

# Ling 795 Seminar: Corpus-based typology

San Diego State University

Schedule # 22337

Fall 2022

We 4:00pm–6:40pm

PSFA-437

Linguistic typology is the study of cross-linguistic variation. Modern typology has evolved past a search for linguistic ‘types’ to broader questions about linguistic structures: as Bickel (2013) puts it, “What’s where why?” This has led to interest in exploring new sources of data, including large multi-lingual text databases. This allows us to consider the typological consequences of both inter- and intra-linguistic variation. We will begin the semester by reading Greenberg’s (1966) seminary paper on word order typology followed by some of the corpus-based developments from that work. Then, based on issues that come up during class discussion, we will dig into the primary literature on other key topics corpus-based typology.

## Instructor

Prof. Rob Malouf

Website: [malouf.sdsu.edu](http://malouf.sdsu.edu)

Email: [rmalouf@sdsu.edu](mailto:rmalouf@sdsu.edu)

Office hours: Tu Th 1:00–2:00 or by appointment

Real office: SHW 244

Zoom office: [SDSU.zoom.us/j/88663149131](https://SDSU.zoom.us/j/88663149131)

Class rosters are provided to the instructor with the student’s legal name. Please let me know if you would prefer an alternate name and/or gender pronoun.

## Student Learning Outcomes

The goals of this course are for participants to gain experience in:

- understanding current theoretical approaches to typology using multi-lingual databases,
- reading and evaluating the primary literature,
- presenting and discussing research material with peers,

- identifying open research questions,
- applying descriptive and theoretical concepts to novel linguistic data.

Through the term, participants (including auditors!) will present and discuss articles from the reading list, which cover a number of aspects of constructional change.

## Course Materials

There is no textbook for this course. Readings for each will be available through the library or posted on the class Canvas site. The specific readings will depend on the interests of the participants.

## Course Design

Each week, we will discuss one or more papers from the research literature on corpus-based typology. We will begin by looking at word order, a central area of typological interest since Greenberg (1961):

- Schnell, Stefan and Nils Norman Schibor (2022). Crosslinguistic corpus studies in linguistic typology. *Annual Review of Linguistics* 8: 171–191.
- Greenberg, Joseph H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. In J.H. Greenberg (ed.), *Universals of language (2nd edition)*, MIT Press. Pp. 72–113.
- Dryer, Matthew S. (1992). The Greenbergian word order correlations. *Language* 68: 81–138.
- Wälchli, Bernhard (2009). Data reduction typology and the bimodal distribution bias. *Linguistic Typology* 13: 77–94.
- Lui, Haitau (2010). Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua* 120: 1567–1578.
- Gerdes, Kim, Sylvain Kahane and Xinying Chen (2021). Typometrics: From implicational to quantitative universals in word order typology. *Glossa: a journal of general linguistics* 6(1): 17. 1–31.
- Hee-Soo Choi, Bruno Guillaume, Karën Fort (2021). Corpus-based language universals analysis using universal dependencies. *SyntaxFest Quasy 2021*

- *Quantitative Syntax*, Mar 2022, Sofia, Bulgaria.

After this, we will choose additional topics to explore based on the interests of the group. We'll all take turns leading the discussions, but everyone should be sure to read the papers in advance. To help focus on what's most relevant, I will give you discussion questions to keep in mind while you are reading.

In addition to leading and participating in discussions, students taking the class for a grade will also prepare a final research project. The final grade will be based on class participation, discussion questions, and on the final project:

Project draft	Dec 10	20%
Final project	Dec 17	50%
Class presentations	—	30%

## Final reading list

### • Introduction

#### – Week 1 (8/24)

- \* Schnell, Stefan and Nils Norman Schiborr (2022). Crosslinguistic corpus studies in linguistic typology. *Annual Review of Linguistics* 8: 171–191.

### • Word order typology

#### – Week 2 (8/31)

- \* Greenberg, Joseph H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. In J.H. Greenberg (ed.), *Universals of language (2nd edition)*, MIT Press. Pp. 72–113.
- \* Dryer, Matthew S. (1992). The Greenbergian word order correlations. *Language* 68: 81–138.

#### – Week 3 (9/7)

- \* Wälchli, Bernhard (2009). Data reduction typology and the bimodal distribution bias. *Linguistic Typology* 13: 77–94.
- \* Lui, Haitau (2010). Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua* 120: 1567–1578.

#### – Week 4 (9/14)

- \* Gerdes, Kim, Sylvain Kahane and Xinying Chen (2021). Typometrics: From implicational to quantitative universals in word order typology. *Glossa: a journal of general linguistics* 6(1): 17. 1–31.
- \* Hee-Soo Choi, Bruno Guillaume, Karën Fort (2021). Corpus-based language universals analysis using universal dependencies. *SyntaxFest Quasy 2021 - Quantitative Syntax*, Mar 2022, Sofia, Bulgaria.

– Week 5 (9/21)

- \* Östling, Robert (2015). Word order typology through multilingual word alignment. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 205–211.
- \* Guzmán Naranjo, Matías and Laura Becker (2015). Quantitative word order typology with UD. 17th International Workshop on Treebanks and Linguistic Theory, Oslo.
- \* Hahn, Michael, Dan Jurafsky, and Richard Futrell (2020). Universals of word order reflect optimization of grammars for efficient communication. *PNAS* 117 (5) 2347–2353.

• **Speech rate**

– Week 6 (9/28)

- \* Roach, Peter (1998). Some languages are spoken more quickly than others. In Bauer & Trudgill (eds) *Language Myths*. Penguin Books. Pp. 150–158.
- \* Pellegrino, François, Christophe Coupé, and Egidio Marsico (2011). Across-language perspective on speech information rate. *Language* 87(3) 539–558.
- \* Coupé, Christopher, Yoon Oh, Dan Dediu, and François Pellegrino (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances* 5(9): eaaw2594.

• **Phonology**

– Week 7 (10/5)

- \* Salesky, Elizabeth, Eleanor Chodroff, Tiago Pimentel, Matthew Wiesner, Ryan Cotterel, Alan W Black, and Jason Eisner (2020). A corpus for large-scale phonetic typology. *ACL 2020*, pp. 4526–4546.
- \* Ahn, Emily P. and Eleanor Chodroff (2022). VoxCommunis: A Corpus for Cross-linguistic Phonetic Analysis. *LREC 2022*, pp. 5286–5294.

- \* Hutin, Mathilde and Marc Allassonnière-Tang (2022). Operation LiLi: Using Crowd-Sourced Data and Automatic Alignment to Investigate the Phonetics and Phonology of Less-Resourced Languages. *Languages* 7:234.
- \* Seifart, Frank (2022). Combining documentary linguistics and corpus phonetics to advance corpus-based typology. In Haig, Schnell, and Seifart (eds.), *Doing corpus based typology with spoken language data: State of the art*, pp. 115–139.

## • Lexical semantics

### – Week 8 (10/12)

- \* Wälchli, Bernhard & Michael Cysouw (2012). Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3): 671–710.
- \* Levshina, Natalia (2015). European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles. *Folia Linguistica* 49(2): 487–520.
- \* Beekhuizen, Barend, Julia Watson & Suzanne Stevenson (2017). Semantic typology and parallel corpora: Something about indefinite pronouns. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)*, 112–117.

### – Week 9 (10/19)

- \* François, Alexandre (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Martine Vanhove (ed.), *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations*, Benjamins, 163–215.
- \* Paradis, Carita (2016). Corpus methods for the investigation of antonyms across languages. In Päivi Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts*, 131–156. Berlin: De Gruyter Mouton.
- \* Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. *Cognition* 201:104280.

### – Week 10 (10/26)

- \* Östling, Robert (2016). Studying colexification through massively parallel corpora. In Päivi Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts*, 157–176. Berlin: De Gruyter Mouton.
- \* Regier T, Carstensen A, Kemp C (2016) Languages Support Efficient Communication about the Environment: Words for Snow Revisited. *PLoS ONE* 11(4): e0151138.
- \* Bradford, Laurestine, Thomas, Guillaume, and Xu, Yang (2022). Communicative need modulates lexical precision across semantic domains: A domain-level ac-

count of efficient communication. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44.

- **Discourse**

- Week 11 (11/2)

- \* DuBois, John W (1987). The discourse basis of ergativity. *Language* 63: 805–855
    - \* Haig, Geoffrey and Schnell, Stefan. 2016. The discourse basis of ergativity revisited. *Language* 92: 591–618

- Week 12 (11/9)

- \* Haig, Geoffrey (2018). The grammaticalization of object pronouns: Why differential object indexing is an attractor state. *Linguistics* 56(4): 781–818. travis
    - \* Haig, Geoffrey, Schnell, Stefan & Schiborr, Nils N. (2021). Universals of reference in discourse and grammar: Evidence from the Multi-CAST collection of spoken corpora. In Haig, Schnell, & Seifart (eds.), *Doing corpus-based typology with spoken language corpora*. Honolulu: University of Hawai'i Press. 141–177. bryan

- Week 13 (11/16)

- \* San Roque, Lila, Lauren Gawne, Darja Hoenigman, Julia C. Miller, Alan Rumsey, Stef Spronck, Alice Carroll & Nicholas Evans. 2012. Getting the story straight: Language fieldwork using a narrative problem-solving task. *Language documentation and conservation* 6:135-174.
    - \* Pratiwi, Desak Putu Eka, I Wayan Arka, & Asako Shiohara (2018). On the situated socio-cultural meaning of benefactives in Balinese. *Linguistik Indonesia* 117–128.

- Week 14 (11/30)

- \* Barth D, Evans N. 2017. SCOPIC design and overview. In *The Social Cognition Parallax Interview Corpus (SCOPIC): A Cross-Linguistic Resource*, ed. D Barth, N Evans, pp. 1–23. Honolulu: Univ. Hawai'i Press
    - \* Barth, Danielle, Nicholas Evans, Arka I Wayan, Henrik Bergvisk, et al. (2021). Language vs. individuals in cross-linguistic corpus typology. In Geoffrey Haig, Stefan Schnell & Frank Seifart (eds.), *Doing corpus based typology with spoken language data: State of the art*, 179–232. Honolulu, HI: University of Hawai'i Press.