On the Relationships Between the Grammatical Genders of Inanimate Nouns and Their Co-Occurring Adjectives and Verbs

Adina Williams\textsuperscript{1} Ryan Cotterell\textsuperscript{1,2,3} Lawrence Wolf-Sonkin\textsuperscript{4} Damián Blasi\textsuperscript{5} Hanna Wallach\textsuperscript{6}

\textsuperscript{1}Facebook AI Research \textsuperscript{2}ETH Zürich \textsuperscript{3}University of Cambridge \textsuperscript{4}Johns Hopkins University \textsuperscript{5}Universität Zürich \textsuperscript{6}Microsoft Research

adinawilliams@fb.com ryan.cotterell@inf.ethz.ch lawrencews@jhu.edu damian.blasi@uzh.ch wallach@microsoft.com

Abstract

We use large-scale corpora in six different gendered languages, along with tools from NLP and information theory, to test whether there is a relationship between the grammatical genders of inanimate nouns and the adjectives used to describe those nouns. For all six languages, we find that there is a statistically significant relationship. We also find that there are statistically significant relationships between the grammatical genders of inanimate nouns and the verbs that take those nouns as direct objects, as indirect objects, and as subjects. We defer deeper investigation of these relationships for future work.

1 Introduction

In many languages, nouns possess grammatical genders. When a noun refers to an animate object, its grammatical gender typically reflects the biological sex or gender identity of that object (Zubin and Köpcke, 1986; Corbett, 1991; Kramer, 2014). For example, in German, the word for a boss is grammatically feminine when it refers to a woman, but grammatically masculine when it refers to a man—\textit{Chef} and \textit{Chefin}, respectively. But inanimate nouns (i.e., nouns that refer to inanimate objects) also possess grammatical genders. Any German speaker will tell you that the word for a bridge, \textit{Brücke}, is grammatically feminine, even though bridges have neither biological sexes nor gender identities. Historically, the grammatical genders of inanimate nouns have been considered more idiosyncratic and

\textsuperscript{*}Equal contribution in this scientific whirlwind.

less meaningful than the grammatical genders of animate nouns (Brugmann, 1889; Bloomfield, 1933; Fox, 1990; Aikhenvald, 2000). However, some cognitive scientists have reopened this discussion by using laboratory experiments to test whether speakers of gendered languages reveal gender stereotypes (Sera et al., 1994)—for example, and most famously, when choosing adjectives to describe inanimate nouns (Boroditsky et al., 2003).

Although laboratory experiments are highly informative, they typically involve small sample sizes. In this paper, we therefore use large-scale corpora and tools from NLP and information theory to test whether there is a relationship between the grammatical genders of inanimate nouns and the adjectives used to describe those nouns. Specifically, we calculate the mutual information (MI)—a measure of the mutual statistical dependence between two random variables—between the grammatical genders of inanimate nouns and the adjectives that describe them (i.e., share a dependency arc labeled AMOD) using large-scale corpora in six different gendered languages (specifically, German, Italian, Polish, Portuguese, Russian, and Spanish). For all six languages, we find that the MI is statistically significant, meaning that there is a relationship.

We also test whether there are relationships between the grammatical genders of inanimate nouns and the verbs that take those nouns as direct objects, as indirect objects, and as subjects. For all six languages, we find that there are statistically significant relationships for the verbs that take those nouns as direct objects and as subjects. For five of the six languages, we also find that there is a statistically significant relationship for the verbs that take those nouns as indirect objects, but...
because of the small number of noun–verb pairs involved, we caution against reading too much into this finding.

To contextualize our findings, we test whether there are statistically significant relationships between the grammatical genders of inanimate nouns and the cases and numbers of these nouns. A priori, we do not expect to find statistically significant relationships, so these tests can be viewed as a baseline of sorts. As expected, for each of the six languages, there are no statistically significant relationships.

To provide further context, we also repeat all tests for animate nouns—a "skyline" of sorts—finding that for all six languages there is a statistically significant relationship between the grammatical genders of animate nouns and the adjectives used to describe those nouns. We also find that there are statistically significant relationships between the grammatical genders of animate nouns and the verbs that take those nouns as direct objects, as indirect objects, and as subjects. All of these relationships have effect sizes (operationalized as normalized MI values) that are larger than the effect sizes for inanimate nouns.

We emphasize that the practical significance and implications of our findings require deeper investigation. Most importantly, we do not investigate the characteristics of the relationships that we find. This means that we do not know whether these relationships are characterized by gender stereotypes, as argued by some cognitive scientists. We also do not engage with the ways that historical and sociopolitical factors affect the grammatical genders possessed by either animate or inanimate nouns (Fodor, 1959; Ibrahim, 2014).

2 Background

2.1 Grammatical Gender

Languages lie along a continuum with respect to whether nouns possess grammatical genders. Languages with no grammatical genders, like Turkish, lie on one end of this continuum, while languages with tens of gender-like classes, like Swahili (Corbett, 1991), lie on the other. In this paper, we focus on six different gendered languages for which large-scale corpora are readily available: German, Italian, Polish, Portuguese, Russian, and Spanish—all languages of Indo-European descent. Three of these languages (Italian, Portuguese, and Spanish) have two grammatical genders (masculine and feminine), while the other two have three grammatical genders (masculine, feminine, and neuter).

All six languages exhibit gender agreement, meaning that words are marked with morphological suffixes that reflect the grammatical genders of their surrounding nouns (Corbett, 2006). For example, consider the following translations of the sentence, "The delicate fork is on the cold ground."

(1) *Die zierliche Gabel steht auf dem kalten Boden.*


The delicate fork is on the cold ground.

(2) *El tenedor delicado está en el suelo frío.*


The delicate fork is on the cold ground.

Because the German word for a fork, *Gabel*, is grammatically feminine, the German translation uses the feminine determiner, *die*. Had *Gabel* been masculine, the German translation would have used the masculine determiner, *der*. Similarly, because the Spanish word for a fork, *tenedor*, is grammatically masculine, the Spanish translation uses the masculine determiner, *el*, instead of the feminine determiner, *la*. As we explain in Section 3, we lemmatize each corpus to ensure that our tests do not simply reflect the presence of gender agreement.

2.2 Grammatical Gender & Meaning

Although some scholars have described the grammatical genders possessed by inanimate nouns as "creative" and meaningful (Grimm, 1890; Wheeler, 1899), many scholars have considered them to be idiosyncratic (Brugmann, 1889; Bloomfield, 1933) or arbitrary (Maratsos, 1979, p. 317). In an overview of this work, Dye et al. (2017) wrote, "As often as not, the languages of the world assign [inanimate] objects into seemingly arbitrary [classes]... William of
Ockham considered gender to be a meaningless, unnecessary aspect of language.’’ Bloomfield (1933) shared this viewpoint, stating that ‘‘[t]here seems to be no practical criterion by which the gender of a noun in German, French, or Latin [can] be determined.’’ Indeed, adult language learners often have particular difficulty mastering the grammatical genders of inanimate nouns (Franceschina 2005, Ch. 4, DeKeyser 2005; Montrul et al. 2008), which suggests that their meanings are not straightforward.

Even if the grammatical genders possessed by inanimate nouns are meaningless, ample evidence suggests that gender-related information may affect cognitive processes (Sera et al., 1994; Cubelli et al., 2005, 2011; Kurinski and Sera, 2011; Boutonnet et al., 2012; Saalbach et al., 2012). Typologists and formal linguists have argued that grammatical genders are an important feature for morphosyntactic processes (Corbett, 1991, 2006; Harbour et al., 2008; Harbour, 2011; Kramer, 2014, 2015), while some cognitive scientists have shown that grammatical genders can be a perceptual cue—for example, human brain responses exhibit sensitivity to gender mismatches in several different languages (Osterhout and Mobley, 1995; Hagoort and Brown, 1999; Vigliocco et al., 2002; Wicha et al., 2003, 2004; Barber et al., 2004; Barber and Carreiras, 2005; Bañón et al., 2012; Caffarra et al., 2015), and the grammatical genders of determiners and adjectives can prime nouns (Bates et al., 1996; Akhutina et al., 1999; Friederici and Jacobsen, 1999). However, the precise nature of the relationship between grammatical gender and meaning remains an open research question.

In particular, the grammatical genders possessed by inanimate nouns might affect the ways that speakers of gendered languages conceptualize the objects referred to by those nouns (Jakobson, 1959; Clarke et al., 1981; Ervin-Tripp, 1962; Konishi, 1993; Sera et al., 1994, 2002; Vigliocco et al., 2005; Bassetti, 2007)—although we note that this viewpoint is somewhat contentious (Hofstätter, 1963; Bender et al., 2011; McWhorter, 2014). Neo-Whorfian cognitive scientists hold a particularly strong variant of this viewpoint, arguing that the grammatical genders possessed by inanimate nouns prompt speakers of gendered languages to rely on gender stereotypes when choosing adjectives to describe those nouns (Boroditsky and Schmidt, 2000; Boroditsky et al., 2002; Phillips and Boroditsky, 2003; Boroditsky, 2003; Boroditsky et al., 2003; Semenuks et al., 2017). Most famously, Boroditsky et al. (2003) claim to have conducted a laboratory experiment showing that speakers of German choose stereotypically feminine adjectives to describe, for example, bridges, while speakers of Spanish choose stereotypically masculine adjectives, reflecting the fact that in German, the word for a bridge, *Brücke*, is grammatically feminine, while in Spanish, the word for a bridge, *puente*, is grammatically masculine. Boroditsky et al. (2003) took these findings to be a relatively strong confirmation of the existence of a stereotype effect—that is, that speakers of gendered languages reveal gender stereotypes when choosing adjectives to describe inanimate nouns. That said, the experiment has not gone unchallenged. Indeed, Mickan et al. (2014) reported two unsuccessful replication attempts.

### 2.3 Laboratory Experiments vs. Corpora

Traditionally, studies of grammatical gender and meaning have relied on laboratory experiments. This is for two reasons: 1) laboratory experiments can be tightly controlled, and 2) they enable scholars to measure speakers’ immediate, real-time speech production. However, they also typically involve small sample sizes and, in many cases, somewhat artificial settings. In contrast, large-scale corpora of written text enable scholars to measure even relatively weak correlations via writers’ text production in natural, albeit less tightly controlled, settings. They also facilitate the discovery of correlations that hold across languages with disparate histories, cultural contexts, and even gender systems. As a result, large-scale corpora have proven useful for studying a wide variety of language-related phenomena (e.g., Featherston and Sternewald, 2007; Kennedy, 2014; Blasi et al., 2019).

In this paper, we assume that a writer’s choice of words in written text is as informative as a speaker’s choice of words in a laboratory experiment, despite the obvious differences between these settings. Consequently, we use large-scale corpora and tools from NLP and information theory, enabling us to test for the presence of even relatively weak relationships involving the grammatical genders of inanimate nouns across multiple different gendered languages. We
therefore argue that our findings complement, rather than supersede, laboratory experiments.

2.4 Related Work

Our paper is not the first to use large-scale corpora and tools from NLP to investigate gender and language. Many scholars have studied the ways that societal norms and stereotypes, including gender norms and stereotypes, can be reflected in representations of distributional semantics derived from large-scale corpora, such as word embeddings (Bolukbasi et al., 2016; Caliskan et al., 2017; Garg et al., 2018; Zhao et al., 2018). More recently, Williams et al. (2019) found that the grammatical genders of inanimate nouns in eighteen different languages were correlated with their lexical semantics. Dye et al. (2017) used tools from information theory to reject the idea that the grammatical genders of nouns separate those nouns into coherent categories, arguing instead that grammatical genders are only meaningful in that they systematically facilitate communication efficiency by reducing nominal entropy. Also relevant to our paper is the work of Kann (2019), who proposed a computational approach to testing whether there is a relationship between the grammatical genders of inanimate nouns and the words that co-occur with those nouns, operationalized via word embeddings. However, in contrast to our findings, they found no evidence for the presence of such a relationship. Finally, many scholars have proposed a variety of computational techniques for mitigating gender norms and stereotypes in a wide range of language-based applications (Dev and Phillips, 2019; Dinan et al., 2019; Ethayarajh et al., 2019; Hall Maudslay et al., 2019; Stanovsky et al., 2019; Tan and Celis, 2019; Zhou et al., 2019; Zmigrod et al., 2019).

3 Data Preparation

We use the May 2018 dump of Wikipedia to create a corpus for each of the six different gendered languages (i.e., German, Italian, Polish, Portuguese, Russian, and Spanish). Although Wikipedia is not the most representative data source, this choice yields language-specific corpora that are roughly parallel—that is, they refer to the same objects, but are not direct translations of each other (which could lead to artificial word choices). We use UDPipe to tokenize each corpus (Straka et al., 2016).

We dependency parse the corpus for each language using a language-specific dependency parser (Andor et al., 2016; Alberti et al., 2017), trained using Universal Dependencies treebanks (Nivre et al., 2017). An example dependency tree is shown in Figure 1. We then extract all noun–adjective pairs (dependency arcs labeled AMOD) and noun–verb pairs from each of the six corpora; for verbs, we extract three types of pairs, reflecting the fact that nouns can be direct objects (dependency arcs labeled DOBJ), indirect objects (dependency arcs labeled IOBJ), or subjects (dependency arcs labeled NSUBJ) of verbs. We discard all pairs that contain a noun that is not present in WordNet (Princeton University, 2010). We label the remaining nouns as “animate” or “inanimate” according to WordNet.

Next, we lemmatize all words (i.e., nouns, adjectives, and verbs). Each word is factored into a set of lexical features consisting of a lemma, or canonical morphological form, and a bundle of three morphological features corresponding to the grammatical gender, number, and case of that word. For example, the German word for a fork, Gabel, is grammatically feminine, singular, and genitive. For nouns, we discard the lemmas themselves and retain only the morphological
features; for adjectives and verbs, we retain the lemmas and discard the morphological features.

For adjectives and verbs, lemmatizing is especially important because it ensures that our tests do not simply reflect the presence of gender agreement, as we describe in Section 2.1. However, this means that if the lemmatizer fails, then our tests may simply reflect gender agreement despite our best efforts. To guard against this, we use a state-of-the-art lemmatizer (Müller et al., 2015), trained for each language using Universal Dependencies treebanks (Nivre et al., 2017). We expect that when the lemmatizer fails, the resulting lemmata will be low frequency. We try to exclude lemmatization failures from our calculations by discarding low-frequency lemmata. For each language, we rank the adjective lemmata by their token counts and retain only the highest-ranked lemmata (in rank order) that account for 90% of the adjective tokens; we then discard all noun–adjective pairs that do not contain one of these lemmata. We repeat the same process for verbs.

Finally, to ensure that our tests reflect the most salient relationships, we also discard low-frequency inanimate nouns and, separately, low-frequency animate nouns using the same process. We provide counts of the remaining noun–adjective and noun–verb pairs in Table 3 (for inanimate nouns) and Table 4 (for animate nouns).

4 Methodology

For each language $\ell \in \{de, it, pl, pt, ru, es\}$, we define $V^\ell_{\text{ADJ}}$ to be the set of adjective lemmata represented in the noun–adjective pairs retained for that language as defined above. We similarly define $V^\ell_{\text{VERB}}$ to be the set of verb lemmata represented in the noun–verb pairs retained for that language, as described above. We then define $V^\ell_{\text{VERB-DOBJ}} \subset V^\ell_{\text{VERB}}$, $V^\ell_{\text{VERB-IOBJ}} \subset V^\ell_{\text{VERB}}$, and $V^\ell_{\text{VERB-SUBJ}} \subset V^\ell_{\text{VERB}}$ to be the sets of verbs that take the nouns as direct objects, as indirect objects, and as subjects, respectively. We also define $G^\ell$ to be the set of grammatical genders for that language (e.g., $G^{de} = \{ \text{MASC, FEM} \}$), $C^\ell$ to be the set of cases (e.g., $C^{de} = \{ \text{NOM, ACC, GEN, DAT} \}$), and $N^\ell$ to be the set of numbers (e.g., $N^{pt} = \{ \text{PL, SG} \}$). Finally, we define fourteen random variables: $A^\ell$ and $A^\ell_0$ are $V^\ell_{\text{ADJ}}$-valued random variables, $D^\ell$ and $D^\ell_0$ are $V^\ell_{\text{VERB-DOBJ}}$-valued random variables, $I^\ell$ and $I^\ell_0$ are $V^\ell_{\text{VERB-IOBJ}}$-valued random variables, $S^\ell$ and $S^\ell_0$ are $V^\ell_{\text{VERB-SUBJ}}$-valued random variables, $G^\ell_1$ and $G^\ell_0$ are $G^\ell$-valued random variables, $C^\ell_1$ and $C^\ell_0$ are $C^\ell$-valued random variables, and $N^\ell_1$ and $N^\ell_0$ are $N^\ell$-valued random variables. The subscripts ‘‘$1$’’ and ‘‘$0$’’ denote inanimate and animate nouns, respectively.

To test whether there is a relationship between the grammatical genders of inanimate nouns and the adjectives used to describe those nouns for language $\ell$, we calculate the MI (mutual information)—a measure of the mutual statistical dependence between two random variables—between $G^\ell_1$ and $A^\ell_1$:

$$\text{MI}(G^\ell_1, A^\ell_1) = \sum_{g \in G^\ell} \sum_{a \in A^\ell} P(g, a) \log_2 \frac{P(g, a)}{P(g) P(a)}, \quad (1)$$

where all probabilities are calculated with respect to inanimate nouns only. If $G^\ell_1$ and $A^\ell_1$ are independent (i.e., there is no relationship between them) then $\text{MI}(G^\ell_1, A^\ell_1) = 0$; if $G^\ell_1$ and $A^\ell_1$ are maximally dependent then $\text{MI}(G^\ell_1, A^\ell_1) = \min\{H(G^\ell_1), H(A^\ell_1)\}$, where $H(G^\ell_1)$ is the entropy of $G^\ell_1$ and $H(A^\ell_1)$ is the entropy of $A^\ell_1$. For simplicity, we use plug-in estimates for all probabilities (i.e., empirical probabilities), deferring the use of more sophisticated estimators for future work. We note that $\text{MI}(G^\ell_1, A^\ell_1)$ can be calculated in $O(\mid G^\ell \cdot \mid V^\ell_{\text{ADJ}})$ time; however, $\mid G^\ell \mid$ is negligible (i.e, two or three) so the main cost is $\mid V^\ell_{\text{ADJ}} \mid$.

To test for statistical significance, we perform a permutation test. Specifically, we permute the grammatical genders of the inanimate nouns 10,000 times and, for each permutation, recalculate the MI between $G^\ell_1$ and $A^\ell_1$ using the permuted genders. We obtain a $p$-value by calculating the percentage of permutations that have a higher MI than the MI obtained using the non-permuted genders; if the $p$-value is less than 0.05, then we treat the relationship between $G^\ell_1$ and $A^\ell_1$ as statistically significant.

Because the maximum possible MI between any pair of random variables depends on the entropies of those variables, MI values are not comparable across pairs of random variables. We therefore also calculate the normalized MI (NMI)
between $G_i^\ell$ and $A_i^\ell$ by normalizing $\text{MI}(G_i^\ell, A_i^\ell)$ to lie between zero and one. The most obvious choice of normalizer is the maximum possible MI—that is, $\min \{H(G_i^\ell), H(A_i^\ell)\}$—however, various other normalizers have been proposed, each of which has different advantages and disadvantages (Gates et al., 2019). We therefore calculate six different variants of $\text{NMI}(G_i^\ell, A_i^\ell)$ using the following normalizers:

\[
\text{MI}(G_i^\ell, A_i^\ell) = \frac{\text{MI}(G_i^\ell, A_i^\ell)}{H(G_i^\ell) + H(A_i^\ell)}
\]

\[
\text{MI}(G_i^\ell, A_i^\ell) = \frac{\text{MI}(G_i^\ell, A_i^\ell)}{2\max \{H(G_i^\ell), H(A_i^\ell)\}}
\]

\[
\text{MI}(G_i^\ell, A_i^\ell) = \frac{\text{MI}(G_i^\ell, A_i^\ell)}{\log M_i^\ell}
\]

where $M_i^\ell$ is the number of non-unique (inanimate) noun–adjective pairs retained for that language.

To test whether there are relationships between the grammatical genders of inanimate nouns and the verbs that take those nouns as direct objects, as indirect objects, and as subjects, we calculate $\text{MI}(G_i^\ell, D_i^\ell)$, $\text{MI}(G_i^\ell, I_i^\ell)$, and $\text{MI}(G_i^\ell, S_i^\ell)$. Again, all probabilities are calculated with respect to inanimate nouns only, and we perform permutation tests to test for statistical significance. We also calculate six NMI variants for each of the three pairs of random variables, using normalizers that are analogous to those in Eq. (2) through Eq. (7).

As a baseline, we test whether there are relationships between the grammatical genders of inanimate nouns and the cases and numbers of those nouns—that is, we calculate $\text{MI}(G_i^\ell, C_i^\ell)$ and $\text{MI}(G_i^\ell, N_i^\ell)$ using probabilities that are calculated with respect to inanimate nouns only. Again, we perform permutation tests (but we do not expect that there will be statistically significant relationships), and we calculate six NMI variants for each pair of random variables using normalizers that are analogous to those in Eq. (2) through Eq. (7).

Finally, we calculate $\text{MI}(G_a^\ell, A_a^\ell)$, $\text{MI}(G_a^\ell, D_a^\ell)$, $\text{MI}(G_a^\ell, I_a^\ell)$, $\text{MI}(G_a^\ell, S_a^\ell)$, $\text{MI}(G_a^\ell, C_a^\ell)$, and $\text{MI}(G_a^\ell, N_a^\ell)$ using probabilities calculated with respect to animate nouns only. The first five of these are intended to serve as a “skylines,” while the last two are intended to serve as a sanity check (i.e., we expect them to be close to zero, as with inanimate nouns). Again, we perform permutation tests to test for statistical significance, and we calculate six NMI variants for each pair of random variables.

### 5 Results

In the first row of Table 1, we provide the MI between $G_i^\ell$ and $A_i^\ell$ for each language $\ell \in \{de, it, pl, pt, ru, es\}$. For all six languages, $\text{MI}(G_i^\ell, A_i^\ell)$ is statistically significant (i.e., $p < 0.05$), meaning that there is a relationship between the grammatical genders of inanimate nouns and
the adjectives used to describe those nouns. Rows 2–4 of Table 1 contain $\text{MI}(G^i_\ell, D^i_\ell)$, $\text{MI}(G^i_\ell, I^i_\ell)$, and $\text{MI}(G^i_\ell, S^i_\ell)$ for each language. For all six languages, $\text{MI}(G^i_\ell, I^i_\ell)$ is statistically significant (i.e., $p < 0.05$). For five of the six languages, $\text{MI}(G^i_\ell, I^i_\ell)$ is statistically significant, but because of the small number of noun–verb pairs involved, we caution against reading too much into this finding. We note that direct objects are closest to verbs in analyses of constituent structures, followed by subjects and then indirect objects (Chomsky, 1957; Adger, 2003). Finally, the last two rows of Table 1 contain $\text{MI}(G^i_\ell, C^i_\ell)$ and $\text{MI}(G^i_\ell, N^i_\ell)$, respectively, for each language. We do not find any statistically significant relationships for either case or number.

To facilitate comparisons, each subplot in Figure 2 contains six variants of $\text{NMI}(G^i_\ell, A^i_\ell)$, $\text{NMI}(G^i_\ell, D^i_\ell)$, and $\text{NMI}(G^i_\ell, S^i_\ell)$—one per normalizer—for a single language $\ell \in \{de, it, pl, pt, ru, es\}$.
and NMI\((G_{a}^\ell, S_{i}^a)\) for the normalizer in Eq. (2), while NMI\((G_{d}^\ell, A_{i}^d)\), NMI\((G_{de}^\ell, D_{i}^{de})\), and NMI\((G_{de}^\ell, S_{i}^{de})\) are all roughly comparable for the other five normalizers. Finally, NMI\((G_{it}^\ell, A_{i}^u)\) and NMI\((G_{it}^\ell, D_{i}^{it})\) are roughly comparable and larger than NMI\((G_{it}^\ell, S_{i}^{it})\), regardless of the normalizer.

In other words, the relationship between the grammatical genders of inanimate nouns and the adjectives used to describe those nouns is generally stronger than, but sometimes roughly comparable to, the relationships between the grammatical genders of animate nouns and the verbs that take those nouns as direct objects and as subjects. However, the relative strengths of the relationships between the grammatical genders of inanimate nouns and the verbs that take those nouns as direct objects, as indirect objects, and as subjects depend on the language.

In Table 2, we provide MI\((G_{a}^\ell, A_{i}^d)\), MI\((G_{d}^\ell, D_{i}^{d})\), MI\((G_{de}^\ell, I_{i}^{de})\), MI\((G_{a}^\ell, S_{i}^a)\), MI\((G_{a}^\ell, C_{i}^a)\), and MI\((C_{a}^\ell, N_{i}^a)\) for each language \(\ell \in \{de, it, pl, pt, ru, es\}\). As with inanimate nouns, we find that there is a statistically significant relationship between the grammatical genders of animate nouns and the adjectives used to describe those nouns. We also find that there are statistically significant relationships between the grammatical genders of animate nouns and the verbs that take those nouns as direct objects, as indirect objects, and as subjects. Again, the relationship for the verbs that take those nouns as indirect objects involves a small number of noun–verb pairs. As expected, we do not find any statistically significant relationships for either case or number.

Table 2: The mutual information (MI) between the grammatical genders of animate nouns and a) the adjectives used to describe those nouns (top row), b) the verbs that take those nouns as direct objects, as indirect objects, and as subjects (rows 2–4, respectively), and c) the cases and numbers of those nouns (rows 5 and 6, respectively) for six different gendered languages. Statistical significance (i.e., a p-value less than 0.05) is indicated using bold. MI values are not comparable across pairs of random variables.
Figure 3: The normalized mutual information (NMI) between the grammatical genders of animate nouns and a) the adjectives used to describe those nouns and b) the verbs that take those nouns as direct objects and as subjects for six different gendered languages. Each subplot contains six variants of NMI($G^\ell_a, A^\ell_a$), NMI($G^\ell_a, D^\ell_a$), and NMI($G^\ell_a, S^\ell_a$)—one per normalizer—for a single language $\ell \in \{de, it, pl, pt, ru, es\}$. 
Figure 4: The normalized mutual information (NMI) between the grammatical genders of a) inanimate and b) animate nouns and the adjectives used to describe those nouns. Each subplot contains NMI($G^f_{\ell}, A^f_{\ell}$) and NMI($G^a_{\ell}, A^a_{\ell}$), calculated using a single normalizer, for each language $\ell \in \{de, it, pl, pt, ru, es\}$. 

**Eq. 2**

**Eq. 3**

**Eq. 4**

**Eq. 5**

**Eq. 6**

**Eq. 7**

Figure 4: The normalized mutual information (NMI) between the grammatical genders of a) inanimate and b) animate nouns and the adjectives used to describe those nouns. Each subplot contains NMI($G^f_{\ell}, A^f_{\ell}$) and NMI($G^a_{\ell}, A^a_{\ell}$), calculated using a single normalizer, for each language $\ell \in \{de, it, pl, pt, ru, es\}$. 
Figure 5: The normalized mutual information (NMI) between the grammatical genders of a) inanimate and b) animate nouns and the verbs that take those nouns as direct objects. Each subplot contains NMI($G^\ell_i, D^\ell_i$) and NMI($G^\ell_a, D^\ell_a$), calculated using a single normalizer, for each language $\ell \in \{de, it, pl, pt, ru, es\}$. 
Figure 6: The normalized mutual information (NMI) between the grammatical genders of a) inanimate and b) animate nouns and the verbs that take those nouns as subjects. Each subplot contains NMI($G_{a, \ell}, S_{a, \ell}$) and NMI($G_{i, \ell}, S_{i, \ell}$), calculated using a single normalizer, for each language $\ell \in \{de, it, pl, pt, ru, es\}$. 
than the NMI values for inanimate nouns. The only exception is Polish, where $\text{NMI}(G_{i}^{pl}, A_{i}^{pl})$ is larger than $\text{NMI}(G_{a}^{pl}, A_{a}^{pl})$, regardless of the normalizer.

6 Discussion

We find evidence for the presence of a statistically significant relationship between the grammatical genders of inanimate nouns and the adjectives used to describe those nouns for six different gendered languages (specifically, German, Italian, Polish, Portuguese, Russian, and Spanish). We also find evidence for the presence of statistically significant relationships between the grammatical genders of inanimate nouns and the verbs that take those nouns as direct objects, as indirect objects, and as subjects. However, we caution against reading too much into the relationship for the verbs that take those nouns as indirect objects because of the small number of noun–verb pairs involved. The effect sizes (operationalized as NMI values) for all of these relationships are smaller than the effect sizes for animate nouns. As expected, we do not find any statistically significant relationships for either case or number.

We emphasize that our findings complement, rather than supersede, laboratory experiments, such as that of Boroditsky et al. (2003). We use large-scale corpora and tools from NLP and information theory to test for the presence of even relatively weak relationships across multiple different gendered languages—and, indeed, the relationships that we find have effect sizes (operationalized as NMI values) that are small. In contrast, laboratory experiments typically focus on much stronger relationships by tightly controlling experimental conditions and measuring speakers’ immediate, real-time speech production. Moreover, although we find statistically significant relationships, we do not investigate the characteristics of these relationships. This means that we do not know whether they are characterized by gender stereotypes, as argued by some cognitive scientists, including Boroditsky et al. (2003). We also do not know whether the relationships that we find are causal in nature. Because MI is symmetric, our findings say nothing about whether the grammatical genders of inanimate nouns cause writers to choose particular adjectives or verbs. We defer deeper investigation of this for future work.

We note that each of our tests can be viewed as a comparison of the similarity of two clusterings of a set of items—specifically, a “clustering” of nouns into grammatical genders and a “clustering” of the same nouns into, for example, adjective lemmata. Although (normalized) MI is a standard measure for comparing clusterings, it is not without limitations (see, e.g., Newman et al. [2020] for an overview). For future work, we therefore recommend replicating our tests using other information-theoretic measures for comparing clusterings.

Acknowledgments

We thank Lera Boroditsky, Hagen Blix, Eleanor Chodoroff, Andrei Cimpian, Zach Davis, Jason Eisner, Richard Futrell, Todd Gureckis, Katharina Kann, Peter Klecha, Zhiwei Li, Ethan Ludwin-Peery, Alec Marantz, Arya McCarthy, John McWhorter, Sabrina J. Mielke, Elizabeth Salesky, Arturs Semenuks, and Colin Wilson for discussions at various points related to the ideas in this paper. Katharina Kann approves this acknowledgment.

A Appendix A: Counts

Counts of the noun–adjective and noun–verb pairs for all six gendered languages are in Table 3 (for inanimate nouns) and Table 4 (for animate nouns).
<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>it</th>
<th>pl</th>
<th>pt</th>
<th>ru</th>
<th>es</th>
</tr>
</thead>
<tbody>
<tr>
<td># noun–adj. tokens</td>
<td>6443907</td>
<td>6246856</td>
<td>11631913</td>
<td>640558</td>
<td>32900200</td>
<td>3605439</td>
</tr>
<tr>
<td># noun–adj. types</td>
<td>770952</td>
<td>666656</td>
<td>640107</td>
<td>638774</td>
<td>1633963</td>
<td>368795</td>
</tr>
<tr>
<td># noun types</td>
<td>10712</td>
<td>6410</td>
<td>5533</td>
<td>5672</td>
<td>9327</td>
<td>6157</td>
</tr>
<tr>
<td># adj. types</td>
<td>4129</td>
<td>3607</td>
<td>4080</td>
<td>3431</td>
<td>11028</td>
<td>1907</td>
</tr>
<tr>
<td># noun–verb (subj.) tokens</td>
<td>3191030</td>
<td>1432354</td>
<td>2179396</td>
<td>1871941</td>
<td>6007063</td>
<td>1534211</td>
</tr>
<tr>
<td># noun–verb (subj.) types</td>
<td>445536</td>
<td>292949</td>
<td>297996</td>
<td>337262</td>
<td>864480</td>
<td>376888</td>
</tr>
<tr>
<td># noun (subj.) types</td>
<td>10741</td>
<td>6318</td>
<td>5522</td>
<td>5780</td>
<td>9129</td>
<td>7470</td>
</tr>
<tr>
<td># verb types</td>
<td>707</td>
<td>702</td>
<td>874</td>
<td>758</td>
<td>1803</td>
<td>875</td>
</tr>
<tr>
<td># noun–verb (dobj.) tokens</td>
<td>3440922</td>
<td>2855037</td>
<td>3964828</td>
<td>4850012</td>
<td>6738606</td>
<td>2859135</td>
</tr>
<tr>
<td># noun–verb (dobj.) types</td>
<td>427441</td>
<td>393246</td>
<td>236849</td>
<td>541347</td>
<td>713703</td>
<td>576835</td>
</tr>
<tr>
<td># noun (dobj.) types</td>
<td>10504</td>
<td>6407</td>
<td>4359</td>
<td>5896</td>
<td>8998</td>
<td>11567</td>
</tr>
<tr>
<td># verb types</td>
<td>805</td>
<td>806</td>
<td>708</td>
<td>738</td>
<td>1539</td>
<td>9746</td>
</tr>
<tr>
<td># noun–verb (iobj.) tokens</td>
<td>163935</td>
<td>71</td>
<td>54138</td>
<td>95009</td>
<td>1570273</td>
<td>56038</td>
</tr>
<tr>
<td># noun–verb (iobj.) types</td>
<td>50133</td>
<td>53</td>
<td>18214</td>
<td>39738</td>
<td>300703</td>
<td>24830</td>
</tr>
<tr>
<td># noun (iobj.) types</td>
<td>5520</td>
<td>59</td>
<td>2258</td>
<td>3757</td>
<td>8150</td>
<td>3574</td>
</tr>
<tr>
<td># verb types</td>
<td>386</td>
<td>68</td>
<td>417</td>
<td>357</td>
<td>1816</td>
<td>464</td>
</tr>
<tr>
<td># noun–case tokens</td>
<td>14681293</td>
<td>N/A</td>
<td>15300621</td>
<td>N/A</td>
<td>51641929</td>
<td>N/A</td>
</tr>
<tr>
<td># noun–case types</td>
<td>2252632</td>
<td>N/A</td>
<td>1465314</td>
<td>N/A</td>
<td>5028075</td>
<td>N/A</td>
</tr>
<tr>
<td># noun types</td>
<td>11998</td>
<td>N/A</td>
<td>5839</td>
<td>N/A</td>
<td>9692</td>
<td>N/A</td>
</tr>
<tr>
<td># case types</td>
<td>4</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td># noun–number tokens</td>
<td>14681293</td>
<td>11588448</td>
<td>15300621</td>
<td>14631732</td>
<td>51641929</td>
<td>5672790</td>
</tr>
<tr>
<td># noun–number types</td>
<td>2252632</td>
<td>1748927</td>
<td>1465314</td>
<td>2042626</td>
<td>5028075</td>
<td>1034307</td>
</tr>
<tr>
<td># noun types</td>
<td>11998</td>
<td>7014</td>
<td>5839</td>
<td>6256</td>
<td>9692</td>
<td>1593</td>
</tr>
<tr>
<td># number types</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3: Counts of the inanimate noun–adjective and noun–verb pairs for all six gendered languages.
Table 4: Counts of the animate noun–adjective and noun–verb pairs for all six gendered languages.

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>it</th>
<th>pl</th>
<th>pt</th>
<th>ru</th>
<th>es</th>
</tr>
</thead>
<tbody>
<tr>
<td># noun–adj. tokens</td>
<td>662760</td>
<td>818300</td>
<td>1137209</td>
<td>712101</td>
<td>3225932</td>
<td>387025</td>
</tr>
<tr>
<td># noun–adj. types</td>
<td>99332</td>
<td>92424</td>
<td>97847</td>
<td>90865</td>
<td>264117</td>
<td>50173</td>
</tr>
<tr>
<td># noun types</td>
<td>1998</td>
<td>1078</td>
<td>954</td>
<td>1006</td>
<td>2098</td>
<td>1320</td>
</tr>
<tr>
<td># adj. types</td>
<td>3587</td>
<td>3507</td>
<td>3836</td>
<td>3176</td>
<td>9833</td>
<td>1828</td>
</tr>
<tr>
<td># noun–verb (subj.) tokens</td>
<td>637801</td>
<td>399747</td>
<td>526894</td>
<td>456349</td>
<td>1516740</td>
<td>310569</td>
</tr>
<tr>
<td># noun–verb (subj.) types</td>
<td>113308</td>
<td>77551</td>
<td>89819</td>
<td>89959</td>
<td>253150</td>
<td>93586</td>
</tr>
<tr>
<td># noun (subj.) types</td>
<td>2056</td>
<td>1066</td>
<td>969</td>
<td>1013</td>
<td>2020</td>
<td>1477</td>
</tr>
<tr>
<td># verb types</td>
<td>707</td>
<td>702</td>
<td>874</td>
<td>758</td>
<td>1799</td>
<td>874</td>
</tr>
<tr>
<td># noun–verb (dobj.) tokens</td>
<td>321400</td>
<td>388187</td>
<td>456824</td>
<td>527259</td>
<td>494534</td>
<td>850234</td>
</tr>
<tr>
<td># noun–verb (dobj.) types</td>
<td>60760</td>
<td>55574</td>
<td>76348</td>
<td>92220</td>
<td>118818</td>
<td>85235</td>
</tr>
<tr>
<td># noun (dobj.) types</td>
<td>1901</td>
<td>1025</td>
<td>867</td>
<td>1028</td>
<td>1912</td>
<td>1023</td>
</tr>
<tr>
<td># verb types</td>
<td>804</td>
<td>805</td>
<td>724</td>
<td>737</td>
<td>1535</td>
<td>745</td>
</tr>
<tr>
<td># noun–verb (iobj.) tokens</td>
<td>51359</td>
<td>7</td>
<td>43187</td>
<td>23139</td>
<td>518540</td>
<td>23955</td>
</tr>
<tr>
<td># noun–verb (iobj.) types</td>
<td>17804</td>
<td>6</td>
<td>8440</td>
<td>110185</td>
<td>11353</td>
<td>9586</td>
</tr>
<tr>
<td># noun (iobj.) types</td>
<td>1149</td>
<td>6</td>
<td>628</td>
<td>773</td>
<td>1858</td>
<td>947</td>
</tr>
<tr>
<td># verb types</td>
<td>378</td>
<td>6</td>
<td>411</td>
<td>340</td>
<td>1769</td>
<td>456</td>
</tr>
<tr>
<td># noun–case tokens</td>
<td>1926614</td>
<td>N/A</td>
<td>1907688</td>
<td>N/A</td>
<td>6357089</td>
<td>N/A</td>
</tr>
<tr>
<td># noun–case types</td>
<td>390672</td>
<td>N/A</td>
<td>299511</td>
<td>N/A</td>
<td>987420</td>
<td>N/A</td>
</tr>
<tr>
<td># noun types</td>
<td>2292</td>
<td>N/A</td>
<td>1024</td>
<td>N/A</td>
<td>2194</td>
<td>N/A</td>
</tr>
<tr>
<td># case types</td>
<td>4</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td># noun–number tokens</td>
<td>1926614</td>
<td>1801285</td>
<td>1907688</td>
<td>1931315</td>
<td>6357089</td>
<td>786177</td>
</tr>
<tr>
<td># noun–number types</td>
<td>390672</td>
<td>306968</td>
<td>299511</td>
<td>356352</td>
<td>987420</td>
<td>200785</td>
</tr>
<tr>
<td># noun types</td>
<td>2292</td>
<td>1135</td>
<td>1024</td>
<td>1072</td>
<td>2194</td>
<td>1593</td>
</tr>
<tr>
<td># number types</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
References


pages 1696–1705, Florence, Italy. Association for Computational Linguistics. DOI: https://doi.org/10.18653/v1/P19-1166


Jacob Grimm. 1890. *Deutsche Grammatik*, C. Bertelsmann.


