GRAMMATICAL GENDER IS REFLECTED IN LANGUAGE'S DISTRIBUTED SEMANTIC REPRESENTATIONS

Alfred Lim (University of Nottingham Malaysia) & Luca Onnis (University of Oslo)
alfred.lim@nottingham.edu.my

It has been proposed that sex-based grammatical gender on inanimate nouns “rubs off” on concepts, such that, for instance, Italian speakers conceptualize ‘spoon’ as more masculine than Spanish or French speakers do. Here, we provide an explicitly computational account of this interaction of grammatical gender and conceptual representations. We examined semantic embeddings generated from distributional semantics models, wherein the semantics of words is represented by vectors of real number values extracting their contextual co-occurrence with other words. As such, word embeddings sharing similar contexts have more similar semantic vectors (e.g., 'flight' and 'plane' versus ‘apple’ and ‘orange’). Embeddings are compatible with a dynamical systems approach to language and mimic the semantic structure found in the cognitive architecture of humans. We propose that gender biases on inanimate nouns may originate by associations from both their surrounding context like gender-marked determiners (in Italian, ‘il/gli’ and ‘la/le’ indicate respectively masculine and feminine animate and inanimate nouns) or adjectival agreement morphemes, and word-internal similarities (e.g., both animate and inanimate nouns ending in -o and -a refer respectively to male and female nouns).

We analyzed pre-trained FastText embeddings of inanimate nouns in 16 languages with grammatical gender (e.g., Arabic, Hindi, Italian, Spanish), and 1 non-gendered language as control (English). For each language combination (total n = 136), we selected inanimate nouns in Wiktionary (min set = 253, max set = 2821) that were of opposite grammatical gender, and of course neutral in English, resulting in word triplets [e.g., "spoon", "cucchiaio" (masculine) in Italian, "cuchara" (feminine) in Spanish]. This controlled for words’ intrinsic semantics, providing a comprehensive and fair cross-language comparison.

The Semantic Proximity Difference (SPD) is a metric that captures the disparity between the word embedding's similarity to 8 reference words of intrinsic female semantics (‘woman’, ‘girl’, ‘sister’, etc.) and its similarity to 8 male reference words (‘man’, ‘boy’, ‘brother’, etc.). For example, SPD can be computed as: *SPD = cosine('cucchiaio', 'femmina') – cosine('cucchiaio', 'maschio')*. A positive value indicates a stronger similarity to female reference words, while a negative value suggests a stronger similarity to male reference words. We utilized linear mixed models with SPD scores as dependent variable, grammatical gender as a three-level factor (neutral, feminine, masculine), and word triplets as a random effect nested in language pairings. Across all gendered languages, FastText word embeddings were closer to the set of reference words aligning with their noun's grammatical gender (χ^2 = 101610.49, p < .001, η^2 = 0.25). Specifically, the effect size of the female-neutral gender contrast was 0.822 (95% CI [0.815, 0.829]), and the male-neutral gender contrast effect size was -0.296 (95% CI [-0.303, -0.288]), both in the expected direction.

To compare the gender effect size between Word2Vec and FastText models, we conducted an analysis on the intersection of the two sets of embeddings, encompassing 10 languages and a total of 17,408 inanimate nouns. The results show that Word2Vec models, which didn't incorporate subword information during training, had a smaller gender effect size (η^2 = 0.20) compared to the FastText models (η^2 = 0.45) that did include subword information in their training. This is in line with our hypothesis that grammatical gender may shape lexical semantics by leveraging both sub-lexical and contextual distributional similarities among words. These masculine/feminine semantic features are showing up in language statistics that do not *directly* incorporate sensorimotor experience, suggesting that much of the embodiment in language processing might become statistically embedded in how words are used.