Title

Testing a Distributional Semantics Account of Grammatical Gender Effects on Semantic Gender Perception

Permalink

https://escholarship.org/uc/item/44j035tc

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Flint, George Rocco
Ivanova, Anna

Publication Date

2024

Copyright Information

Testing a Distributional Account of Grammatical Gender Effects on Semantic Gender Perception

George Flint
georgeflint@berkeley.edu
Cognitive Science
University of California, Berkeley

Anna A. Ivanova
a.ivanova@gatech.edu
School of Psychology
Georgia Institute of Technology

Abstract

One well-known prediction of linguistic relativity theories is the effect of a noun's grammatical gender on its semantics; for instance, ”key” is feminine in Spanish but masculine in German and thus might be associated with feminine traits for Spanish speakers but with masculine traits for German speakers. Experimental and corpus evidence for these effects has been mixed. In this work, we considered a distributional semantics account of putative grammatical gender effects on semantics and tested its predictions in Spanish, German, and English (control). In Part 1, we hypothesized that grammatical gender of concrete nouns affects the similarity of noun embeddings to embeddings of adjectives semantically associated with men or with women. We found support for this hypothesis in *fastText* embeddings, showing that nouns with the same meaning but with opposite genders in Spanish and German show opposite attraction effects both for words ”man” and ”woman” and for adjectives associated with men and women, although the effect size was weaker for German than for Spanish. *BERT* embeddings also showed consistent effects for Spanish but mixed results for German, suggesting possible variation across languages. In Part 2, we asked whether people systematically choose adjectives associated with women/men for grammatically feminine/masculine nouns, respectively. In a noun-adjective matching experiment (432 participants total), we found predicted grammatical gender effects for Spanish but not for German. Cosine similarity between the noun and the adjectives in *fastText* embeddings significantly predicted trial-level responses in all 3 languages; however, Spanish showed an additional effect of grammatical gender, indicating that participant noun-adjective associations are not fully explained by distributional semantics.

Keywords: linguistic relativity; grammatical gender; distributional semantics; cross-linguistic comparison

Introduction

The effects of language on thought (linguistic relativity, or the Sapir-Whorf hypothesis) have been the subject of many heated debates. One reason why the topic has been controversial is the fact that the evidence for some Whorfian claims has failed to replicate, casting doubt on the whole research program. Here, we provide a new framework for investigating a commonly discussed linguistic relativity effect—-the effect of grammatical gender on semantic content.

We are re-examining the effect of grammatical gender on semantic perception of concrete nouns reported by Boroditsky, Schmidt, and Phillips (2003), who explored Spanish-English and German-English bilingual speakers' perception of nouns that have opposite genders in Spanish or German by asking participants which adjectives they associated with each noun and quantifying the extent to which these adjectives were associated with men or women. The authors claimed that speakers used semantically gendered adjectives corresponding to the grammatical gender of the noun in their native language (e.g., ”elegant” for the feminine ”llave”[”key”] in Spanish). However, the full study was never published, and a later study by Mickan, Schiefke, and Stefanowitsch (2014) failed to replicate these findings. Another study failed to replicate a related experiment where participants paired pictures of objects with pictures of either a man or a woman (Elpers, Jensen, & Holmes, 2022), but showed a significant effect in an artificial learning paradigm.

A complementary way to examine the relationship between grammatical gender of nouns and their semantics is corpus analyses. Corpus analyses can be used to test whether there is significant co-occurrence between nouns of a given grammatical gender (e.g., feminine) and adjectives stereotypically associated with that gender (e.g., “delicate”). Using this approach, Williams, Cotterell, Wolf-Sonkin, Blasi, and Wallach (2021) indeed found a significant association between the two, although a recent follow-up by Stańczak, Du, Williams, Augenstein, and Cotterell (2023) suggests that this relationship disappears once the lexical semantics of the nouns is controlled for. Both the experimental and the corpus evidence therefore remains mixed.

Here, we approach the question of grammatical gender effects on semantics from a distributional semantics angle (Lenci et al., 2008; Bhatia, Richie, & Zou, 2019). We propose that one possible way by which a noun's grammatical gender might have semantic effects is by warping the distributional semantic space, such that adjectives semantically related to man/woman would be located closer to grammatically masculine/feminine nouns respectively. This distributional effect might then influence human behavior, such that people would be more likely to select a woman-related adjective (over a man-related adjective) for a grammatically feminine noun simply because they are closer to each other and are therefore perceived as more similar. In contrast to Boroditsky et al. (2003), we test speakers of gendered languages on materials from that same language. As such, our results can provide insights into intra-linguistic effects of grammatical gender, but not into cross-linguistic effects.

The contributions of this work are fourfold. First, we explore a procedure for identifying adjectives that are semantically associated with men or women by selecting words in

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

the distributional semantic neighborhood of man vs. woman. Second, we test the hypothesis that grammatical gender of nouns affects their embeddings' cosine similarity with embeddings of adjectives that are semantically associated with man vs. woman. Third, we use a binary choice noun-adjective matching paradigm to test the effects of noun grammatical gender on adjective choice in an intralinguistic setting. And fourth, we test whether participant responses in the noun-adjective matching paradigm can be predicted from noun-adjective embedding similarities. We examine these relationships in two gendered languages—German and Spanish—and a non-gendered language, English, as a control.

Part 1: Distributional semantics analysis

Here, we test the hypothesis that grammatical gender and semantic gender are linked in the distributional semantic space. We break down this hypothesis into sub-hypotheses as follows (a) there is an interaction between noun grammatical gender (masculine vs. feminine) and their embedding similarity with words "man" and "woman", such that grammatically masculine nouns are closer to "man" and vice versa; (b) person-describing adjectives sampled from the distributional semantic neighborhood of "man" and "woman" are semantically associated with men and women, as measured via behavioral ratings; (c) grammatical gender of nouns modulates their similarity to adjectives semantically associated with men and women.

Method

We used *fastText* (Bojanowski, Grave, Joulin, & Mikolov, 2017) and *BERT* (Devlin, Chang, Lee, & Toutanova, 2019) word embedding models for English, Spanish, and German, and cosine similarity between vectors as our measure of semantic similarity.

Adjective selection. We used a *BeautifulSoup* crawler to extract adjectives from Wiktionary's list of adjectives for each language. We then selected a set of adjectives with the highest similarity scores to the word for "man" and "woman" in each language, filtering out rare, archaic, colloquial, offensive, racial, ethnic, or national adjectives. As a result, we obtained 154, 148, and 140 adjectives for English, Spanish, and German. Half of the adjectives for each language were in the masculine group and half were in the feminine group. For German, adjectives were in their dictionary, gender-neutral forms. For Spanish, adjectives of both masculine and feminine forms could be included. To prevent the influence of the confound of grammatical gender of adjectives on their cosine similarity to nouns, whenever a Spanish adjective had different masculine and feminine forms, we included the opposite grammatical gender form of that adjective along with the original.

Noun selection. We use 36 nouns that have opposite genders in Spanish and German (feminine vs. masculine; German's neuter nouns were not included). Nouns were selected from the highest rated words for concreteness from a set of English lemmas in (Brysbaert, Warriner, & Kuperman, 2014)., filtered via *NLTK* part-of-speech-tagger to include only nouns. We also included some nouns from Study 1 of Elpers et al. (2022).

Table 1: Sample nouns, chosen to have the same meanings but opposite grammatical genders in Spanish and German.

Group	English	Spanish	German
es.M-de.F	bridge	puente	Brücke
es.F-de.M	key	llave	Schlüssel
es.M-de.F	sun	sol	Sonne
es.F-de.M	moon	luna	Mond

Human ratings. We ran an Qualtrics survey via the Prolific crowdsourcing platform where participants (n=22 for each language) rated the selected adjectives for English, Spanish, and German on a scale of 1 to 7 (1=most feminine, 4=neutral, 7=most masculine). Participants also rated whether each adjective could be used to describe a person or to describe an object. For each question, participants could indicate that they did not know the word and abstain from responding. For Spanish, if a given adjective had different masculine and feminine forms, participants were presented with both forms of that adjective, separated by a slash and in randomized order. After excluding participants who failed the attention checks or gave inconsistent "I don't know" responses, 19, 20, and 21 participants were left for English, Spanish, and German, respectively. Participants for English were selected from the United States and prescreened to only speak English; for Spanish and German, they were selected worldwide and prescreened to have a primary language of Spanish and German respectively. Participants were sampled to be half male and half female. All participants consented to the study; the study protocol was approved by the McGovern Institute for Brain Research at MIT. Participants were paid at the rate of $12/hour.

Analyses. We analyze the data with mixed effects regression models using the *lme4* R package (Bates, Mächler, Bolker, & Walker, 2014) and report the p-values derived with the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). We use dummy coding with English as the reference level for the language predictor; for other categorical predictors, we use sum coding. We treat individual nouns and adjectives as random effects. We fit the following models: (1) *similarity* ∼ *language* * *target concept* * *noun group* + *(1 + language* | *noun concept)*, where target concept is "man"/"woman", noun group is the group based on grammatical gender (see above), and noun concept are individual noun meanings; (2) *mean rating* ∼ *language* * *adjective group* + *(1* | *adjective)*, where adjective group are adjectives in the semantic neighborhood of man/woman; (3) *similarity* ∼ *language* * *noun group* * *adjective group* + *(1 + language* | *noun concept)* + *(1* | *adjective)*, where adjective group has been additionally filtered for gender association ratings (see below).

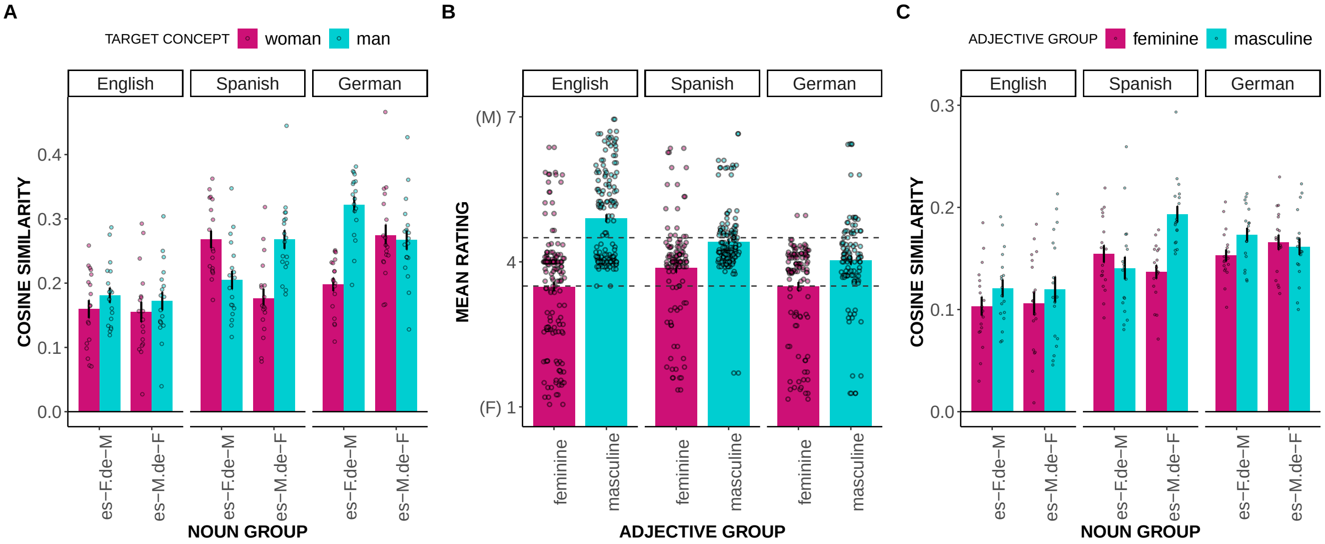

Figure 1: The relationship between grammatical gender and semantic gender associations in the distributional semantic space of *fastText* embeddings. Noun groups: feminine in German, masculine in Spanish (*es-M.de-F*) and masculine in German, feminine in Spanish (*es-F.de-M*). (A) Grammatical gender modulates cosine similarity between concrete inanimate nouns and words ”man” and ”woman”; as expected, the effect goes in opposite direction for words of opposite genders in Spanish and German. (B) Adjectives in the semantic neighborhood of words “man” and “woman” are associated with men and women, respectively. Dashed lines denote our selection thresholds for the next analysis (above 4.5 for masculine and below 3.5 for feminine). (C) Grammatical gender modulates cosine similarity between concrete inanimate nouns and adjectives stereotypically associated with men and women. In all plots, errorbars show standard error of the mean.

Results

Nouns. We compared cosine similarities to two target words, “man” and “woman”, for 2 groups of nouns: feminine in German, masculine in Spanish (*es-M.de-F*) and masculine in German, feminine in Spanish (*es-F.de-M*). In English, words from both groups were closer to “man” than to “woman” (β=-.01, *p*=.006), but there was no difference between groups, no difference in the man-woman effect between English and Spanish, and an even higher difference in similarity scores for man vs. woman in German (β=-.02, *p* <.001).

Critically, we observed an interaction between language, noun group, and relative distance to target words for both Spanish (β=.04, *p* <.001) and German (β=-.03, *p* <.001) relative to English (Figure 1A). As predicted, these effects go in opposite directions, such that nouns with the same semantic content are pulled toward “man” in a language where they are grammatically masculine and toward “woman” in a language where they are grammatically feminine.

Adjectives. We tested a new way to source adjectives that have masculine and feminine semantic associations by selecting adjectives with the highest similarity scores to either “man” or “woman”. Our method yielded a set of adjectives, many of which can be used to describe people (English: 88.31%, Spanish: 83.78%, German: 65.71%). Adjectives that do not describe people were excluded from subsequent analyses.

A comparison of human semantic gender association ratings for selected adjective groups (Figure 1B) showed that, as predicted, adjectives closest to “man” were more strongly associated with men than adjectives closes to “woman” for English (β=-.53, *p* <.001); the effect was present even more strongly for Spanish (interaction with English: β=-.04, *p*=.002) and somewhat, but not fully, dampened for German (interaction with English: β=.37, *p* <.001). That said, this adjective selection process was noisy and sometimes yielded counter-intuitive results, such as ”patriarchal” being selected for the English-feminine list and ”fraulich” (”feminine”) being selected for the German-masculine list. For the next analysis, we have selected a subset of adjectives that was both in the semantic neighborhood of “man” and “woman” and had high gender association scores (>4.5 for masculine and <3.5 for feminine).

Noun-adjective similarities. We compared cosine similarities for nouns from our opposing grammatical gender groups (*es-M.de-F* and *es-F.de-M*) and adjectives with feminine or masculine semantic associations sourced in the previous step. In English, nouns were on average closer to masculine-associated adjectives than to feminine-associated adjectives (β=-.007, *p*.032), with no difference between noun groups.

In accordance with our hypothesis, we found an interaction between language, adjective group, and noun group for both Spanish relative to English (β=-.02, *p* <.001) and German

Table 2: Adjectives used in Part 2. Participants had to match a noun with an adjective from the masculine group or an adjective from the feminine group.

Assoc.	English	Spanish	German
Masc.	grandfatherly	masculino	männisch
	manly	macho	hochtechnologisch
	hunky	forzudo	ungestalt
	handsome	primitivo	sterbend
	brawny	vaquero	allein
	dapper	poderoso	beruflich
Fem.	sexy	sentimental	sentimentalisch
	statuesque	despampanante	spindeldürr
	full-figured	sexy	zärtlich
	voluptuous	sensual	bildhübsch
	feminine	virgen	fraulich
	grandmotherly	femenino	mütterlich

relative to English (β=.005, $p <$.004) in *fastText* embeddings. The effect goes in opposite directions for the two languages because the *es-M.de-F* group is more attracted to semantically masculine adjectives for Spanish relative to English but more attracted to semantically feminine adjectives for German relative to English. Although the effect is statistically significant for both languages, the effect size is much lower for German. In*BERT* embeddings, we observed the same interaction at significance in Spanish but mixed results in German. In most configurations, the interaction between adjective group and noun group in German did not reach significance.

In summary, we find support for our hypothesis: grammatical gender and semantic gender association interact in the distributional semantic space—for Spanish and German in *fastText* embeddings and Spanish in *BERT* embeddings.

Part 2: Noun-adjective matching experiment

Here, we test an experimental prediction of linguistic relativity theories with respect to grammatical gender: that people will associate concrete inanimate nouns of a given gender with an adjective that is semantically associated with that gender. To test it, we conducted an experiment where, in each trial, participants were presented with a noun, a masculine-associated adjective and a feminine-associated adjective, and asked to select an adjective that goes better with that noun. If grammatical gender of the noun affects binary choice responses, we expect grammatically feminine adjectives to be paired more often with feminine-associated nouns and vice versa for grammatically masculine adjectives. By using words with the same meanings but opposite genders in Spanish and German, we ensure that the effects we might observe are not simply driven by semantic differences between the two noun groups.

Method

Materials selection. We used human adjective ratings from Study 1 to select the top 6 most strongly rated adjectives for each gender association group for each language, filtering also for person and object describability (adjectives must be rated below 1.6 on average for both measures, where 1 is "Yes" and 2 is "No"). In Spanish, if a given adjective had different masculine and feminine forms, when paired with a given noun, the grammatical gender of that adjective agreed with that of the noun. German adjectives remained in dictionary form in all noun pairings. These adjectives are shown in Table 2. We used the same set of nouns as Part 1.

Pairing creation. Within each language, we created each combination of masculine and feminine adjective, and then match one noun to each of these combinations to create 36 triplets of a noun, masculine adjective, and feminine adjective. We create 5 more of these sets of 36 triplets, each with one Latin square design shift of concepts.

Survey format. We deployed another XM Qualtrics survey via Prolific (n=150, 170, 180 for English, Spanish, and German respectively) wherein participants are presented with one of the sets of 36 triplets and, for each triplet, were asked to choose which of the two presented adjectives they would use to describe the presented noun. Participants were able to select one adjective, the other, or a third option, ”I don't know one or more of these words.”

Participant selection. Participants were selected on Prolific via the following criteria: For English, participants were selected the United States provided their first language was English; For Spanish, participants were selected worldwide provided the first language was Spanish; for German, participants were selected worldwide provided their first language was German. Participants were sampled to be half male and half female. All participants consented to the study; the study protocol was approved by the [redacted institution]. Participants were paid at the rate of $12/hour.

Data exclusion. In addition to the 36 triplets in each set of the stimulus data for the survey of each language, we added two attention check questions where participants were asked to choose between ”masculine” or ”feminine” for ”man” and ”woman” (and equivalents in each language), dispersed randomly throughout each survey. If a participant did not answer either one of these questions with the expected answers (i.e. ”masculine” for ”man” and vice versa), or if a participant selected the unknown word option for more than one third of responses, their responses were excluded from analyses.

After exclusions of 10, 21, and 37 participants, 140, 149, and 143 participants' responses remained in English, Spanish, and German respectively for analyses.

Analyses. We used mixed effects modeling as described in Part 1. Given that participants' responses in this experiment are binary, we used logistic regression models. When comparing two nested models, we used the likelihood ratio test as implemented with the *anova* function in R. The full model is specified as *Rating* $\sim$ *language* * *noun group* + *lan-*

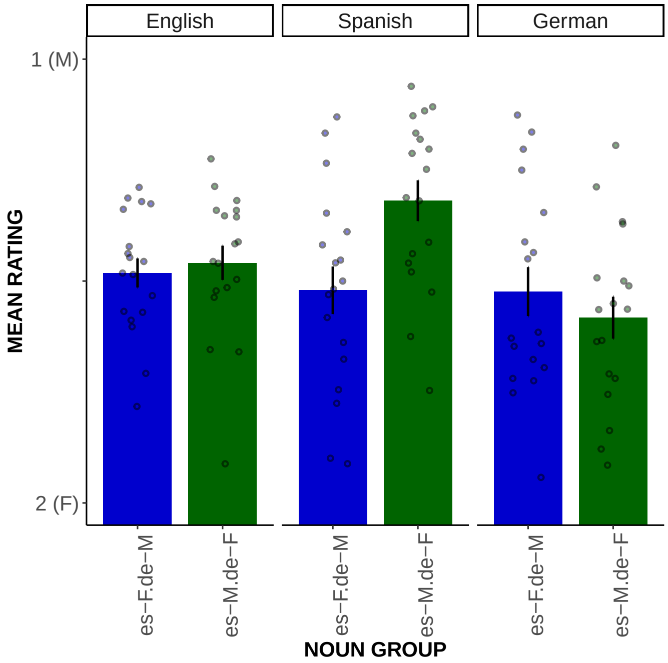

Figure 2: Results of the noun-adjective matching experiment. Each dot denotes a noun; the y axis indicates a ratio of masculine vs. feminine adjectives chosen for that noun (top: all masculine, bottom: all feminine).

Table 3: Participants' preference for masculine vs. feminine adjective can be predicted by the cosine distance between the noun and the adjectives. cosine2masc and cosine2fem denote cosine similarity (*fastText*) between the noun and the masculine and feminine adjectives, respectively. The data are modeled on a per-trial level. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

	Estimate	SE	p value
(Intercept)	-0.03	0.12	0.796
Language - Spanish	-0.38	0.15	0.011*
Language - German	0.49	0.18	0.006**
Noun Group	-0.06	0.10	0.542
Question #	-0.08	0.04	0.021*
Spanish:Noun Group	-0.38	0.13	0.003**
German:Noun Group	0.07	0.16	0.660
English:cosine2masc	-0.17	0.08	0.034*
Spanish:cosine2masc	-0.20	0.06	<.001***
German:cosine2masc	-0.84	0.09	<.001***
English:cosine2fem	0.39	0.04	<.001***
Spanish:cosine2fem	0.33	0.06	<.001***
German:cosine2fem	0.14	0.04	0.001***

guage : cosine2masc + language : cosine2fem + question number + (1 + language | noun concept) + (1 + question | participant), where cosine2masc and cosine2fem are cosine similarities between the noun and the masculine and feminine adjectives in each trial. Cosine similarities and question number were centered and scaled prior to fitting.

Results

Effect of noun's grammatical gender on adjective selection. Participants had no general preference for semantically masculine vs. semantically feminine adjectives for English (intercept β=-.11, n.s.), although there was a stronger preference for masculine adjectives for Spanish relative to English (β=-.31, p=.025) and the reverse for German (β=.38, p=.004).

With respect to our main hypothesis, we found an interaction between language and noun group for Spanish relative to English (β=-.48, $p <$=.001), indicating a substantial effect of a noun's grammatical gender on the choice of an adjective semantically associated with that gender. The effect for German was present numerically (Figure 2), but did not reach significance, a result that is in line with weaker attraction effects in the distributional semantic space observed in Part 1 (Figure 1C).

Effect of cosine similarity. Finally, we tested the relative contributions of (a) the noun's grammatical gender and (b) the cosine similarities between a noun and masculine/feminine adjectives as predictors of the participants' adjective selection in each trial. (Embeddings were extracted from the *fastText* model.) A model that included both predictors outperformed both the model without the cosine similarities (χ^2=226, $p <$.001) and the model without the noun group information (χ^2=9.0, p=.029). The full model (Table 3) shows significant predictive power of cosine similarity to both masculine and feminine adjectives for all 3 languages, a significant effect of question number (whereby participants choose the masculine adjective more often in later trials) and, interestingly, a significant effect of noun group in Spanish even when the cosine similarities are accounted for, suggesting that the cosine similarities do not fully mediate the effect of a noun's grammatical gender on adjective choice.

Discussion

Grammatical gender is an intriguing linguistic feature in that it is frequently yet inconsistently related to the semantics of the nouns to which it is assigned. As such, grammatical gender and semantic gender associations are expected to interact in the distributional semantic space. Here, we empirically tested the presence of such an interaction in two gendered languages, Spanish and German, and additionally investigated whether these distributional semantic effects might predict human behavior in a noun-adjective matching task.

An analysis of the distributional semantic space showed an interaction between language, adjective group, and noun group for both Spanish and German, such that grammatically masculine nouns in each language were more similar to men-associated adjectives (relative to women-associated adjectives) than grammatically feminine nouns. These results are in accordance with the hypothesis that grammatical gender and semantic gender associations are linked in the distributional semantic space. The matching task showed effects of cosine similarity between nouns and adjectives on matching choices; for Spanish but not for German, we also found effects of grammatical gender that explained additional variance relative cosine similarities. The grammatical gender effect was not significant in German, a fact perhaps related to the small effect of grammatical gender on similarity scores between nouns and adjectives in Part 1. We conclude that the distributional account can provide valuable insights into the relationship between grammatical and semantic gender, although perhaps it does not fully capture grammatical gender effects that affect human behavior.

Different results (both distributional and behavioral) observed for Spanish and German open up questions about the factors that might affect the entanglement of grammatical and semantic gender across languages. Distinctions have been made in the literature between the congruence of semantic gender and grammatical gender specifically between Spanish and German: (Sera et al., 2002) noted that there seems to be a correlation between perceptual cues of semantic gender and grammatical gender, whereas (Comrie, 1999) considers the German grammatical gender system to be more arbitrary; (Landor, 2014, p. 46-47) provides some examples of mismatches between semantic gender and grammatical gender in German, which might weaken the association between these features in the distributional semantics space and thus lead to weaker linguistic relativity effects. Investigating the relationship between grammatical and semantic gender across a more diverse set of languages, both typologically and behaviorally, could help build a systematic framework for quantifying the links between language structure and its effects on cognition.

Examining linguistic relativity effects from a distributional semantics perspective opens up a variety of exciting avenues for future work. In our analyses, we restricted our adjective set to adjectives with high cosine similarity scores to ”man” and ”woman”, a feature that was predictive of semantic association of these adjectives with men and women, but much variance was left unexplained. A strong test of the distributional semantics account would entail checking whether adjectives controlling for semantic association scores but systematically varying cosine similarity (and vice versa) would modulate grammatical gender effects.

In addition, this paper examined intralinguistic effects of grammatical gender on semantic gender associations, yet Whorfian accounts are often in search of cross-linguistic or even extra-linguistic effects. Thus, it might then be of interest to expand our hypothesis to one more resembling that of Boroditsky et al. (2003). However, research on cross-linguistic effects of this nature is complicated by a variety of confounds. (Bassetti & Nicoladis, 2016) found that the knowledge of other languages (and, further, the particular combination of languages known) might reduce the effect of grammatical gender on thought. One way to test these relationships is by leveraging contemporary large language models, in particular multilingual and multimodal models.

Finally, although word embedding similarity was a significant predictor of noun-adjective matching choices, effects of grammatical gender were not fully mediated by it. One explanation is that *fastText* embeddings are an imperfect model of a distributional semantic space, and a better model would capture more effects pertaining to the grammatical-semantic gender interaction. Yet an alternative account is that grammatical gender exerts influence on human behavior not only by warping the distributional semantic space, but also by some other mechanism. Testing mechanistic accounts of grammatical gender effects on semantic cognition—and of linguistic relativity effects more broadly—is a critical step toward clarifying the relationship between language and thought.

Future directions. There are several possibilities for why results are observed in all experiments for Spanish but only some for German. One of which is that the proposed warping effect is simply weaker or more confounded in German than in Spanish–differences between languages or language families could then explain this effect. Another is that our noun and adjective materials have too high a degree of infrequency, regionality, or other idiosyncrasies that would confound the observation of an effect. The selection of only those nouns which have opposite grammatical genders in Spanish and German could also confound an effect. It would thus be beneficial to test if our results generalize to another set of materials and to investigate other languages to explore whether there are patterns in where this effect shows up.

Acknowledgements

We would like to thank Eric Martinez and Carina Kauf for their assistance in verifying the translations of instructions and materials for Spanish and German, respectively. We would also like to thank Evelina Fedorenko and Ted Gibson for providing the infrastructure and material support for conducting the behavioral studies.

References

Bassetti, B., & Nicoladis, E. (2016). Research on grammatical gender and thought in early and emergent bilinguals. *International Journal of Bilingualism*, *20*(1), 3–16. doi: 10.1177/1367006915576824

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bhatia, S., Richie, R., & Zou, W. (2019). Distributed semantic representations for modeling human judgment. *Current Opinion in Behavioral Sciences*, *29*, 31–36.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. *Language in mind: Advances in the study of language and thought*, *22*, 61–79.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Comrie, B. (1999). Grammatical gender systems: A linguist's assessment. *Journal of Psycholinguistic Research*, *28*, 457–466. doi: 10.1023/A:1023212225540

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *Bert: Pre-training of deep bidirectional transformers for language understanding.*

Elpers, N., Jensen, G., & Holmes, K. J. (2022). Does grammatical gender affect object concepts? registered replication of phillips and boroditsky (2003). *Journal of Memory and Language*, *127*, 104357.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(13).

Landor, R. (2014). Grammatical categories and cognition across five languages: The case of grammatical gender and its potential effects on the conceptualisation of objects. *Griffith University, Brisbane, Australia*, 46-47.

Lenci, A., et al. (2008). Distributional semantics in linguistic and cognitive research. *Italian journal of linguistics*, *20*(1), 1–31.

Mickan, A., Schiefke, M., & Stefanowitsch, A. (2014). Key is a llave is a schlüssel: A failure to replicate an experiment from boroditsky et al. 2003. *Yearbook of the German Cognitive Linguistics Association*, *2*, 39–50.

Sera, M., Elieff, C., Forbes, J., Burch, M., Rodríguez, W., & Dubois, D. (2002, Sep). When language affects cognition and when it does not: an analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, *131*(3), 377-397.

Stańczak, K., Du, K., Williams, A., Augenstein, I., & Cotterell, R. (2023). Grammatical gender's influence on distributional semantics: A causal perspective. *arXiv preprint arXiv:2311.18567*.

Williams, A., Cotterell, R., Wolf-Sonkin, L., Blasi, D., & Wallach, H. (2021). On the relationships between the grammatical genders of inanimate nouns and their co-occurring adjectives and verbs. *Transactions of the Association for Computational Linguistics*, *9*, 139–159.