

The Diverse Names Database: A Tool for Creating More Equitable, Diverse, and Inclusive Linguistic Example Sentences

Lex Konnelly, Pocholo Umbal, and Nathan Sanders
University of Toronto

Linguists generally agree that all language varieties are valid, not just as objects of study within linguistics, but as ways of using language. While there is significant linguistic scholarship dedicated to advocating for more affirming language representations, this is unfortunately not always reflected in how linguistics is taught and practiced. In this talk, we describe a project intended to respond to one such salient concern in the linguistics classroom: stereotyping and language biases in example sentences.

In addition to representing a main source of data in linguistic research within various sub-fields, example sentences are a core vehicle for linguists in teaching a wide range of phenomena to our students. However, in both cases, it has been observed that the content of these sentences largely reflect the biases of the researchers who construct them: referents are typically given generic, Anglocentric proper names like *John* and *Mary*, reflecting dominant white culture and conforming to heteronormative gender roles. Prior research has shown that, particularly in syntax textbooks and journals, example sentences systematically under-represent women and perpetuate harmful gender stereotypes, such as presenting feminine-gendered arguments as non-subjects, as more likely to be unnamed, and more often referred to in kinship terms in relation to masculine-gendered referents (Macaulay and Brice 1997, Pabst et al. 2018, Richy and Burnett 2019, Kotek et al. 2020). The field of linguistics is becoming increasingly attuned to these and other language-related biases, and considering the role of representation in example sentences is embedded within a larger disciplinary project of interrogating how linguistics can be more affirming of human diversity (see, for example, recent calls to action by Rickford and King 2016, Leonard 2018, Conrod 2019, Charity Hudley 2020, and Sanders et al. 2020).

In the classroom, linguists may not realize that they are relying on their own biases in coming up with examples, particularly in ‘on-the-fly’ situations such as in lectures and tutorials. To this end, we aim to share our progress on the construction of the Diverse Names Database, a database of names for every letter of the English alphabet, from a variety of languages and cultures, categorized by gender and confirmed with native speakers (see Fig. 1 for excerpt).

Our hope is that this tool will not only be supportive to instructors in constructing more diverse and inclusive examples in their course material, but that it will also be a helpful resource for linguistics students, enabling them to create more affirming examples in their assignments and other coursework. In this presentation, we outline the goals for the project in relation to the above concerns, introduce our process of developing and adjusting the design, and present some additional issues for consideration and ideas for using the database. Once publicly accessible, the Diverse Names Database represents an additional resource for increasing gender and cultural diversity and representation in our linguistics example sentences and for providing a more equitable experience for linguistics students, and ultimately, for the future of the field.

Figure 1: Sample entries from the Diverse Names Database

feminine-leaning	masculine-leaning	all-gender
Anuheā (Hawaiian, Austronesian)	Aimo (Finnish, Uralic)	Ameyalli (Nahuatl, Uto-Aztecan)
⋮	⋮	⋮
Zuriñe (Basque, isolate)	Zaharia (Romanian, Indo-European)	Zhyrgal (Kyrgyz, Turkic)

References

- Charity Hudley, Anne H. 2020. Fostering a culture of racial inclusion in linguistics: For the children of the 9th Ward circa 2005. Plenary address at the 94th annual meeting of the Linguistic Society of America.
- Conrod, Kirby. 2019. Doing gender and linguistics. Keynote talk given at They, Hirsself, Em, and You: Nonbinary Pronouns in Research and Practice.
- Kotek, Hadas, Sarah Babinski, Rikker Dockum, and Christopher Geissler. 2020. Gender representation in linguistic example sentences. *Proceedings of the Linguistic Society of America* 5(1): 514–528. <https://doi.org/10.3765/plsa.v5i1.4723>
- Leonard, Wesley Y. 2018. Reflections on (de)colonialism in language documentation. In *Reflections on language documentation 20 years after Himmelmann 1998*, ed. Bradley McDonnell, Andrea L. Berez-Kroeker, and Gary Holton. Honolulu: University of Hawai‘i Press, 55–65.
- Macaulay, Monica and Colleen Brice. 1997. Don’t touch my projectile: Gender bias and stereotyping in syntactic examples. *Language* 73(4): 798–825.
- Pabst, Katharina, Paola Cépeda, Hadas Kotek, Krysten Syrett, Katharine Donelson, and Miranda McCarvel. 2018. Gender bias in linguistics textbooks: Has anything changed since Macaulay & Brice (1997)? Talk presented at the 92nd annual meeting of the Linguistic Society of America.
- Richy, Célia and Heather Burnett. 2019. Jean does the dishes while Marie fixes the car: A qualitative and quantitative study of social gender in French syntax articles. *Journal of French Language Studies* 30(1): 1–26.
- Rickford, John R. and Sharese King. 2016. Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language* 92(4): 948–988.
- Sanders, Nathan, Pocholo Umbal, and Lex Konnelly. 2020. Methods for increasing equity, diversity, and inclusion in linguistics pedagogy. *Actes du Congrès de l’ACL 2020 | 2020 CLA Conference Proceedings*. https://cla-acl.artsci.utoronto.ca/wp-content/uploads/actes-2020/Sanders_Umbal_Konnelly_CLA-ACL2020.pdf