interaction of CONGRUENCY and ANIMACY, DISTANCE (namely pronoun position) and as a function of the RESPONSE (to *AJ* task, 'accept'/'reject'), trial as well as item-specific measures influencing rts (Zipf frequency and graphemic length of possessee nouns) and random intercepts for nouns (items) and subjects. The model further allowed for individual differences in *AJ* response by participant with random slopes to account for the varying judgment behavior, i.e., the observed group differences among participants (elaborated in Section 4.4).[3]

[3]Model term: *logrt ~ Congruency * Animacy * Distance + Response + Frequency + (1|Graphemic Length) + (1|Item) + (1|Trial) + (Response|Subject).*

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

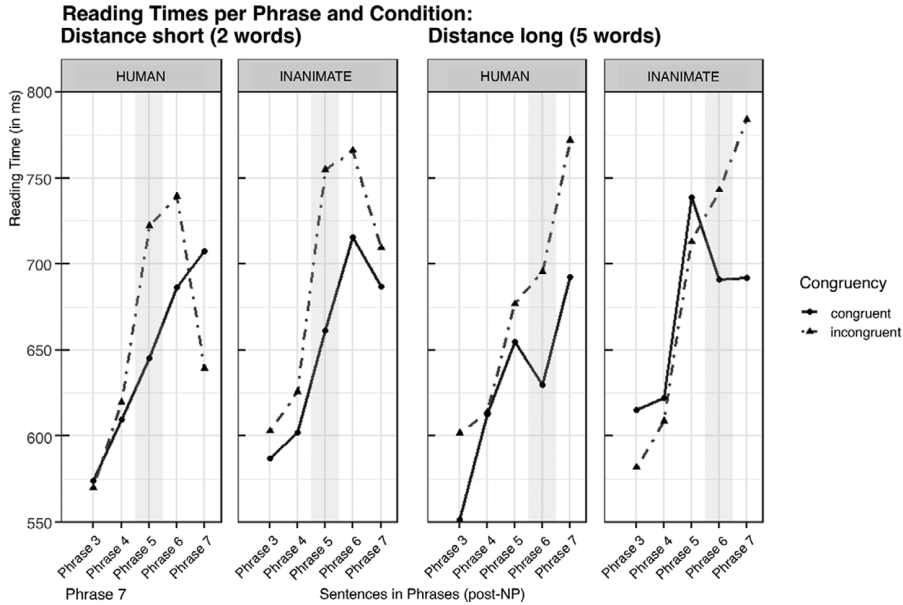

Figure 3. Mean reading times following a human (left side of the panels) or an inanimate antecedent (right side of the panels) through sentences with a short (left panel) or long distance (right panel) between the noun and the anaphoric pronoun. The possessive phrase containing congruent *ihr* or incongruent *sein* pronouns is shaded in grey. Dots indicate the congruent condition; triangles the incongruent.

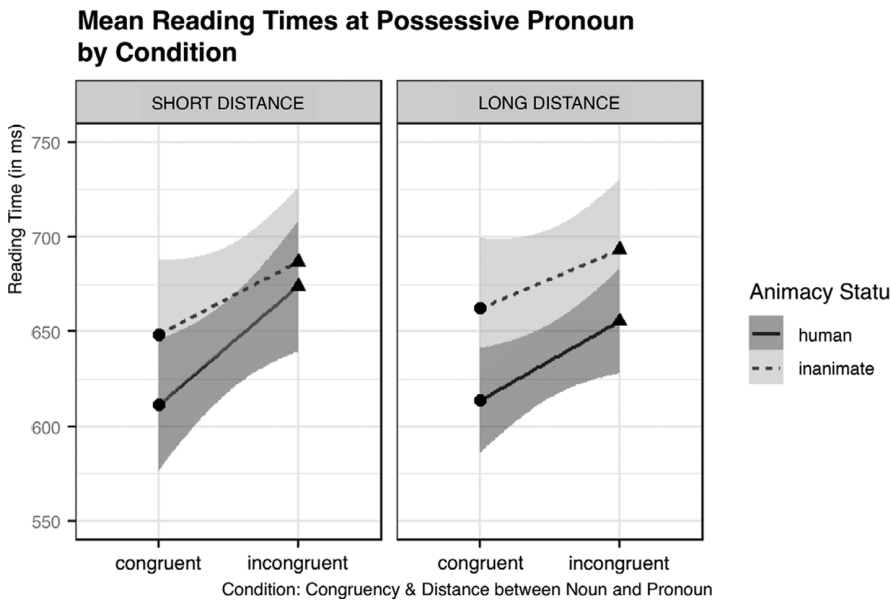

Figure 4. Mean reading times at the possessive phrase containing congruent *ihr* or incongruent *sein* pronouns with human (solid line) or inanimate (dashed line) antecedents in short (left panel) or long distance (right panel) between noun and anaphoric pronoun. Dots indicate the congruent condition; triangles the incongruent; shades represent 95% CIs based on the standard error.

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

Within this model, a main effect of CONGRUENCY emerged signaling that incongruent possessive pronoun phrases were read more slowly than congruent ones (M = 632 ms, SD = 291 versus M = 675 ms, SD = 323), the effect approaching significance (β = 0.04, SE = 0.02, t = 2.02, p = 0.04). Subsequent post-hoc tests for interactions showed that the slowdown imposed by incongruent *sein* pronouns was significant for both human and inanimate antecedents (more than 50 ms increase from congruent human (M = 611 ms, SD = 272) to incongruent human anaphors (M = 663 ms, SD = 310; *est* = −0.04, SE = 0.02, t = −2.61, p = 0.04); inanimates had a smaller difference of 34 ms between congruent (M = 653 ms, SD = 308) versus incongruent sentences (M = 687 ms, SD = 335); *est* = −0.04 for, SE = 0.01, t = −2.67, p = 0.03)).

In the short distance condition, when possessor and possessee were two words apart, incongruent anaphors slowed down rts at the possessive phrase by nearly 50 ms (M = 677 ms, SD = 324) compared to congruent ones (M = 628 ms, SD = 294), and post-hoc tests point at a significant CONGRUENCY difference therein (*est* = −0.06, SE = 0.02, t = −3.47, p = .003), whereas no effect was found in the long distance. A distance effect occurred only for inanimate, not for human, anaphors. For the former, rts were faster with two words than five words before the possessive pronoun, the post-hoc contrast tending toward marginal significance (*est* = −0.04, SE = 0.013, t = −2.67, p = 0.04). Moreover, post-hoc comparisons for interactions with the given RESPONSE revealed that whenever a sentence was subsequently rejected, rts were significantly higher (for 'accepted' cases: M = 630 ms, SD = 281, for 'rejected' ones: M = 697 ms, SD = 348; *est* = −0.04, SE = 0.013, t = −3.275, p = .001), hinting at the correlation between the outcome of sentence judgments and reading speed, which can either be interpreted as being hampered by earlier processes during reading or as a penalty for sentences with an augmented processing cost.

4. Discussion

While corpora can 'record grammatical knowledge as realized in language production', judgments provide intuitive information about such knowledge while rts tap into parsing processes during language comprehension in real time; they thus 'complement each other' methodologically (Myers, 2009, 413). What has been found in corpus data by Fleischer (2022) can be confirmed experimentally too: There is unexpectedly much variability in the pronominal gender assignment of German possessive constructions. Audring and Booij (2009, 13) interpret this as speakers' 'uncertainty', 'a source of language variation and change'. From our perspective, a certain diffusion is to be restricted to the grammatical agreement of pronouns with inanimate possessors. The variability exhibited by human possessors, we believe, is rooted in the socio-semantic agreement with referent gender indexed by pronouns. While some cases may be borderline, and some speakers may have less certainty about linguistic norms, based on our evidence, we are inclined to follow the claims that a mental representation of possessive relations cannot be built from grammatical structures alone, but draws on extra-linguistic knowledge and discourse information, which has formed cognitive accounts of situational models (Zwaan et al., 1998) or the mental model theory (Garnham, 2001).

From the results, we infer that readers attend to both syntactic and socio-semantic cues, i.e., rely on grammatical rules and real-world knowledge about likely referents.

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

Processing involves a combination of both (Cacciari et al., 1997; Caffarra et al., 2014; Kennison, 2003), because in the reading and judgment experiment, referent characteristics (human versus inanimate) and agreement (congruent versus incongruent) have led to asymmetrical effects.

Our experiment pinpointed slowdowns on the specific part of the sentence containing the possessive, and the incremental method used – albeit admittedly a more artificial reading than, for example, full sentence presentation – thus qualifies as a reliable experimental presentation technique.

4.1. Possessive agreement

In each of the analyses, pronoun agreement significantly affected the outcome and speed of the sentence judgment decision, as well as the reading of possessive anaphors. By contrasting grammatically congruent *ihr* with incongruent *sein* pronouns referring to feminine antecedent nouns, we were able to confirm that readers do detect these feature mismatches (Carreiras et al., 1996; Osterhout & Mobley, 1995) but nevertheless may come to an acceptability decision of such sentences in German. We further found differential incongruity effects because violating the gender feature of the possessive pronoun was less severe for inanimate nouns and in long antecedent–pronoun distance than for human nouns and in short distance. The factorial design helped disentangle what supports and what hinders tolerance of a possessive pronoun that does not agree with its antecedent gender features: Evidently, processing difficulty for a referring expression such as an anaphoric possessive pronoun seems to be sensitive to the animacy status of the antecedent and, under specific circumstances, the distance between the possessive relation, which will be addressed in the next sections, respectively.

Although we observed a clear acceptability advantage and acceleration for congruent as compared to incongruent pronominalizations, grammatical gender features appear less binding than expected. These results can be taken as evidence for a 'blurred' gender distinction (De Vogelaer et al., 2020, 288; similarly discussed in Fleischer, 2022, 280–283) in German, in which *sein* assumably acts as the unmarked, underspecified form (Fleischer, 2022, 283; Oppenrieder & Thurmair, 2002, 119). Importantly, 'underspecified' does not mean 'completely unresolved' (Karimi & Ferreira, 2016, 1014): Even though their acceptability was surprisingly high, incongruent possessive pronouns clearly magnified rts and thus hampered online comprehension. Referential processing was not delayed until the decision-making with the task at hand; rather competition occurred at the pronoun during reading and was potentially revised when judging the sentences. Given the missing effects of ANIMACY in rts but not AJs, we infer that the final pronominal integration of possessives can be delayed until the complete sentence has been read (Garnham, 2001, 83), after which filters for candidates of the best fit can be applied (Chow et al., 2014), resulting in approval or disapproval. Interestingly, some of the factors we investigated – aside from congruency – affected only AJs, that is, the conscious decision after having read the sentence, but not on the immediate measures during reading. This finding alludes to a late commitment to reference relations when all available information has been parsed (in full support of Stewart et al., 2007).

We have unveiled that even in a grammatical gender language with a pervasive morphological marking like German, in an anaphoric sentence, the constraint

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

imposed by a morphosyntactic feature alone is neither 'primarily relevant' (Ackerman, 2019, 9) nor does it suffice to provide unambiguous cues as to pronoun gender, or else incongruent *sein* anaphors referring to feminine nouns should have been disapproved entirely, in particular for inanimate antecedents for which agreement should be completely formal.

4.2. Animacy-based sources of gender variation

Animacy is deemed a pivotal distinctive trait in language, resulting in 'morphological splits based on animacy and humanness' (Ortmann, 1998, 83). Indeed, our study revealed the influence of antecedent animacy on sentence acceptability. Ratings of sentences with incongruent *sein* instead of congruent *ihr* anaphors were lower overall, but the frequent insensitivity to the grammatical violation in inanimates was striking. As expected, these appeared far more acceptable under incongruent *sein* reference than human antecedents.

There is evidence on how inanimate entities are more often referred to with neutral (i.e., it), less specific or default forms, rather than specific pronouns (Sorlin & Gardelle, 2018), which could have boosted the acceptability of masculine/neuter *sein* for inanimates, but due to the syncretism, this cannot be definitely determined. At the same time, referring back to previously activated inanimate possessors was associated with costs, observable in the considerable increase in rts. Together, this suggests a restriction of possessive constructions, even when used grammatically correctly, which leads us to the impression that prototypical possessors are animate; inanimate possessors, though grammatically absolutely viable, are obviously much less common in natural language. Two assumptions can explain the observed asymmetries depending on the animacy status of the possessor noun: Either the use of 'possessive' pronouns may be unusual with inanimate nouns (consistent with prior findings of fewer productions of pronouns for inanimate referents, see Exp. 1 from Fukumura & van Gompel, 2011, as well as with research on variation of genitive attributes with inanimates, see Kopf & Bildhauer, 2024), or the gender feature loses activation much faster in nonhuman contexts, making it more difficult to maintain hence more effortful to reactivate the discourse reference.

The latter intriguing line of thought, aligning with our findings, draws on a cognitive account of activation and prominence. Anaphoric reference as a coherence device requires to i) create a mental discourse representation in which two entities must be related meaningfully, and ii) keep the introduced referent(s) activated until the coreference with further expressions has successfully been established (Garnham, 2001). It has been argued that cognitive activation, perceptual prominence and discursive relevance are greater for human entities as they are most likely agents of events (Bader et al., 2023, 669), carry language-external gender-based meanings (Kalverkämper, 1979, 63 cited in Leiss, 1997, 326) and are more concrete (thus easier to visualize) given their identifying characteristics (such as [+human], [−young], [+teaching] for *Lehrkraft* [teaching person]), which makes personal nouns more relevant to everyday life of us as social beings. In contrast, referential processes involving inanimates underlie syntactic automatisms (Kalverkämper, 1979, 63 in Leiss, 1997, 326), hence receiving less memory activation (Fukumura & van Gompel, 2011), ultimately leading to declining levels of activation of an inanimate controller in language comprehension (Köpcke & Zubin, 2009: 142) that come with lower salience (Hammer et al., 2008) namely lower cognitive prominence and, therefore, accelerated decay (Frank, 2019: 103). Such less enhanced

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

processing depth has also been described as shallow processing or a shallow commitment to the initially considered referent (Karimi & Ferreira, 2016: 1024). Shallow processing of inanimate possessives, however, is challenged by our findings as incongruent inanimate references were not processed at the same speed as those with congruent pronouns. Furthermore, the rt penalty for inanimate referred to with *sein* conflicts with an account of gender insensitivity (Fleischer, 2022). Likewise, this selected gender insensitivity phenomenon is not a case of a 'good enough' representation (Karimi & Ferreira, 2016) either, since rts were clearly negatively affected by the deviating pronoun referring to inanimate nouns regardless of a later positive judgment of a sentence with incongruent *sein.* A more heretic interpretation of the high incongruency acceptability results overall is to claim that grammatical gender information is simply not relevant enough to affect comprehension. If that were the case, however, *AJ* results among human and inanimate anaphors should be comparable, which is not corroborated. To explain this discrepancy, another, additional, notion of grammatical gender must be assumed – a function it takes over on top of syntactic agreement, which leads us to aspects of social meaning in the human domain. Reconsider that the nature of third-person pronouns lies in the identification of an antecedent indicated by the pronoun. When comparing sentences with a potentially gendered subject (human nouns, e.g., *Stellvertretung* [deputy]) with those that have purely grammatical gender (inanimate nouns, e.g., *Jahreszeit* [season]), we essentially compare whether conceptual relevance complements grammatical information or not.

Among human referents, some candidates may arrive at a better socio-semantic fit with the masculine/neuter *sein* pronoun than others because of a referential gender convergence whereby activated gendered knowledge or inferred stereotypical associations (Irmen et al., 2010; Misersky et al., 2019) can override syntactic cues (Cacciari et al., 1997; Caffarra et al., 2014). Although the differentiation between epicene and typically female referents within human possessors could in principle affect the overall influence of the factor animacy with respect to human antecedents; the results confirm that the animacy effect was robust. We know from the subgroup analysis that incongruent possessive pronouns with epicenes are more accepted, suggesting elevated *sein*-acceptance under human reference, but not completely discarded for female referents either. Comparing gender-nonspecific to gender-specific human reference in the in-depth analysis it deserves goes beyond the scope of this article, but is covered by Schütze (accepted) in broader detail.

The gender of possessives may thus function as a (partly negligible) grammatical category in inanimate anaphors and as an additional gender-classifying feature in animates, especially human nouns. Given the different ways to establish agreement, different reasons for pronoun variation might collude as an interplay of multiple factors in reference processing. Whether these engage different cognitive processes, whether the source of irritation is grammatical or semantic or both, and whether processes differ for each type of antecedent such that different (timings of) components are at work (as reported by Hammer et al., 2005, 236) remains a question open for future investigation.

4.3. Dependency distance

In our study, an increase of intervening words from two to five between the possessor and possessive pronoun resulted in a small yet significant decline in acceptability

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

ratings. Still, the role of linear distance between an antecedent and its pronoun remains somewhat vague: Against our predictions, distance did not affect acceptability ratings for inanimate or human nouns as expected. In both cases, sentence judgments were not more positive when the distance was longer. This outcome can be attributed to the amount of the intervening material chosen for the present experiment, i.e., two versus five words, for which we opted to leverage German word order flexibility while preserving adequate sentence length for a reading task already challenging the comprehension system. In this exact range, Panther (2009, 80) reported small, if any, differences in pronominal references (in conceptual rather than grammatical agreement with *Mädchen* (girl [neut.]) – in stark contrast to the clear increase from zero to two words distance to antecedents. Kennison (2003) noted how pronouns embedded 'in more complex syntactic structures were processed more slowly', and that possessive pronouns might be processed more slowly in general compared to the more frequent personal pronouns, inferring that different syntactic configurations may come with a different salience or expectancy of a possessive. For some items in the present experiment, the word order resulting from greater distance configurations, farther separating the subject from the object, is less typical for German than for other items. In this concern, the grammatical structure of the stimulus sentences used here may have affected anaphoric processing too.

4.4. Further factors and avenues

With the present experiment, it became evident that among German speakers, 'a remarkable tolerance is observed toward combinations that are rare in spontaneous production' (De Vogelaer et al., 2020, 291, though on Dutch data). The strict matching criterion dictates a rejection of a coreference dependency when features do not match and obey the 'identity relation' between φ-features (Ackerman, 2019, 13, 14, cf. Section 1.3). In this respect, and in the context of the pronominal anaphors examined here, our findings do not conclusively support the claim that German consistently and solely 'access[es] only the [grammatical] feature[s] during coreference resolution' (Ackerman, 2019, 19). Below, we discuss an alternative explanation to the unfulfilled rejection hypothesis, namely that readers have overly and readily accepted sentences after diverting to a non-coreferential interpretation, in which feature identity would no longer be necessary.

Given our design – obstructing regressions to previous words and using isolated single-referent sentences with limited context – one could argue that the possessive pronoun's deictic function to indicate discourse units may not necessarily rule out other antecedents if the gender feature does not match that of the given candidate(s). Instead, if a pronoun cannot be linked to – and cannot be reinspected for (in) congruency with – the only explicitly mentioned antecedent in the presence of nonmatching features, they may further trigger antecedent search processes as a repair strategy to establish coreference to some other discourse referent than the possessor when the parser encounters a lack of satisfactory candidates (Chow et al., 2014). This offers another path to resolve a gender-deviant pronoun: 'as coreferential with the available antecedent and ungrammatical, or as linking to an unmentioned referent and grammatical', a coping mechanism which has been attested for 'particularly skilled readers, [who] may come up with an additional, unmentioned referent' (Piepers & Redl, 2018, 98). Rather than relying solely on the gender information cued

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

by the pronoun in an attempt to parse an ungrammatical construction, an alternative explanation is that participants may read the sentences without successfully establishing a coreference relation between the pronoun and the given antecedent (as proposed by van Gompel & Liversedge, 2003: 7). In fact, shifting to a sentence or discourse external, competing antecedent to integrate an unexpected pronoun could explain the frequent acceptance of incongruent sentences despite the formal agreement violation as well as the amplified rts at the possessive phrase, as this strategy would be costly (Chow et al., 2014, 3). However, such interpretation would have to account for the observed animacy effect. Moreover, Karimi and Ferreira (2016, 1017) note that, even if feature mismatches pose a risk to impede successful reference resolution, the referent of a referring expression is still reliably identified when prompted to read for comprehension – similar to our design.

When investigating possessive references, one might conceive some referents to be more likely to '"possess' something, i.e., as more prototypical possessors, like we advocated above (Section 4.2). While this is certainly true for human referents, it could also pertain to inanimate referents that are concrete objects rather than abstract concepts, yielding functional differences in processing (Martín-Loeches et al., 2001; also, note that the version of the *Animacy Hierarchy* proposed by Corbett (2006: 185) is subdivided accordingly: *human* > *other animate* > *concrete inanimate* > *abstract inanimate*). In the context of the type of semantic relation between the in−/animate possessor and the possessed, the antecedent's conceptual role (ownership, agent- or patient-like status) may interact with patterns of variation, too. In addition, the definiteness of the antecedent—here, preceded by the feminine definite article *die* (the [fem.]) – could be driving gendered perceptions (Imai, 2014). Notably, fixed and conventionalized expressions like *etwas hat seinen Reiz/Zweck/Grund* (something has its [masc./neut.] charm/purpose/reason) may presumably be stored as constructions and hence accepted more readily due to their frequent co-occurrence (Stefanowitsch, 2011, but see Fleischer, 2022, 264–267 for discussion). However, since neither of these post-hoc explored factors was an experimental criterion, meaning items were not controlled accordingly, further work is desirable to assess their impact and advance the understanding of *sein*-possessives in feminine references. Furthermore, although possessee gender was not explicitly tested, its approximate balance across feminine and masculine/neuter forms in the stimulus set (alongside possessee number) mitigated potential impact on the present findings. Given that prior corpus research (Fleischer, 2022: 275–276) found no evidence for gender- or number-driven agreement attraction with possessives, these features were not treated as variables in the present design, and no specific analyses were conducted.

Subjective ratings like AJs can vary for many reasons besides the grammaticality of a sentence. Although one third of distractors with grammatical inconsistencies and idiosyncrasies to counter the conspicuousness (Section 2.2) had been added as a 'red herring' to lay false trails, and despite a thorough randomization, one out of six subjects (16 of 96, 16.7%) indicated the experimental focus on gender and (pro)nouns in the post-experimental debriefing. Following Osterhout and Mobley (1995), who detected a similar group effect in AJs, we discovered that their performance based on the overall mean acceptability of incongruent items was affected by group, regardless of antecedent animacy, which is indicative of a general agreement sensitivity. We therefore assume that the extent to which someone is responsive to grammatical gender saliency might differ on an individual level (alongside individual differences in reading strategies; see Mitchell (1984), and Garnham (2001: 80), where the pattern

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

of results was largely confined to those who read more slowly in the experiments, suggesting that strategic processing could be decisive). Deeming individual variation an essential characteristic in language perception and comprehension research, we did not exclude these datasets. In fact, by-subject differences in dealing with anaphora have previously been reported, e.g., preference groups in Hübner (2021a): 14) that exclusively adhered to grammatical or pragmatic agreement patterns. Future experiments could focus on how real-time comprehension strategies of referential relations vary systematically across subjects; here, we addressed this issue through the random effect structure for participants (Sections 3.1–3.3). Stimulus repetition (pronominal incongruencies with *sein*) and a forced choice decision task may have also upweighted structural cues and may have highlighted the violation more prominently than a mere content question or probe may not have (Leeser et al., 2011).

Finally, besides individual (dis)approval of referential mismatches, an influence of participant age on preferences in referential agreement type has been reported by Oelkers (1996: 11–12), Shake and Stine-Morrow (2011), Foley and Ahn (2024), and Steriopolo and Schütze (2025). When inspecting generational effects on sentence judgments, Schütze (accepted) showed that age indeed emerged as an influential factor, with higher acceptability rates of incongruent possessive pronouns by younger than older individuals. The question of whether younger participants are more open, flexible, and even innovative in using gender-deviant pronouns or adults in later life stages are more resistant to grammatical deviations (in general, or in pronominal reference specifically) is unresolved at this point (see Eckert, 1998, for discussion). Notably, age-related variability in anaphora resolution, retrieval and interpretation is also tied to cognitive abilities, such as working memory capacity (Karimi & Ferreira, 2016, 1033–1034), as well as gender-sensitive language development. On an individual level, evidence suggests that readers with better language processing skills are generally more sensitive to formal referential ambiguity than less proficient ones (Nieuwland & van Berkum, 2006). While we accounted for subject variation in the analyses, measuring these skills could be highly informative.

5. Conclusion

A two-fold task structure involving SPR and judgment of German sentences explored the acceptability and processing of possessive anaphoric pronouns that do not align with the grammatical gender of their referents (grammatically masculine/neuter *sein* instead of feminine *ihr* for feminine antecedents). Our manipulations introduced a gender violation in inanimate and human references and thereby created a potential conflict on formal and/or conceptual grounds.

We found that native readers of German were sensitive to gender agreement, reflected in a decrease of acceptability ratings and an increase of rts at the site of the pronominal violation. Effects varied by antecedent animacy, with different consequences for inanimate versus human antecedents, confirming that both linguistic structure and conceptual factors influence language processing.

A central implication is that research on gender agreement in languages with gender systems must take the type of reference (grammatical or gender based) into account and must not neglect the difference nor the overlaps between formal and conceptual gender features that guide agreement patterns. By and large, the referential inconsistencies induced by the pronoun led to markedly distinct effects largely

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

depending on the animacy property of the possessor, which elucidates one source of gender conflicts in real-time reading comprehension.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/langcog.2025.10032.

Data availability statement. The complete dataset, containing preprocessed data files, an R script for statistical analyses and the stimulus set, is stored in the corresponding Open Science Framework repository (https://osf.io/g9kbt/).

Acknowledgments. We wish to thank the anonymous reviewers for their valuable suggestions and constructive feedback. We are also very thankful to Elisa Schmied for the invaluable assistance in stimulus design and control and to Ronja Eberhard and Anne Heinzmann for their support in creating stimulus materials. We are grateful to Simon Kasper for sharing materials on German language varieties (REDE project). Many thanks to Sol Lago and the psycholinguistics colloquium at Frankfurt University for fruitful discussions on this topic.

Funding statement. This work was funded by a research grant to the first author as part of the Research Training Group 2700: 'Dynamics and Stability of Linguistic Representations' (DFG-441119738).

Competing interests. The authors declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Ethical approval. Ethical approval for this study was obtained from Ethikkommission der Deutschen Gesellschaft für Sprachwissenschaften (Approval Number: 2023–04).

Consent to participate. Informed consent to participate was written. Participants read the consent statement and explicitly clicked to agree if they wished to participate and continue to the experiment.

Consent for publication. In the consent form, we have asked for and obtained written informed consent to publish their data anonymously, explaining that identifying details would be removed.

References

Ackerman, L. (2019). Syntactic and cognitive issues in investigating gendered coreference. *Glossa: A Journal of General Linguistics*, 4(1), 117–127. https://doi.org/10.5334/gjgl.721

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407.

Audring, J. (2008). Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology*, 18(2), 93–116. https://doi.org/10.1007/s11525-009-9124-y.

Audring, J., & Booij, G. (2009). Genus als probleemcategorie. *Taal en tongval*, 61(3), 13–37.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28. https://doi.org/10.21500/20112084.807.

Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgements. *Journal of Linguistics*, 46(2), 273–330. https://doi.org/10.1017/S0022226709990260.

Bader, M., & Portele, Y. (2025). Animacy outweighs topichood when choosing pronouns and word order. *Journal of Memory and Language*, 142, 104615. https://doi.org/10.1016/j.jml.2025.104615.

Bader, M., Torregrossa, J., & Rinke, E. (2023). Pinning down the interaction between animacy and syntactic function in the interpretation of German and Italian personal and demonstrative pronouns. *Discourse Processes*, 60(9), 655–673. https://doi.org/10.1080/0163853X.2023.2252699.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://CRAN.R-project.org/package=lme4

Binanzer, A., Schimke, S., & Schunack, S. (2022). Syntaktische Domäne oder lineare Distanz – Welcher Faktor steuert semantische Kongruenz im Kontext von hybrid nouns und Epikoina in stärkerem Maß? In

G. Diewald & Nübling (Eds.), *Genus – Sexus – Gender* (pp. 193–218). de Gruyter. https://doi.org/10.1515/9783110746396-006.

Birkenes, M. B., Chroni, K., & Fleischer, J. (2014). Genus- und Sexuskongruenz im Neuhochdeutschen: Ergebnisse einer Korpusuntersuchung zur narrativen Prosa des 17. Bis 19. *Jahrhunderts. Deutsche Sprache*, 42(1), 1–24. https://doi.org/10.37307/j.1868-775X.2014.01.02.

Birkenes, M. B., & Fleischer, J. (2022). Genus- und Sexuskongruenz im Mittelhochdeutschen: Eine Paralleltextanalyse zum lexical hybrid kint. In G. Diewald & D. Nübling (Eds.), *Genus – Sexus – Gender* (pp. 241–265). de Gruyter. https://doi.org/10.1515/9783110746396-008.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8, e9414. https://doi.org/10.1027/1618-3169/a000123.

Cacciari, C., Carreiras, M., & Cionini, C. B. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, 37, 517–532. https://doi.org/10.1006/jmla.1997.2528.

Caffarra, S., Janssen, N., & Barber, H. A. (2014). Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing. *Neuropsychologia*, 63, 124–134. https://doi.org/10.1016/j.neuropsychologia.2014.08.016.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49A(3), 639–663. https://doi.org/10.1080/713755647.

Cauldron Science (2022). *Gorilla experiment builder*. https://gorilla.sc.

Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, 5, 630. https://doi.org/10.3389/fpsyg.2014.00630.

Corbett, G. G. (2006). *Agreement. Cambridge textbooks in linguistics*. Cambridge University Press.

Czech, H. (2014). Zur variation grammatischer und semantischer Genuskongruenz bei personal- und Relativpronomen im Deutschen: Variablenbindung und lineare Distanz als potenzielle Einflussfaktoren. *Tagungsband zur StuTS*, 55, 1–40. https://doi.org/10.1515/zgl-2015-0015.

Dahl, Ö. (2000). Animacy and the notion of semantic gender. In B. Unterbeck, M. Rissanen, T. Nevalainen, & M. Saari (Eds.), *Gender in grammar and cognition. Trends in linguistics: Studies and monographs* ***124***. Mouton de Gruyter.

De Vogelaer, G., Fanta, J., Poarch, G., Schimke, S., & Urbanek, L. (2020). Syntactic or semantic gender agreement in Dutch, German and German learner Dutch: A speeded grammaticality judgement task. In G. De Vogelaer, D. Koster, & T. Leuschner (Eds.), *German and Dutch in contrast: Synchronic, diachronic and psycholinguistic perspectives* (Vol. 11, pp. 271–298). De Gruyter.

Duden (2016). *Das Wörterbuch der sprachlichen Zweifelsfälle. Richtiges und gutes Deutsch*. (M. Hennig (Ed.); with R. Osterwinter, J. G. Schneider, A. Steinhauer and the Dudenredaktion, 8th ed.). Duden Verlag (Duden – Deutsche Sprache in 12 Bänden).

Duden (2022). *Die Grammatik. Struktur und Verwendung der deutschen Sprache. Sätze - Wortgruppen – Wörter*. (A. Wöllstein (Ed.); *With the collaboration of the Dudenredaktion*, 10th ed.). Duden Verlag (Duden – Deutsche Sprache in 12 Bänden).

Eckert, P. (1998). Age as a sociolinguistic variable. In F. Coulmas (Ed.), *The handbook of sociolinguistics* (Vol. 9, pp. 151–167). Blackwell Handbooks in Linguistics. https://doi.org/10.1002/9781405166256.CH9.

Esaulova, Y., & von Stockhausen, L. (2022). Empirical evidence for subtle gender biases in language. In D. Ayoun (Ed.), *The Acquisition of Gender* (pp. 49–70). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Fleischer, J. (2022). Qualität hat seinen Preis": Genus-insensitives "sein" im Gegenwartsdeutschen. *Linguistische Berichte*, 271, 251–288.

Foley, S., & Ahn, B. (2024). Processing cataphoric *they* amidst pronominal innovation. *Journal of Cultural Cognitive Science*, 8, 289–308. https://doi.org/10.1007/s41809-024-00153-7.

Frank, M. (2019). Das anaphorische Referenzsystem. In M. Frank (Ed.), *Phorische Verkettung im Deutschen: Eine exemplarische Untersuchung anhand von Diskursrelationen der kausalen Gruppe*, Linguistik – Impulse & Tendenzen, Vol. 79, Kapitel 3, 80–281. de Gruyter.

Fukumura, K., & van Gompel, R. P. (2011). The effect of animacy on the choice of referring expression. *Language and Cognitive Processes*, 26(10), 1472–1504. https://doi.org/10.1080/01690965.2010.506444.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

Gries, S. (2021). *Statistics for linguistics with R: A practical introduction*. De Gruyter. https://doi.org/10.1515/9783110718256.

Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2005). Pronominal reference in sentences about persons or things: An electrophysiological approach. *Journal of Cognitive Neuroscience*, 17(2), 227–239.

Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Research*, 1230, 177–191. https://doi.org/10.1016/j.brainres.2008.07.004.

Hellinger, M., & Bußmann, H. (2015). Gender across languages: The linguistic representation of women and men. In M. Hellinger & H. Motschenbacher (Eds.), *Gender across languages* (Vol. 4, pp. 1–25). John Benjamins Publishing Company.

Hübner, J. (2021a). Genus und Sexus im Konflikt. Kongruenzformen hybrider Nomina im Sprachproduktionsprozess. *Linguistik online*, 107(2), 9–19. https://doi.org/10.13092/lo.107.7682.

Hübner, J. (2021b). Das Mädchen und ihr Liebhaber. Pragmatik als motivierender Faktor von Sexuskongruenz. In A. Binanzer, J. Gamper, & V. Wecker (Eds.), *Prototypen – Schemata – Konstruktionen* (pp. 31–52). De Gruyter. https://doi.org/10.1515/9783110710595-003.

Imai, M., Schalk, L., Saalbach, H., & Okada, H. (2014). All giraffes have female-specific properties: Influence of grammatical gender on deductive reasoning about sex-specific properties in German speakers. *Cognitive Science*, 38(3), 514–536. https://doi.org/10.1111/cogs.12074.

Irmen, L., Holt, D. V., & Weisbrod, M. (2010). Effects of role typicality on processing person information in German: Evidence from an ERP study. *Brain Research*, 1353, 133–144. https://doi.org/10.1016/j.brainres.2010.07.018.

Irmen, L., & Kurovskaja, J. (2010). On the semantic content of grammatical gender and its impact on the representation of human referents. *Experimental Psychology*, 57(5), 367–375. https://doi.org/10.1027/1618-3169/a000044.

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. Van Patten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). Routledge.

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951.

Kennison, S. M. (2003). Comprehending the pronouns *her*, *him*, and *his*: Implications for theories of referential processing. *Journal of Memory and Language*, 49(3), 335–352. https://doi.org/10.1016/S0749-596X(03)00071-8.

Köpcke, K. M., Panther, K. U., & Zubin, D. A. (2010). Motivating grammatical and conceptual gender agreement in German: empirical studies. In H.-J. Schmid & S. Handl (Eds.), *Cognitive Foundations of Linguistic Usage Patterns* (Vol. 13, pp. 171–194). Berlin, New York: de Gruyter.

Köpcke, K. M. & Zubin, D. A. (1997). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. In H. Sieburg (Ed.), Sprache – Genus/Sexus. Dokumentation germanistischer Forschung 3, 86–115. Peter Lang.

Köpcke, K. M., & Zubin, D. A. (2009). Genus. In E. Hentschel & P. M. Vogel (Eds.), *Deutsche Morphologie* (pp. 132–154). De Gruyter.

Kopf, K., & Bildhauer, F. (2024). The genitive alternation in German. *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/cllt-2024-0017.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58(2), 239–261.

Leeser, M., Brandl, A., & Weissglass, C. (2011). Task effects in second language sentence processing research. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching, and research: Insights from psycholinguistics* (pp. 179–198). John Benjamins. https://doi.org/10.1075/lllt.30.13lee?locatt=mode:legacy.

Leiss, E. (1997). Genus und Sexus. Kritische Anmerkungen zur Sexualisierung der Grammatik. In H. Sieburg (Ed.), *Sprache – Genus/Sexus. Dokumentation germanistischer Forschung* (Vol. 3, pp. 322–345). Peter Lang.

Lenth, R. V., Banfai, B., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Piaskowski, J., Riebl, H., & Singmann, H. (2024). emmeans: Estimated Marginal Means, aka Least-Squares Means (Version 1.10.6) [R package]. https://CRAN.R-project.org/package=emmeans

https://doi.org/10.1017/langcog.2025.10032 Published online by Cambridge University Press

Martín-Loeches, M., Hinojosa, J. A., Fernández-Frías, C., & Rubia, F. J. (2001). Functional differences in the semantic processing of concrete and abstract words. *Neuropsychologia*, 39(10), 1086–1096. https://doi.org/10.1016/S0028-3932(01)00033-1.

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

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. A. Just (Eds.), *New methods in reading comprehension*. Earlbaum.

Myers, J. (2009). Syntactic judgement experiments. *Language and Linguistics Compass*, 3(1), 406–423. https://doi.org/10.1111/j.1749-818X.2008.00113.x.

Nieuwland, M. S., & van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098.

Niu, R., & Liu, H. (2022). Effects of syntactic distance and word order on language processing: An investigation based on a psycholinguistic Treebank of English. *Journal of Psycholinguistic Research*, 51, 1043–1062. https://doi.org/10.1007/s10936-022-09878-4.

Oelkers, S. (1996). Der Sprintstar und ihre Freundinnen". Ein empirischer Beitrag zur Diskussion um das generische Maskulinum. *Muttersprache*, 106, 1–15.

Oppenrieder, W., & Thurmair, M. (2002). Doppelt genäht hält besser": Mehrfachmarkierung im Gegenwartsdeutschen. *Jahrbuch für Internationale Germanistik, Reihe A, Band*, 54, 117–123.

Ortmann, A. (1998). The role of [±animate] in inflection. In R. Fabri, A. Ortmann, & T. Parodi (Eds.), *Models of inflection. Linguistische Arbeiten* (Vol. 388, pp. 60–84). https://doi.org/10.1017/S0022226700238267.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34(6), 739–773. https://doi.org/10.1006/jmla.1995.1033.

Panther, K. U. (2009). Grammatische versus konzeptuelle Kongruenz. Oder: Wann siegt das natürliche Geschlecht? In R. Brdar-Szabó, E. Knipf-Komlósi, & A. Péteri (Eds.), *An der Grenze zwischen Grammatik und Pragmatik* (Vol. 3, pp. 67–86). Peter Lang.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006.

Piepers, J., & Redl, T. (2018). Gender-mismatching pronouns in context: The interpretation of possessive pronouns in Dutch and Limburgian. *Linguistics in the Netherlands*, 35(1), 97–110. https://doi.org/10.1075/avt.00007.pie.

Prolific (2014). (Version 08/2023). London, UK. https://www.prolific.com

R Core Team (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

RStudio Team (2024). *RStudio: Integrated development for R. RStudio version 4.4.2*. Boston, MA: PBC. http://www.rstudio.com/

Schmitz, D., & Esser, J. (2021). SfL: Statistics for linguistics (Version 0.4.2) [R package]. https://github.com/dosc91/SfL

Schütze, C. (accepted). "Die Romanfigur hat seine Geschichte": Zur Akzeptanz genusinkongruenter Possessivpronomina bei epizönen Personenbezeichnungen. Germanistische Linguistik.

Schütze, C. T. (2016). *The empirical base of linguistics: Grammaticality judgements and linguistic methodology*. Language Science Press. https://doi.org/10.26530/OAPEN_603356.

Schütze, C. T., & Sprouse, J. (2014). Judgement data. In Podesva, R. J., & Sharma, D. (Eds.). Research methods in linguistics ***27***, Chapter 3, 27–50. Cambridge University Press.

Shake, M. C., & Stine-Morrow, E. A. (2011). Age differences in resolving anaphoric expressions during reading. *Aging, Neuropsychology, and Cognition*, 18(6), 678–707. https://doi.org/10.1080/13825585.2011.607228.

Siewierska, A. (2004). *Person*. Cambridge University Press. https://doi.org/10.1017/CBO9780511812729.

Sorlin, S., & Gardelle, L. (2018). Anthropocentrism, egocentrism and the notion of Animacy hierarchy. *International Journal of Language and Culture, From Culture to Language and Back: The Animacy Hierarchy in Language and Discourse*, 5(2). https://doi.org/10.1075/ijolc.00004.gar.

Stefanowitsch, A. (2011). Keine Grammatik ohne Konstruktionen: Ein logisch-ökonomisches argument für die Konstruktionsgrammatik. In S. Engelberg, A. Holler, & K. Proost (Eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik* (pp. 181–210). De Gruyter. https://doi.org/10.1515/9783110262339.181.

Steriopolo, O., & Schütze, C. (2025). Referential means in German: An experimental study comparing feminine epicene nouns with masculine generic nouns. *Linguistics*. https://doi.org/10.1515/ling-2023-0049.

Stewart, A. J., Holler, J., & Kidd, E. (2007). Shallow processing of ambiguous pronouns: Evidence for delay. *Quarterly Journal of Experimental Psychology*, 60(12), 1680–1696. https://doi.org/10.1080/17470210601160807.

Thurmair, M. (2006). "Das Model und ihr Prinz". Kongruenz und Texteinbettung bei genus-Sexus-Divergenz. *Deutsche Sprache: ds*, 34, 191–220.

van Gompel, R. P., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 128–139. https://doi.org/10.1037/0278-7393.29.1.128.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58, 475–482. https://doi.org/10.1007/BF03395630.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. https://doi.org/10.21105/joss.01686

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, 2(3), 199–220. https://doi.org/10.1207/s1532799xssr0203_2.

Cite this article: Schütze, C., Fleischer, J., & Domahs, U. (2025). Gender conflicts in German possessives: comparing inanimate to human reference reveals asymmetries, *Language and Cognition*, 17, e73, 1–26. https://doi.org/10.1017/langcog.2025.10032