Is there a semantic core to noun agreement classes?
Data from 15 languages spanning 6 families

Sometimes, language carves nature at its joints. Not only do words tend to pick out coherent categories (dogs, cats, mats, lamps) rather than franken-categories (things to take out of the house during a fire), syntax itself often respects categorical distinctions (Hartshorne et al., 2014; Levin, 1993; Talmy, 2000).

Other aspects of language appear, at least at first glance, to be at least mostly arbitrary. These include the lexical gender classes of Indo-European and Semitic languages, classifier systems in East Asian languages, the Bantu noun class system, and animacy classes in many American languages, among others. While these all have some semantic correlates — for instance, separating male and female entities (in gender systems), animates and inanimates (in animate systems), and humans, plants, and animals (some Bantu systems) — in many if not most cases, there is no obvious semantic motivation for a noun's class assignment. The arbitrariness of lexical gender in German was famously lampooned by Mark Twain (1880) in an essay titled, "The awful German language" (among his complaints: the German word for *turnip* is female, whereas the word for *maiden* is neuter).

Twain may have overstated the case. Using word embeddings as a proxy for semantics, Williams et al. (2019) found that words with similar meanings typically have the same lexical gender in 15 of 18 languages studied, even for inanimate nouns (which lack *actual* gender).

Here, we ask whether, to the extent noun agreement classes carve up semantics, do they do so in compatible ways? Do nouns that appear in the same noun class in one language tend to appear in the same noun class in other languages?

We obtained the noun class for 633 animate and 247 inanimate concepts (the number of words varied across languages; all languages lacked words closely matching some concepts) in Indo-European (Spanish, French, Italian, Sicilian, German, Dutch, Ukrainian, Lithuanian), Afro-Asiatic (Arabic, Hebrew), Niger-Congo (Swahili, Kîîtharaka), Algic (Ojibwe), Sino-Tibetan (Mandarin), and Koreanic (Korean) languages. For each language, we calculated pairwise similarity between every noun, based on occurring in the same class(es) or not. Similarity in these similarities was calculated across languages. Missing data, nouns with multiple classes, close synonyms, and other nuances were handled in sensible ways.

For animate nouns, noun classification was consistent within Indo-European languages (Fig). This does not only reflect shared heritage: while there are two large perfect cliques (groups of nouns that are in the same class in all languages) – one consisting of mostly inherently male entities and the other mostly inherently female entities – there were several other cliques: groups of words that may be masculine in some languages and feminine or neuter in others, but always in the same class. This pattern is reduced but still present when all 6 language families are considered.

Inanimate nouns showed much less consistency across languages (though there is some!), even considering only Indo-European (Fig).

We discuss potential explanations for these patterns, including chance agreement. We anticipate adding additional languages and families before September, potentially sharpening results. Implications for acquisition are discussed.
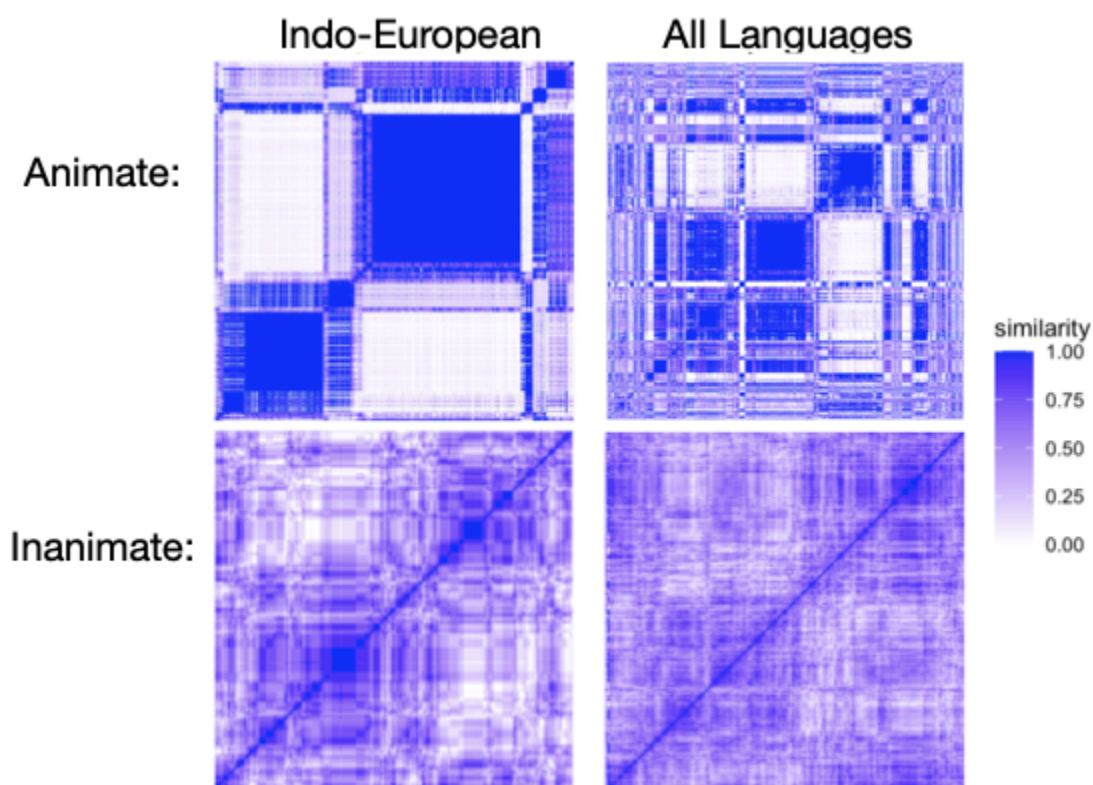
Separate heatmaps for animate nouns (top) and inanimate nouns (bottom), for all languages (right) or Indo-European only (left). Each heatmap depicts pairwise similarity between words. A similarity of 0 (white) means the noun pairs are never in the same noun class in any of the 15 languages. A similarity of 1 (dark blue) means the noun pair are always in the same noun class(es) in every language.

Hartshorne, J. K., Bonial, C., & Palmer, M. (2014). The VerbCorner Project: Findings from Phase 1 of crowd-sourcing a semantic decomposition of verbs. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 397-402).

Levin, B. (1993). English verb classes and alternations: A preliminary investigation. University of Chicago press.

Talmy, L. (2000). Toward a cognitive semantics: Concept structuring systems (Vol. 1). MIT press.

Twain, M. (1880). *A tramp abroad*.

Williams, A., Blasi, D., Wolf-Sonkin, L., Wallach, H., & Cotterell, R. (2019). Quantifying the Semantic Core of Gender Systems. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 5734-5739).