

# Syntactic and cognitive issues in investigating gendered coreference

Lauren Ackerman  
*Newcastle University*  
Newcastle upon Tyne, NE1 7RU  
lauren.ackerman@ncl.ac.uk

**Abstract** This squib examines how the multifaceted nature of “gender” (human and linguistic) interacts with grammatical operations such as coreference dependency formation. It particularly focuses on how personal names and referents who identify as nonbinary can provide insight into the cognitive representations of gender. A three-tiered model of gender is proposed that unites grammatical, cognitive, social, and biological aspects and describes how implications of this model might be tested in future work.

**Keywords:** gender; features; coreference; pronouns; personal names

# 1 Introduction

The term *gender* is fraught in part because it is so difficult to define precisely, even within a specific academic context like coreference resolution. This squib aims to clarify what kinds of gender might be relevant for real-time processing of coreference between an anaphor and a referring expression, particularly noting proper names. It develops the hypothesis that the type of gender involved in coreference checking in English, and possibly many other languages, is primarily a domain-general categorical representation of the referent rather than a formal syntactic or semantic feature. It then suggests a line of research into individual variation that would be able to inform the questions brought up herein.

This squib proposes three kinds of “gender” that should be distinguished in theories of (co)reference: grammatical gender, cognitive gender, and biosocial gender. It describes a three-tiered schema for formalizing the process of gender checking during coreference resolution. While English is the primary focus of this paper, motivation for these three categories can be found cross-linguistically. I draw on biological, social, cognitive, and grammatical evidence for how gender is conceptualized and used in human interaction in order to argue that coreference resolution (in English) relies primarily on a non-syntactic property, cognitive gender, for determining whether or not a pronoun and antecedent match or mismatch, which may additionally be domain-general in origin.

The difference between sentences (1-a) and (1-b) illustrates that English coreference is influenced by discourse-level information and world knowledge. In order to develop a felicitous context for (1-a), one almost must assume the speaker is communicating their disapproval of the referent through misgendering (although the referent’s gender remains ambiguous). In contrast, (1-b) provides a context that immediately allows for a felicitous and not (overtly) transphobic interpretation.

- (1)    a. # At the farmhouse, the cowgirl<sub>i</sub> left his<sub>i</sub> lasso in the kitchen.  
       b.    At the Halloween party, the cowgirl<sub>i</sub> left his<sub>i</sub> lasso in the kitchen.

The difference in apparent acceptability between these two sentences indicates that the property of gender relevant for coreference is, at the very least, more complex than a formal syntactic feature. This observation by itself is not novel (e.g. Collins & Postal 2012; Frazier et al. 2015; Nieuwland & Van Berkum 2006; Hess, Foss & Carroll 1995; Duffy & Keir 2004; Pyykkönen, Hyönä & van Gompel 2010; Garnham, Oakhill & Reynolds 2002). Thus this paper develops a formal treatment of how specific “types” of gender can match or mismatch during coreference dependency resolution and what this means for the linguistic encoding of gender identity.

## 2 Background

This paper presents one possible way to explain how gender is conceived and applied to linguistic referents. It represents a self-consistent, comprehensive model that can be tested empirically. Furthermore, it provides a starting point for interdisciplinary research into the many linguistic facets of gender, particularly those involved in distinguishing an individual's cognitive gender: the way others categorize or classify that individual, from biosocial gender: a more complicated combination of biological, cultural and experiential factors elaborated upon below.

In order to precisely distinguish different types of gender, the following section briefly defines grammatical, cognitive, and biosocial gender. These three types have been derived from syntactic, semantic, typological, social, and neuro-biological work on gender. They are not intended to be all-encompassing, rather they are a terminological starting point for a coherent and precise discussion across fields and subfields in which the word *gender* may be used for multiple distinct concepts.

### 2.1 Grammatical gender

*Grammatical gender* comprises formal morphosyntactic features. It is the property of words that allows the formal grammatical process of agreement to be carried out. This includes agreement of gender categories (henceforth *labels*) such as masculine, feminine, neuter, common, etc.<sup>1</sup> These features are properties of the morphemes themselves, and may be independent from the real-world genders associated with the referents, as in (2), but compare (3) which reflects some “real world” knowledge in the agreement of grammatical gender (both adapted from Comrie 2005).

- (2) *Tsez*: The anthropomorphic rooster (definitionally male), his hen wife (definitionally female), and his frog mistress (gender is never specified, but is assumed female) are all associated with Gender III, the noun class containing all non-human animals:

b - oχix - no    ɬoɦr - ā    eχi - n    wit'-wiš χin  
 III - appear - P.CONV frog - ERG say - P.UNW wit'wish QUOT  
 'The frog appeared and said "witwish".'

- (3) *Russian*: The anthropomorphic (but genderless) mirror either must use or is most acceptable with a neuter verb conjugation when it speaks in a translation of *Snow White*:

<sup>1</sup> Grammatical gender likely includes noun classes as well, although the discussion of noun class semantics is beyond the scope of this paper.

Ja o politike predpočlo by ne govorit'  
 I about politics.PREP prefer.P.NEUT COND not speak.INF  
 'I would prefer not to talk about politics.'

In English, it is unclear whether or not grammatical gender plays a role in syntactic operations or psycholinguistic processes. It has been argued that English has completely lost grammatical gender, based on historical changes and loss of productive gender morphology (Baron 1971). Certainly, there is no overt gender agreement between nouns, adjectives and articles. However, Bjorkman's recent treatment of gender agreement between names and pronouns makes a case for a limited grammatical gender system in English, in which sentences like (4) display a contrast in acceptability (Bjorkman 2017).

- (4) a. That surgeon<sub>i</sub> operated on three of their<sub>i</sub> patients today.  
 b. ?\* Johnathan<sub>i</sub> operated on three of their<sub>i</sub> patients today.

Bjorkman observes that sentences like (4-a) are more acceptable than (4-b), even when the surgeon is known to all parties, and suggests this is due to names having grammatical gender (i.e., a  $\phi$ -feature) in English, which must then agree with the pronoun, at least for some speakers. Whether or not English makes use of grammatical gender to determine gender congruency between coreferring elements, an argument for  $\phi$ -features on names must account for how gender (cognitive and/or grammatical) is associated with their referents, since gender bias of names is wildly variable and mutable.

## 2.2 Cognitive gender

*Cognitive gender* encompasses a large number of closely related terms currently in use in the literature. This includes semantic gender (e.g. Asarina 2009), definitional gender (Kreiner, Sturt & Garrod 2008) and notional gender (i.e. natural gender, but see McConnell-Ginet 2015: for why the term 'natural' is inappropriate), which are ways of associating lexical items with masculine or feminine properties, but without necessarily attributing  $\phi$ -features to them. Returning to sentence (1) for instance, *cowgirl* is definitionally female, but can be used for a male/masculine referent in certain circumstances. The feminine definition associated with *cowgirl* is thus defeasible, since gender agreement between *cowgirl* and *his* should be impossible if the property being checked is a  $\phi$ -feature. This is not incompatible with English having formal gender features for some words, but I argue that it is strong evidence that what is primarily relevant for coreference resolution is not the morphosyntactic feature. This argument will be elaborated upon in Section 4.1, below.

Furthermore, there is evidence from developmental psychology and language acquisition that young children acquire labels for gender categories before they are able to consistently sort people into those categories (Fagot & Leinbach 1993). At this point in development, (at least) two gender categories are present but not enough input has been received to develop a consistent rubric for evaluating the massive variation present in the population. For instance, children may be able to use the proper pronouns for common and canonically gendered referents (e.g., “mommies” and “daddies”) but fail to generalize identification criteria to novel referents that deviate in one or more ways (e.g., men with long hair, women wearing collared shirts) (Fagot & Leinbach 1993). This may indicate that gender categories are developed and refined by repeated exposure to exemplars.

### 2.3 *Biosocial gender*

*Biosocial gender* is, fundamentally, an individual’s gender as it is experienced internally. In addressing this type of gender, a few terminological clarifications are necessary. I will assert a distinction between sex and gender, which are widely confounded terms in linguistics (Cheshire 2002). Herein, *sex* refers to biological properties such as karyotype (XX, XY, etc.) and phenotype (e.g., internal and external anatomy). Even in biological terms, sex is not a binary property since the physical traits contributing to an organism’s sex can vary along multiple dimensions. As an edge case, people with Complete Androgen Insensitivity Syndrome (CAIS) may have XY chromosomes but a predominantly female phenotype (e.g., Hughes et al. 2012). However, sex is still often used as a shorthand for distinguishing the bimodal nature of the male-female spectrum (Johnson & Repta 2012; Lorber 1996).

This definition of sex overlaps with biosocial gender. More precisely, biosocial gender is the multidimensional property of an individual as determined by their biology and cultural norms of identity expression. What distinguishes biosocial gender from other types of gender is that, as an external observer, one’s accuracy of categorization is impossible to assess without the individual’s introspection and medical history. That is, biosocial gender may not be something that can be doubtlessly determined without detailed anthropological, introspective and potentially invasive medical analyses. This is because social pressures and societal norms can contribute to an individual representing themselves in a way that is inconsistent with the way they categorize themselves. A particularly clear illustration is the case of transgender people who are “in the closet” or otherwise representing themselves as a binary gender category to which they were assigned at birth but would not categorize themselves as being. Here, an individual’s biosocial gender might be in direct conflict with the gender with which other people would categorize them.

Our present understanding of gender suggests that the majority of people do fall into a bimodal distribution of biosocial genders (0.4% of respondents in a UK survey reported thinking of themselves as a way other than male' or female'; [Glen & Hurrell 2012](#)). But many individuals do not categorized themselves with a discrete binary label, and it would do the science and the individuals a disservice to gloss over the often subtle and diverse variations present in the population at large, even within one of the stereotypical categories ([Johnson & Repta 2012](#)). Despite the potential complications in identifying the precise biosocial gender of an individual, it is still an important factor for phenomena involving social identity ([Zimman 2017](#)). Therefore, it is important to explicitly define biosocial gender as distinct to ensure it is not confounded during investigation of phenomena associated with either biosocial- or cognitive- genders.

### 3 Further evidence

#### 3.1 *Personal names as antecedents*

Personal names comprise a large portion of antecedents used in empirical investigations and syntactic judgments of English coreference, presumably due to their intuitive gender-specificity. However, English lacks overt morphological marking on names to unambiguously distinguish a correct assessment of the cognitive gender of the referent. A clear example of this problem is illustrated in (5-a), in which the two given pronouns can corefer with the name equally well in the absence of disambiguating context (such as whether the Taylor in question is Taylor Swift, a woman, or Taylor Lautner, a man. As for Taylor Mason, a nonbinary character played by the nonbinary actor Asia Kate Dillon, (5-b) is the appropriate formulation ([Dillon 2017, April 10](#)), although the processing cost and intuitive acceptability of this linguistic structure, in terms of linguistic judgments, is currently a subject of investigation ([Ackerman 2018](#); [Conrod 2018](#); [Konnelly & Cowper 2017](#); [Prasad, Morris & Feinstein 2018](#)).

- (5) a. On the red carpet, Taylor<sub>i</sub>'s fans screamed to get [his<sub>i</sub>/her<sub>i</sub>] attention.  
 b. On the red carpet, Taylor<sub>i</sub>'s fans screamed to get [their<sub>i</sub>] attention.

One possibility is that the name *Taylor* is stored in the lexicon as discrete entries (e.g., Taylor<sub><masc></sub>, Taylor<sub><fem></sub>).<sup>2</sup> In this case, a comprehender may retrieve one of the two entries initially, but have to revise the selection if conflicting information

<sup>2</sup> The possibility of the lexicon containing Taylor<sub><nonbinary></sub> is a logical possibility but will not be discussed further at this point because of the current dearth of empirical studies on nonbinary gender perception and its influence on lexical categories.

is received at a later time during comprehension. The presence of different lexical entries for each string-identical name with a distinct valuation of a gender -feature makes testable predictions regarding the learning and application of new lexical entries. One can quickly learn a new name or a new use of a common name, but if extensive previous experience with a common name (e.g., Michael<sub><masc></sub>) influences the processing of a newly encountered and rare version of the name (Michael<sub><fem></sub>), this might be observable in behavioral or psychophysical measures.<sup>3</sup>

Another possibility is that “unisex” names are underspecified for gender (e.g., Taylor<sub><0></sub>), and whatever gender assumptions are made about the referent are done so without reference to the lexicon. However, it is not immediately clear what the implications of this configuration would be or how this could be tested. At the very least, it would be necessary to conduct extensive evaluation of each individual participant’s experience with the target names and gender nonconformity and examine effects from the perspective individual differences.

### 3.2 *Cross-linguistic and cross-cultural evidence*

More than just a language of convenience, English has certain properties that allow dissociation of the three proposed types of gender. English marks gender (broadly construed) on its third person pronouns (*she*, *he*), but it does not have overt or productive morphological agreement for gender. Numerous studies demonstrate strong biases for the gendering of certain noun phrases (e.g., *surgeon*, *pilot*, *nurse*, *babysitter*), but these are defeasible (Kennison & Trofe 2003; Duffy & Keir 2004; Pyykkönen, Hyönä & van Gompel 2010; Garnham, Oakhill & Reynolds 2002; Kreiner, Sturt & Garrod 2008). Furthermore, English has some remnants of gendered morphology (*actor/lactress*, *aviator/laviatrix*) and definitionally gendered nouns (*mother*, *father*, *cowgirl*, *bellboy*). Finally, in cultural terms, English has been at the international forefront of informal, community-based development of nonbinary language and so-called “neopronouns” (e.g., Spivak pronouns introduced by Spivak (1990: xv), gender variant neologisms described in Centauri (2013), a.o.).

<sup>3</sup> If names are stored generically with stereotypical/proportional/statically determined gender, then by familiarizing a naive participant to an uncommon or novel pairing between a name and gender (e.g., a woman named Michael or a boy named Sue), there should still be a detectable processing cost to forming a coreference dependency between the pronoun and name. However, if names instead receive gendered properties from domain-general or world knowledge, then retrieval of the uncommon entry should be facilitated more by the context and less processing cost should be observed (Pyykkönen, Hyönä & van Gompel 2010). See Cai et al. (2017) for examples of how long- and short-term learning can be tested.

It is relevant to note that many cultures around the world have established and traditional nonbinary, queer, and third-gender categories (e.g., Navajo: [Epple \(1998\)](#), Bugis: [Graham \(2004\)](#), Māori: [Murray \(2003\)](#), a.o.), though apparently none of those cultures speak languages that explicitly grammatically indicate a nonbinary gender *on pronouns*. Rather, pronouns may formally mark animacy or when they do mark gender, nonbinary gender categories are indicated through shifting use of standard binary gender agreement (e.g., [Hall & O'Donovan 1996](#)).

## 4 Gender in coreference resolution

Coreference resolution is said to compare the grammatical features of the anaphoric element and the candidate antecedent ([Garnham, Oakhill & Reynolds 2002](#); [Garnham & Oakhill 1990](#)). Thus, there must be criteria for what counts as ‘matching’ or ‘mismatching’ in order for a coreference dependency to be resolved or rejected. In a case such as (1), restated below in (6) where coreference is resolvable but is not a priori congruent, one might expect the decontextualized mismatch in gender between *cowgirl* and *his* to create a processing slowdown. However, the underlying mechanism for such a prediction is not transparently derivable from syntax-first models of real-time coreference resolution without incorporation of discourse-level knowledge.

- (6) a. # At the farmhouse, the cowgirl<sub>i</sub> left his<sub>i</sub> lasso in the kitchen.  
 b. At the Halloween party, the cowgirl<sub>i</sub> left his<sub>i</sub> lasso in the kitchen.

A strict criterion for matching might look something like this (loosely adapted from definitions of agreement by [Lasnik & Uriagereka \(e.g. 1988\)](#); [Carnie \(2007\)](#); [Payne & Huddleston \(2002\)](#)):

**Strict matching criterion:** Matching gender requires the formal grammatical feature ( $\phi$ -feature) of the anaphor to be identical to the candidate antecedent. If the features are not identical, the coreference dependency is rejected.

However, this strict matching criterion poses problems for some common, well-described types of coreference in which the antecedent is not explicitly or overtly present in the syntax (e.g., the ‘statue rule’: [Jackendoff \(1992\)](#); “impostors” which are superficially 3<sup>rd</sup> person but conceptually 2<sup>nd</sup> or 1<sup>st</sup>: [Collins & Postal \(2012\)](#)):

- (7) a. Regarding a customer [Jackendoff \(1992\)](#):  
 [The ham sandwich in the corner]<sub>i</sub> needs his<sub>i</sub> bill.  
 b. Spoken to a king ([Collins & Postal 2012](#)):  
 [Your majesty]<sub>i</sub> must protect yourself<sub>i</sub>/himself<sub>i</sub>/\*herself<sub>i</sub>/\*themselves<sub>i</sub>.

Even still, in these cases of apparent number mismatch, some formal level of representation should contain number features that can be checked during coreference resolution, i.e. what [Collins & Postal \(2012\)](#) term a ‘source’. This type of argument can only account for the apparent gender mismatch in (6) if we posit that a masculine  $\phi$ -feature is attributed to *cowgirl* only after the candidate antecedent is encountered or identified. Thus, a slightly less strict criterion might be formulated as such, adapted for coreference processing from [Collins & Postal \(2012: 182\)](#)<sup>4</sup>:

**Less strict matching criterion:** The act of resolving a coreference dependency requires an identity relation between the  $\phi$ -features of a pronoun and either (a)  $\phi$ -features of the antecedent, or (b)  $\phi$ -features of the antecedent as determined by the semantic properties of the notional ‘source’. If the features are not identical, the coreference is rejected.

Yet, this still doesn’t quite cover the case of (6), where the cognitive gender of the referent is female but the coreference between the masculine pronoun and the (female) antecedent is licit. Neither does it fully explain (7), in which the antecedents might or might not be interchangeable with lexical items that have matching gender  $\phi$ -features (a: ✓ The man<sub><masc></sub>, ✗ The customer<sub><0></sub>; b: ✓ The king<sub><masc></sub>, ✗ The monarch<sub><0></sub>). This might be covered in two ways: there might be a way to override the feature checking criteria through modeling the parser as having earlier access to pragmatics and world knowledge, or the feature checking process has a broader criterion of what can count as matching. The latter could be formulated as such:

**Broad matching criterion:** Matching gender requires either a formal feature ( $\phi$ -feature) or a cognitive category of the lexical item to be identical to the candidate antecedent in order to match. A cognitive category might include a probabilistic representation of the semantic set of possible referents, but also would be susceptible to environmental context (e.g., [Cai et al. 2017](#); [Arnold et al. 2018](#)).

The alternate option in which discourse factors can override feature checking could also account for some of the cross-linguistic variation observed in the literature ([Comrie 2005](#)). That is, with one of the stricter definitions of gender matching, languages might be able to override the matching criterion when contextual evidence provides conflicting information to different degrees based on the variable strength of the criterion. Languages with very strong or strict matching criteria would then find it difficult to have pragmatic context override the formal gender features of the anaphor which triggered the coreference dependency.

<sup>4</sup> I have taken liberties in adapting this condition in order to present it without requiring extensive background or explanation.

However, all of this assumes that languages that have formal gender features on anaphoric elements also have formal features that can be checked on the candidate antecedents. What then, would happen if the candidate antecedent didn't have a  $\phi$ -feature in any instantiation? Would this cause a processing slowdown because the initial checking operation would automatically fail? If so, we should expect to see processing slowdowns for coreference dependencies which connect generic antecedents and gender-specific anaphora as compared to coreference dependencies which connect gender-specific antecedents to gender-specific anaphora, (Cf. Foertsch & Gernsbacher 1997).

## 4.1 *Feature checking*

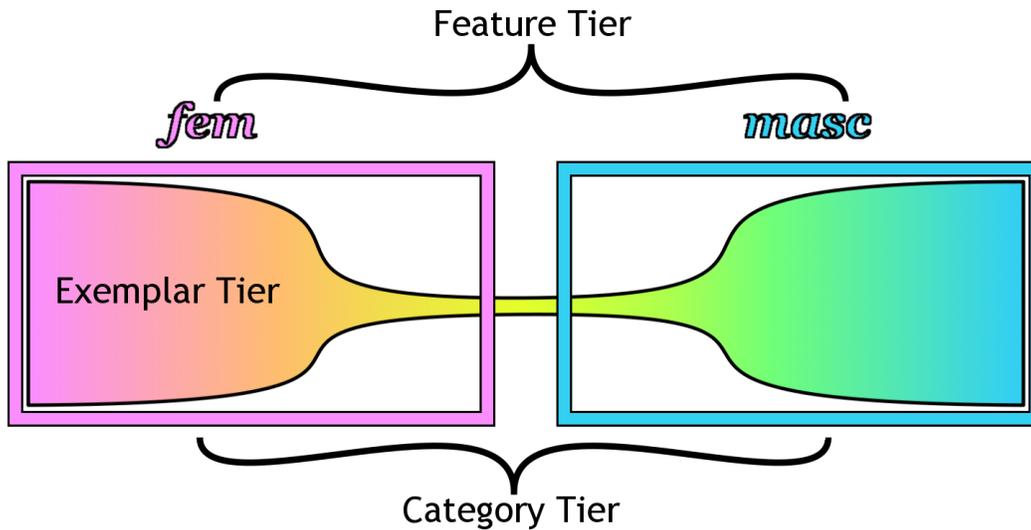
If formal morphosyntactic gender features are present in a language like English, but cannot be used to model how the parser checks for congruency in coreference dependency formation, what purpose do they serve? I will not argue for or against English having formal grammatical features for gender, but rather that such features are irrelevant during coreference dependency formation. Instead, English and languages with similar gender systems rely on cognitive gender for evaluating gender congruency in real time. In order to describe how such a system operates, a three-tiered scheme of linguistically and cognitively encoding gender is posited below.

The three tiers comprise an Exemplar Tier, a Category Tier, and a Feature Tier (Figure 1). These tiers are not meant to represent actual processing mechanisms or structures in the mind. Rather, they are abstract categories of processes or representations that can be used to map behaviors and empirical observations to theoretical properties of grammars. Thus, each tier is designed to be as theory-agnostic as possible.

### 4.1.1 **The Exemplar Tier**

The Exemplar Tier consists of observations from individual's exposure to the variety of observable gender expression. This may include tokens of phenotypic variation, non-conformity of gender expression, and variation of cultural norms. Crucially, most individuals will be primarily exposed to other individuals who have unambiguous binary gender expression, thus will have distinctly bimodal input represented in this tier (Fagot & Leinbach 1993; Glen & Hurrell 2012). Individuals who are members of or adjacent to non-conforming or nonbinary communities may have a different distribution of input, especially if exposure occurs during early acquisition of gender categories.

It cannot be that this tier includes the perceiver's perception of the gender of the person with which they interact with, because that requires a categorization be-



**Figure 1:** A schema depicting the three proposed tiers, overlaid.

haviour that is crucially not a component of this tier. Instead, the tokens in this tier might be conceptualized as matrices of perceived properties that are used downstream to categorize the gender of the individual. For example, hair length and style, face shape, pitch range of voice, clothing style, sociolinguistically marked properties of speech, etc, could be dimensions of each token. These properties can be used to categorize an individual's gender (Fagot & Leinbach 1993), but are not inherently properties of binary biosocial genders. Furthermore, few of these properties are purely linguistic. This tier therefore represents a way of organizing general perceptual input about individuals who a person encounters and interacts with throughout the lifespan.

#### 4.1.2 The Category Tier

The Category Tier consists of categories that are established through cognitive processes relying on input from the exemplar tier and semantics (e.g., gender schema; Bem 1981; Fagot & Leinbach 1993). The categories of gender encoded in this tier may shift if the distribution of input to the Exemplar Tier changes. As an individual accumulates more exemplars over the lifespan, each new token will comprise a smaller proportion of the total input, thus will have less influence on the shape of the Category Tier. The way someone sorts individuals into gender categories should take into account a subset of the dimensions catalogued in the Exemplar Tier. However an individual categorizes people into genders and what information is used to make those determinations, the Category Tier holds information about

the parameters of each gender category. The structure and robustness of this tier relies on the assumption that gender is typically perceived categorically (Fagot & Leinbach 1993).

For example, this could manifest as recognition of variance in feminine gender expression and what it means to self-identify as a gender (e.g., Zimman 2017). However, humans are still eager to and adept at categorizing people with short hair who are wearing pants into *feminine* and *masculine* categories (leaving aside the accuracy or relevance of these categories). This suggests that the categorical perception of gender is complex and culturally specific. The details of this process are beyond the scope of this paper. What remains relevant is that the boundaries of these categories may slightly differ between individuals within a culture or society. Thus the boundaries may differ more between individuals belonging to different cultures or societies (as phonemic boundaries differ more for individuals who have acquired different languages).

These categories are not strictly linguistic, but contribute to assessments of whether linguistic meanings are consistent or felicitous when concerning the gender of referents. For instance, when discussing a known person (who is, say, categorized by both interlocutors as *female*), it may be relevant for the comprehension mechanism to refer to the category when assessing the plausibility of statements (Prasad, Morris & Feinstein 2018; Kreiner, Sturt & Garrod 2008).

The interaction of the Exemplar Tier and the Category Tier may generate and assign probabilities of genderedness to gender-biased (or equi-biased) lexical items, including names. In being exposed to instances of *surgeons* or *Michaels*, the tokens that have *surgeon* or the name *Michael* as a property fall predominantly into the male category. If this is the mechanism for generating gender stereotyping, then the stereotype would be accessed in one of several ways (that all have the same consequence): An aggregate of all *surgeon/Michael* tokens is assessed as a probability; an individual token of *surgeon* or *Michael* is evaluated for gender category (thus drawn at random from all tokens of *surgeon/Michael*); or the evaluation of gender is assessed at an earlier time and is a property that is rarely updated in the lexicon, independent of the content and structure of the Exemplar and Category tiers. Crucially, whatever the process for determining gender bias associated with a lexical item, its meaning, or gender plausibility, this information is separate from the grammatical information stored in the Feature Tier.

#### 4.1.3 The Feature Tier

The Feature Tier consists of discrete  $\phi$ -features which may include *<feminine>* and *<masculine>*, among others. These labels can be mapped onto the cognitive categories in the Category Tier, but need not be. During coreference resolution,

whether or not the Feature Tier is used to determine gender congruency is graded from strict grammatical gender systems to languages without grammatical gender (see Corbett (2015) for examples).

This tier differs from the Category tier in that the  $\phi$ -features are strictly linguistic and are formally encoded in the grammar of a language. That is, where the Category Tier concerns categorization of people and animate gendered referents based on social/cultural norms, the Exemplar Tier does not categorize anything: it consists of linguistic labels that are used in grammatical operations like agreement. These labels do not need to correspond to human gender (e.g. Bantu noun class systems, etc.), and can apply to inanimate lexical items. Furthermore, they do not apply to the *referents* of the relevant lexical items, but to the lexical items themselves. For instance, languages that have strict gender agreement will ignore the cognitive gender of the referents (Category Tier) in using grammatical gender to satisfy agreement relations (Feature Tier).

Together, these tiers are three levels at which the parser could assess gender congruency during coreference resolution. Once an anaphor is linked to a candidate antecedent, the parser may access one of the tiers to check gender congruency (Sturt 2003). If the Feature Tier does not supply relevant formal features for both lexical items (e.g., gender  $\phi$ -features for pronouns but not unisex names or gender-stereotyped nouns in English, Cf. Bjorkman (2017)), it cannot compare like to like. In this case, using the Category Tier as a holistic congruency assessment would be preferable. Speculatively, if the Exemplar Tier were to have a third mode (e.g., a nonbinary human gender), this might affect the structure of the other tiers and provide organic support for the genesis of novel personal pronouns (e.g., Centauri 2013).

## 4.2 *Typological evidence*

Together, these tiers describe three levels of encoding of gender, broadly construed, that a language (or an individual) may draw upon in order to determine the gender congruency of an anaphor and candidate antecedent during real-time coreference resolution.

- (8)    **Strict feature:** Languages with no exception to grammatical gender agreement which access only the Feature Tier during coreference resolution  
           **Strong feature:** Languages that only draw on the Category Tier in specific contexts  
           **Weak feature:** Languages that draw on the Category Tier when the Feature Tier is incongruent, but still have grammatical features

**Absent feature:** Languages without grammatical gender which make use of the Category Tier only where gender plausibility is concerned

Tsez exemplifies strict features, with (2) demonstrating a strict grammatical gender system for anthropomorphic animals (Comrie 2005; Corbett 2015). Languages of the 'absent feature' type would then rely entirely on the cognitive categorization of the antecedent to evaluate coreference feasibility. Languages with intermediate strategies like Spanish and Russian would then show some mixed properties wherein formal features are checked during coreference resolution, but may be overridden given contextually appropriate information (Asarina 2011; 2009).

- (9) Agreement patterns in Russian where *vrach* (*m*) refers to a woman (%=marked in certain registers) (Asarina 2009)
- a. Umnaja vrach prishla  
smart.FEM doctor(I) came.FEM
  - b. %Umnyj vrach prishel  
smart.MASC doctor(I) came.MASC
  - c. %Umnyj vrach prishla  
smart.MASC doctor(I) came.FEM
  - d. \*Umnyj vrach prishel  
smart.FEM doctor(I) came.MASC  
'The smart [female] doctor has come.'

Where does English fit into this hierarchy? As it has been claimed that English no longer has grammatical gender (except, possibly on pronouns) (Baron 1971), it might be an absent feature language. However, Bjorkman (2017) suggests that English does have limited use of grammatical gender agreement, particularly when referring to named individuals. If so, we might expect such cases to evoke psycholinguistic/cognitive behaviors that are similar to those observed in languages that make use of the Feature Tier. However, testing this is made difficult by the limited circumstances in which English could have grammatical gender. The potential environments for detecting grammatical gender in English are wholly overlapping with environments where cognitive gender (as determined by the Category Tier) could be an alternative source for checking during coreference resolution. That is, words that could have formal gender features (as Bjorkman suggests, personal names) should always also receive a gender property from the cognitive gender of the referent, encoded in the Category Tier.

## 5 Future directions and conclusions

There are a few ways to test the hypotheses described in this paper. If definitionally gendered nouns or personal names have formal grammatical gender in English, then there should be a failure in coreference resolution for the link between *cowgirl*<sub><fem></sub> and *his*<sub><masc></sub> in (1)/(6), or *Johnathan*<sub><masc></sub> and *their*<sub><0></sub> in (4). At this stage of processing, the parser may need to draw upon the Category Tier (rather than Feature Tier, as it may have originally attempted). This could presumably cause a processing slowdown or electrophysiological effect comparable to one that might be observed for a plausibility mismatch.

Since the anaphor in (1)/(6) is also definitionally masculine/male, in conjunction with the pragmatic context (a Halloween party, in which costumes allow people some flexibility in identity), the parser may reassign the gender of the lexical item *cowgirl* in a process similar to that of impostor anaphora (Collins & Postal 2012). This should be detectable in behavioral and psychophysiological measures (e.g., Nieuwland & Van Berkum 2006; Kuperberg et al. 2003; Canal, Garnham & Oakhill 2015). However, the tiered schema I propose predicts that individuals who have extensive exposure to third genders or gender nonconforming communities will have differently shaped exemplar distributions, thus also differently shaped category tiers. If the category tier is shaped in such a way that the boundaries between gender categories are overlapping or ambiguous, this may ease the processing cost of reanalysis.

The three types of gender distinguished in this proposal comprise a model for exposure to variance in gender expression, cognition, and linguistic encoding. The model is designed to be broadly applicable and testable across interfaces of linguistic, cognitive, psychological and sociological work. I describe some applications of the model to psycholinguistic topics and suggest future directions for development. Since forays into research on nonbinary gender are few and recent, the three-tiered model is intended to lead to better informed hypotheses about individual variation related to gender, language processing, and experience. Moreover, nonbinary people often suffer social stigma for their identities (McLemore 2015). This puts empirical studies touching on nonbinary issues in a position to set the standard for ethical and compassionate research on and in conjunction with nonbinary people. This squib provides a set of terminology and the beginnings of a framework from which formal, empirical, and experimental linguistic research on nonbinary issues can grow, while incorporating the varied experiences of the people directly affected by it.

## Abbreviations

<MASC> = masculine  $\phi$ -feature, <FEM> = feminine  $\phi$ -feature, <0> = no gender  $\phi$ -feature, ERG = ergative, P.UNW = past unwitnessed, P.CONV = past converb, P.NEUT = past neuter, INF = infinitive, QUOT = quotative, III = noun class III, PREP = prepositional, COND = conditional

## Acknowledgements

The author wishes to thank Joel Wallenberg, Anders Holmberg, and Bronwyn Bjorkman for their helpful and insightful comments on earlier drafts of this paper and Kirby Conrod for their helpful discussion throughout.

## Competing interests

The author has no competing interests to declare.

## References

- Ackerman, Lauren. 2018. Being themselves: processing and resolution of singular (im)personal they. In *The 31st CUNY Conference on Human Sentence Processing*. [osf.io/qba7d](https://osf.io/qba7d).
- Arnold, Jennifer E, Iris M Strangmann, Heeju Hwang, Sandra Zerkle & Rebecca Nappa. 2018. Linguistic experience affects pronoun interpretation. *Journal of Memory and Language* 102. 41–54. <http://dx.doi.org/10.1016/j.jml.2018.05.002>.
- Asarina, Alevtina. 2009. Gender and adjective agreement in Russian. In *The 4th Annual Meeting of the Slavic Linguistics Society*. Zadar, Croatia.
- Asarina, Alevtina. 2011. *Case in Uyghur and beyond*. Massachusetts Institute of Technology dissertation.
- Baron, Naomi S. 1971. A reanalysis of English grammatical gender. *Lingua* 27. 113–140. [http://dx.doi.org/10.1016/0024-3841\(71\)90082-9](http://dx.doi.org/10.1016/0024-3841(71)90082-9).
- Bem, Sandra Lipsitz. 1981. Gender schema theory: a cognitive account of sex typing. *Psychological Review* 88(4). 354–364. <http://dx.doi.org/10.1037/0033-295X.88.4.354>.
- Bjorkman, Bronwyn M. 2017. Singular they and the syntactic representation of gender in English. *Glossa: A Journal of General Linguistics* 2(1). <http://dx.doi.org/10.5334/gjgl.374>.

- Cai, Zhenguang G, Rebecca A Gilbert, Matthew H Davis, M Gareth Gaskell, Lauren Farrar, Sarah Adler & Jennifer M Rodd. 2017. Accent modulates access to word meaning: evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology* 98. 73–101. <http://dx.doi.org/10.1016/j.cogpsych.2017.08.003>.
- Canal, Paolo, Alan Garnham & Jane Oakhill. 2015. Beyond gender stereotypes in language comprehension: self sex-role descriptions affect the brain's potentials associated with agreement processing. *Frontiers in Psychology* 6:1953. <http://dx.doi.org/10.3389/fpsyg.2015.01953>.
- Carnie, Andrew. 2007. *Syntax: a generative introduction*. Wiley-Blackwell.
- Centauri, Widow. 2013. *Gender variant neologisms*. San Diego State University dissertation. <http://dx.doi.org/10211.10/4199>.
- Cheshire, Jenny. 2002. The Handbook of Language Variation and Change. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.). Blackwell Publishing Ltd. Chap. Sex and gender in variationist research, 423–443. <http://dx.doi.org/10.1002/9780470756591.ch17>.
- Collins, Chris & Paul Martin Postal. 2012. *Imposters: a study of pronominal agreement*. MIT Press.
- Comrie, Bernard. 2005. Grammatical gender and personification. In *Perspectives on language and language development*, 105–114. Springer.
- Conrod, Kirby. 2018. Pronouns in motion. In *Lavender Linguistics 2018 (LavLang)*. Providence, RI.
- Corbett, Greville G. 2015. *The expression of gender*. Vol. 6. Walter de Gruyter.
- Dillon, Asia Kate. 2017, April 10. *As It Happens: CBC Radio*. Interview with Carol Off.
- Duffy, Susan A & Jessica A Keir. 2004. Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition* 32(4). 551–559. <http://dx.doi.org/10.3758/BF03195846>.
- Epple, Carolyn. 1998. Coming to terms with Navajo Nádleehí: a critique of “berdache,” “gay,” “alternate gender,” and “two-spirit”. *American Ethnologist* 25(2). 267–290. <http://www.jstor.org/stable/646695>.
- Fagot, Beverly I & Mary D Leinbach. 1993. Gender-role development in young children: from discrimination to labeling. *Developmental Review* 13(2). 205–224. <http://dx.doi.org/10.1006/drev.1993.1009>.
- Foertsch, Julie & Morton Ann Gernsbacher. 1997. In search of gender neutrality: is singular they a cognitively efficient substitute for generic he? *Psychological Science* 8(2). 106–111. <http://dx.doi.org/10.1111/j.1467-9280.1997.tb00691.x>.
- Frazier, Michael, Lauren Ackerman, Peter Baumann, David Potter & Masaya Yoshida. 2015. Wh-filler-gap dependency formation guides reflexive antecedent search. *Frontiers in Psychology* 6(1504). <http://dx.doi.org/10.3389/fpsyg.2015.01504>.

- Garnham, Alan & Jane Oakhill. 1990. Mental models as contexts for interpreting texts: implications from studies of anaphora. *Journal of Semantics* 7(4). 379–393. <http://dx.doi.org/10.1093/jos/7.4.379>.
- Garnham, Alan, Jane Oakhill & David Reynolds. 2002. Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition* 30(3). 439–446. <http://dx.doi.org/10.3758/BF03194944>.
- Glen, Fiona & Karen Hurrell. 2012. *Technical note: measuring gender identity*. Tech. rep. Manchester: Equality & Human Rights Commission.
- Graham, Sharyn. 2004. It's like one of those puzzles: conceptualising gender among bugis. *Journal of Gender Studies* 13(2). 107–116. <http://dx.doi.org/10.1080/0958923042000217800>.
- Hall, Kira & Veronica O'Donovan. 1996. Rethinking Language and Gender Research. In: Victoria Lee Bergvall (ed.). Longman. Chap. Shifting gender positions among Hindi-speaking hijras. <http://dx.doi.org/10.13140/2.1.3369.3760>.
- Hess, David J, Donald J Foss & Patrick Carroll. 1995. Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General* 124(1). 62–82. <http://dx.doi.org/10.1037/0096-3445.124.1.62>.
- Hughes, Ieuan A, John D Davies, Trevor I Bunch, Vickie Pasterski, Kiki Mastroyannopoulou & Jane MacDougall. 2012. Androgen insensitivity syndrome. *The Lancet* 380(9851). 1419–1428. [http://dx.doi.org/10.1016/S0140-6736\(12\)60071-3](http://dx.doi.org/10.1016/S0140-6736(12)60071-3).
- Jackendoff, Ray. 1992. Mme. tussaud meets the binding theory. *Natural Language & Linguistic Theory* 10(1). 1–31. <http://dx.doi.org/10.1007/BF00135357>.
- Johnson, Joy L & Robin Repta. 2012. Designing and conducting gender, sex, and health research. In: John L. Oliffe & Lorraine Greaves (eds.). SAGE: Thousand Oaks, CA. Chap. Sex and gender, 17–37.
- Kennison, Shelia M & Jessie L Trofe. 2003. Comprehending pronouns: a role for word-specific gender stereotype information. *Journal of Psycholinguistic Research* 32(3). 355–378. <http://dx.doi.org/10.1023/A:1023599719948>.
- Konnely, Lex & Elizabeth Cowper. 2017. The future is they: the morphosyntax of an english epicene pronoun. Ms., University of Toronto. <https://ling.auf.net/lingbuzz/003859>.
- Kreiner, Hamutal, Patrick Sturt & Simon Garrod. 2008. Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *Journal of Memory and Language* 58(2). 239–261. <http://dx.doi.org/10.1016/j.jml.2007.09.003>.
- Kuperberg, Gina R, Tatiana Sitnikova, David Caplan & Phillip J Holcomb. 2003. Electrophysiological distinctions in processing conceptual relationships within

- simple sentences. *Cognitive Brain Research* 17(1). 117–129. [http://dx.doi.org/10.1016/S0926-6410\(03\)00086-7](http://dx.doi.org/10.1016/S0926-6410(03)00086-7).
- Lasnik, Howard & Juan Uriagereka. 1988. *A course in GB syntax: lectures on binding and empty categories*. MIT Press.
- Lorber, Judith. 1996. Beyond the binaries: depolarizing the categories of sex, sexuality, and gender. *Sociological Inquiry* 66(2). 143–160. <http://dx.doi.org/10.1111/j.1475-682X.1996.tb00214.x>.
- McConnell-Ginet, Sally. 2015. The Expression of Gender. In: Greville G. Corbett (ed.). *De Gruyter Mouton Berlin*. Chap. Gender and its relation to sex: the myth of ‘natural’ gender, 3–38.
- McLemore, Kevin A. 2015. Experiences with misgendering: identity misclassification of transgender spectrum individuals. *Self and Identity* 14(1). 51–74. <http://dx.doi.org/10.1080/15298868.2014.950691>.
- Murray, David AB. 2003. Who is Takatapui? Maori language, sexuality and identity in Aotearoa/New Zealand. *Anthropologica* 42(2). 233–244. <http://dx.doi.org/10.2307/25606143>.
- Nieuwland, Mante S & Jos JA Van Berkum. 2006. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 18(7). 1098–1111. <http://dx.doi.org/10.1162/jocn.2006.18.7.1098>.
- Payne, John & Rodney D Huddleston. 2002. *The Cambridge Grammar of the English Language*. In: Rodney D Huddleston & Geoffrey Pullum (eds.). Cambridge, UK: Cambridge University Press. Chap. Nouns and noun phrases.
- Prasad, Grusha, Joanna Morris & Mark Feinstein. 2018. The P600 for singular ‘they’: how the brain reacts when John decides to treat themselves to sushi. In *The 31st CUNY Conference on Human Sentence Processing*. <https://osf.io/2vjyp/>.
- Pyykkönen, Pirita, Jukka Hyönä & Roger PG van Gompel. 2010. Activating gender stereotypes during online spoken language processing. *Experimental Psychology* 57(2). 126–133. <http://dx.doi.org/10.1027/1618-3169/a000016>.
- Spivak, Michael. 1990. *The Joy of TEX: a gourmet guide to typesetting with the AMS-TEX macro package*. American Mathematical Society.
- Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48(3). 542–562. [http://dx.doi.org/10.1016/S0749-596X\(02\)00536-3](http://dx.doi.org/10.1016/S0749-596X(02)00536-3).
- Zimman, Lal. 2017. Representing Trans: linguistic, legal and everyday perspectives. In: Evan Hazenberg & Miriam Meyerhoff (eds.). Wellington, New Zealand: Victoria University Press. Chap. Trans People’s Linguistic Self-determination and the Dialogic Nature of Identity, 226–248.