Patterns of Closeness and Abstractness in Colexifications
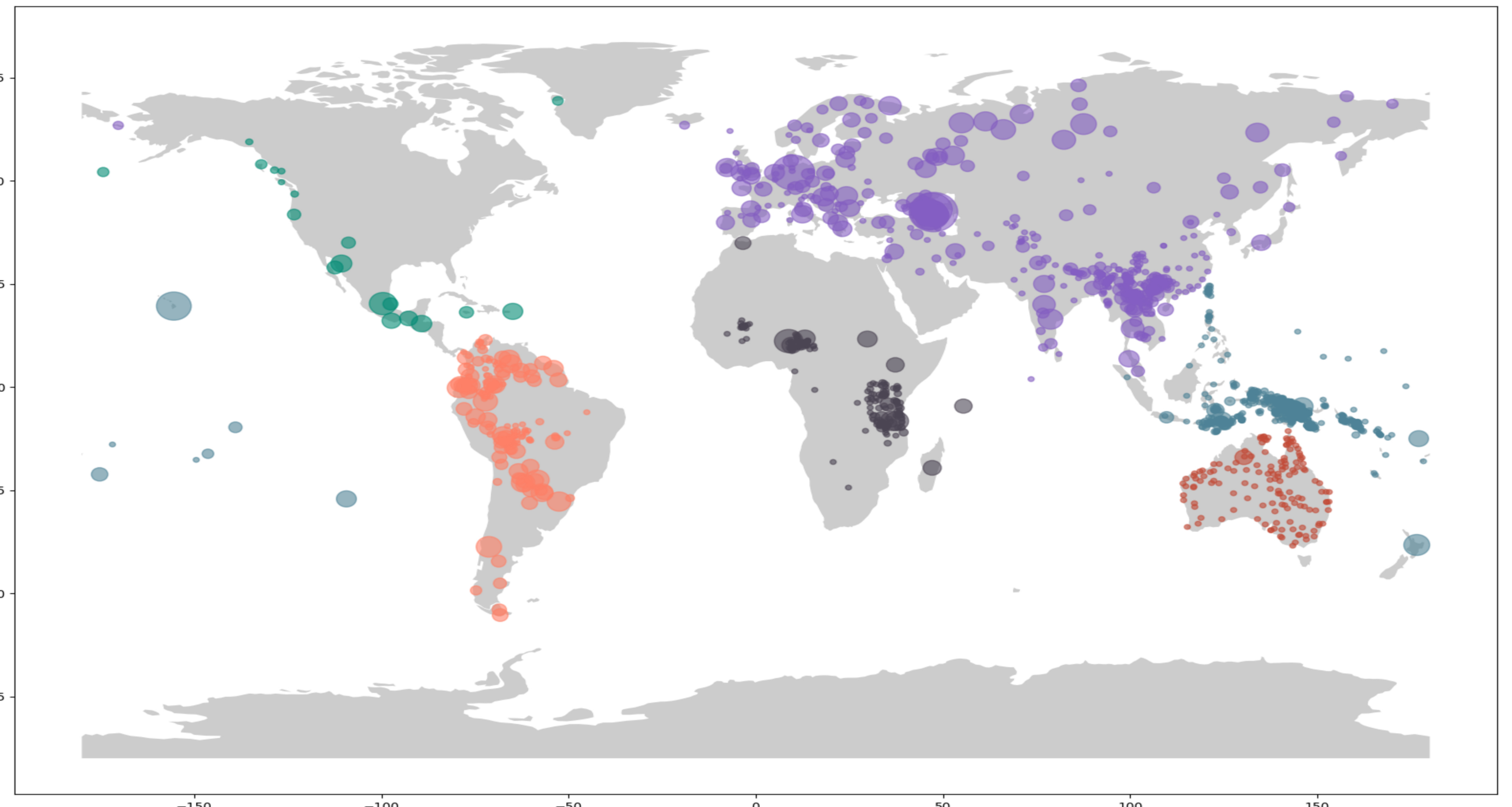## The Case of Indigenous Languages in Americas

Yiyi Chen & Johannes Bjerva
Department of Computer Science, Aalborg University, Copenhagen, Denmark

1. Background

- Colexification refers to the linguistic phenomenon where a single lexical form is used to convey multiple meanings.
- A colexification pattern refers to a case where two concepts are colexified, such as MOON-SUN.
- A colexification is an instance of a colexification pattern, for example, yazune'i in South American language Aikanã.
- Concrete concepts are those that are perceived by the senses, such as CAT and MOUNTAIN.
- Abstract concepts are not perceived by senses, such as RELATIONSHIP and UNDERSTANDING.
- Lakoff and Johnson (2008) shows that abstract concepts are often understood by reference to more concrete concepts.
- Xu et al. (2020) shows that concepts more distant in concreteness are more likely to colexify using Intercontinental Dictionary Series.
- However, studies show that abstract words are often combined with abstract ones, while concrete words are often combined with concrete ones (Frassinelli et al., 2017; Naumann et al., 2018).
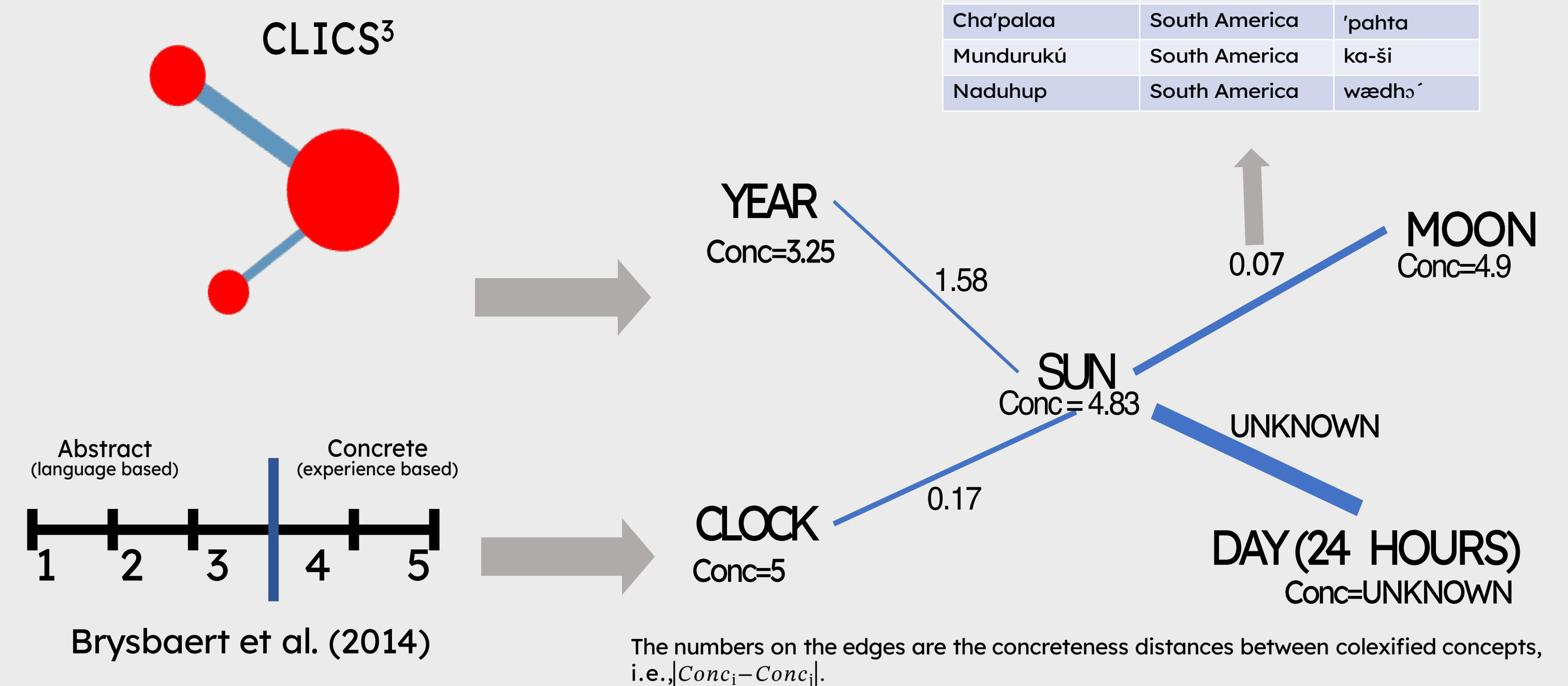
2. Hypothesis

Concepts *closer* in concreteness/abstractness are more likely to colexify.

In other words, concrete concepts are more likely to colexify with concrete concepts, while abstract concepts are more likely to colexify with abstract concepts.
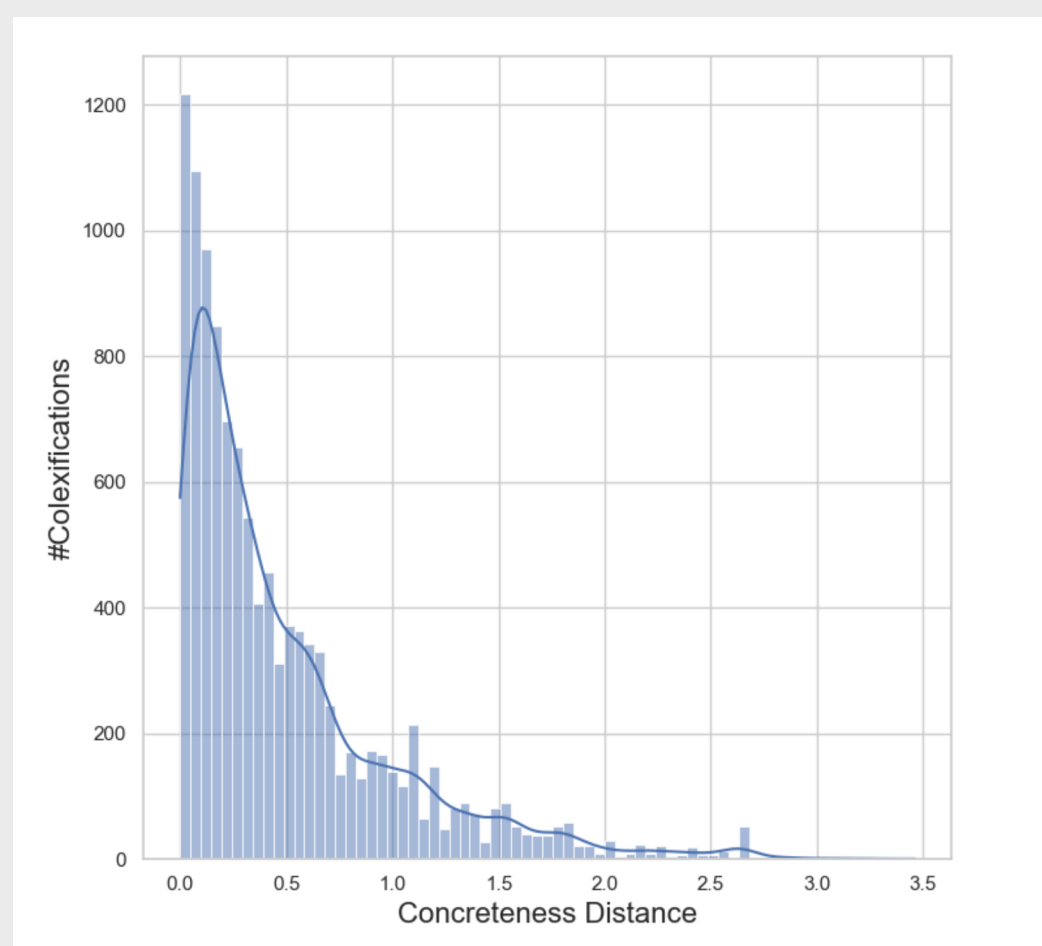
4. Analyses

Across 170 Indigenous Languages | Across 1989 Languages
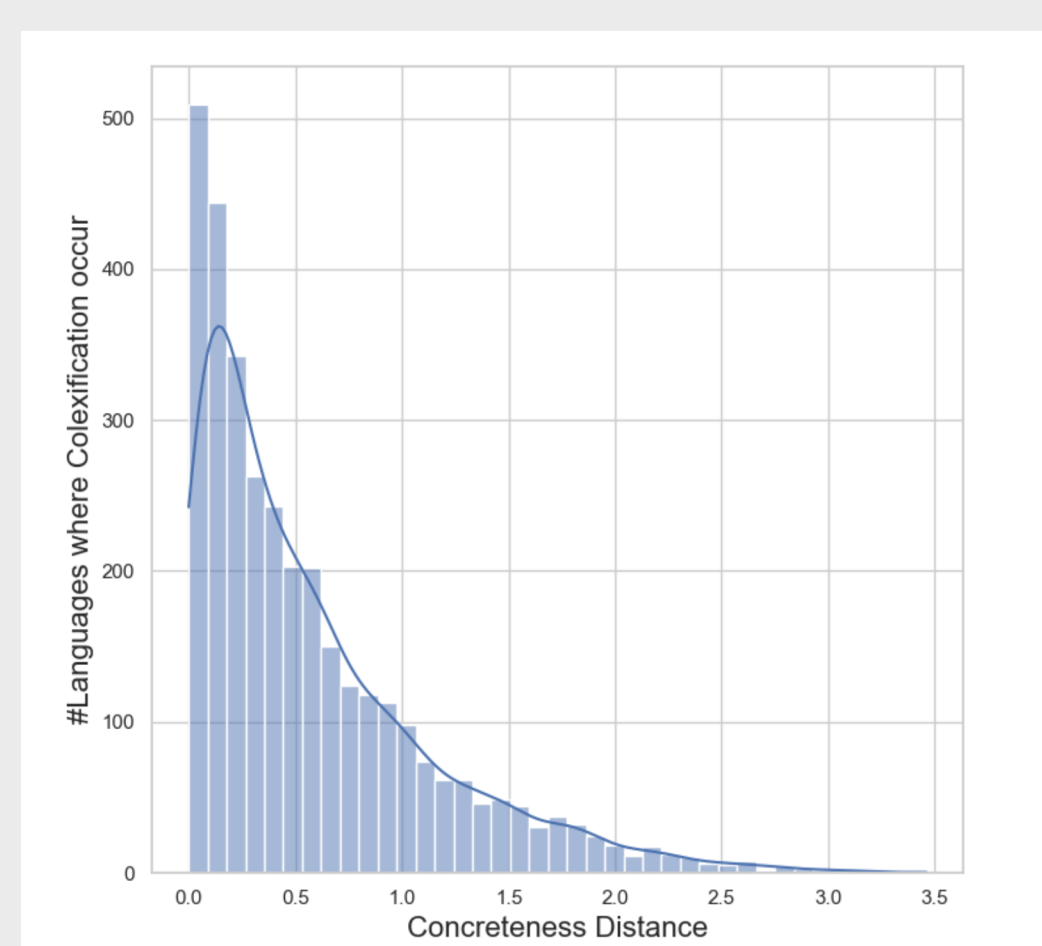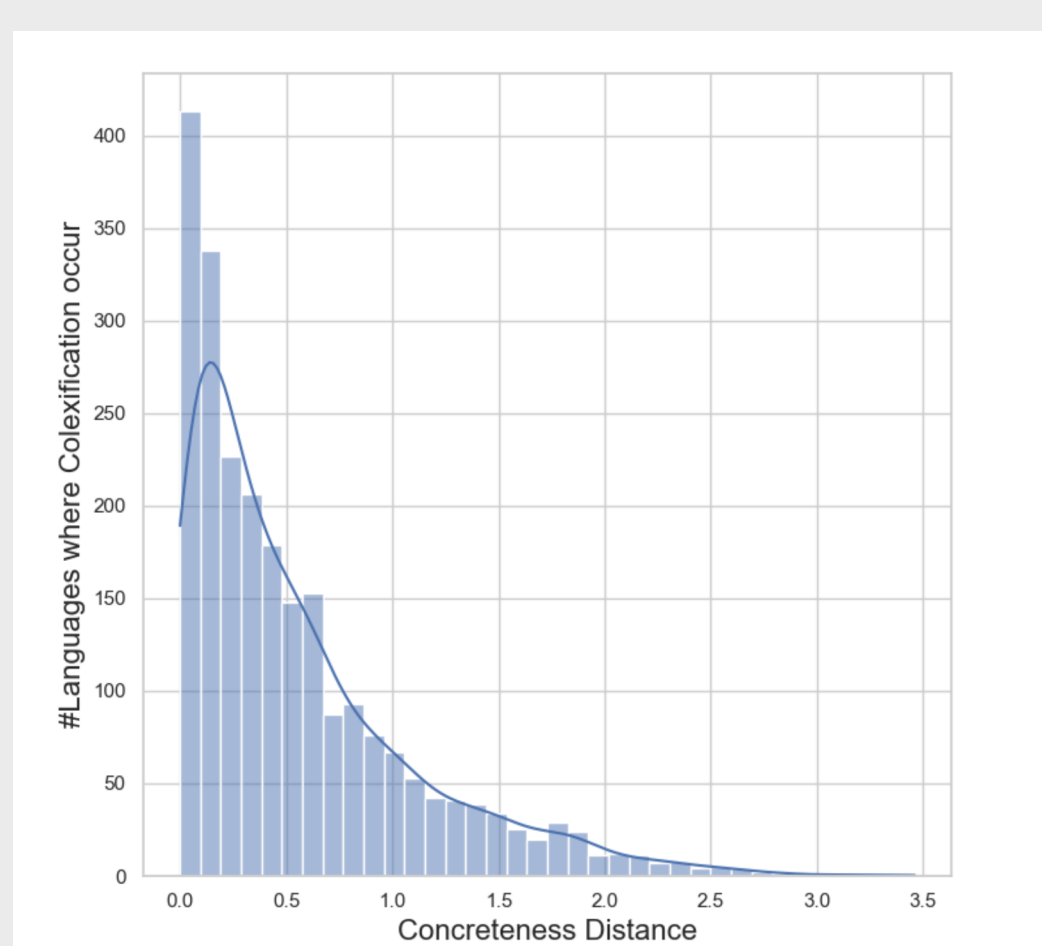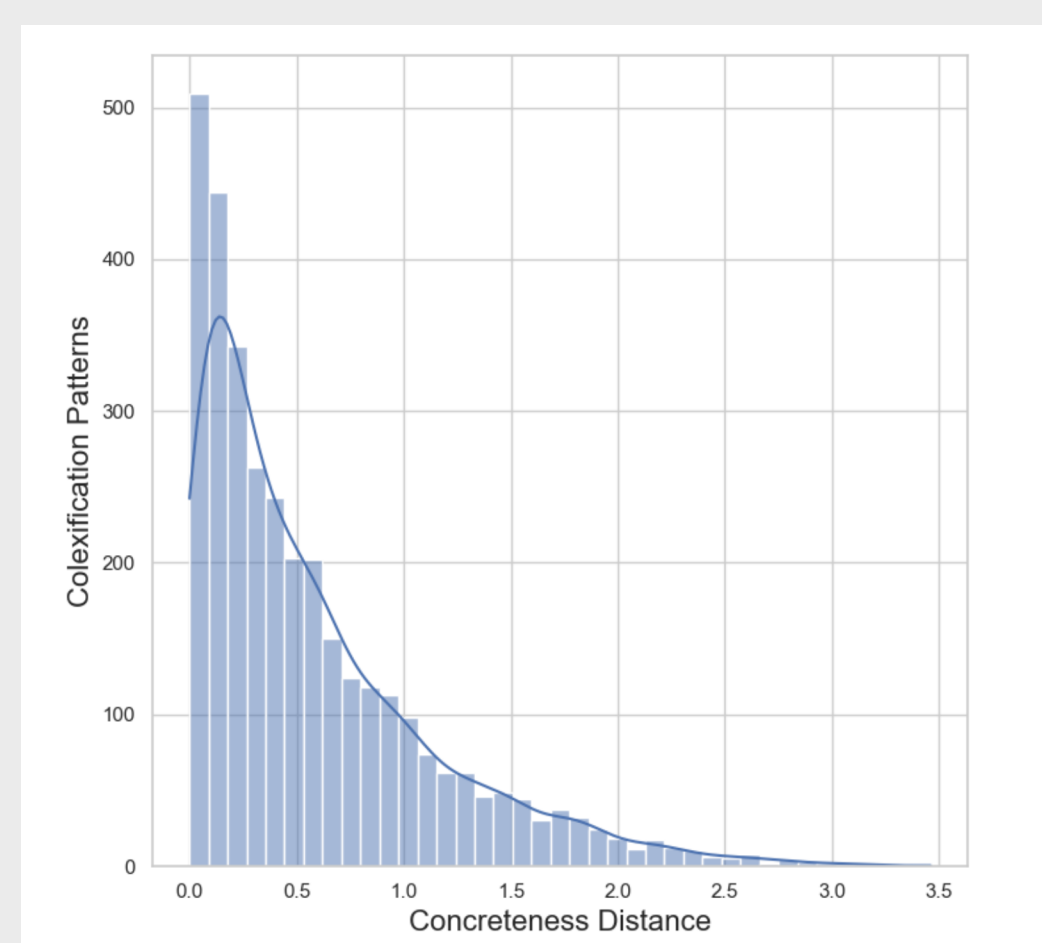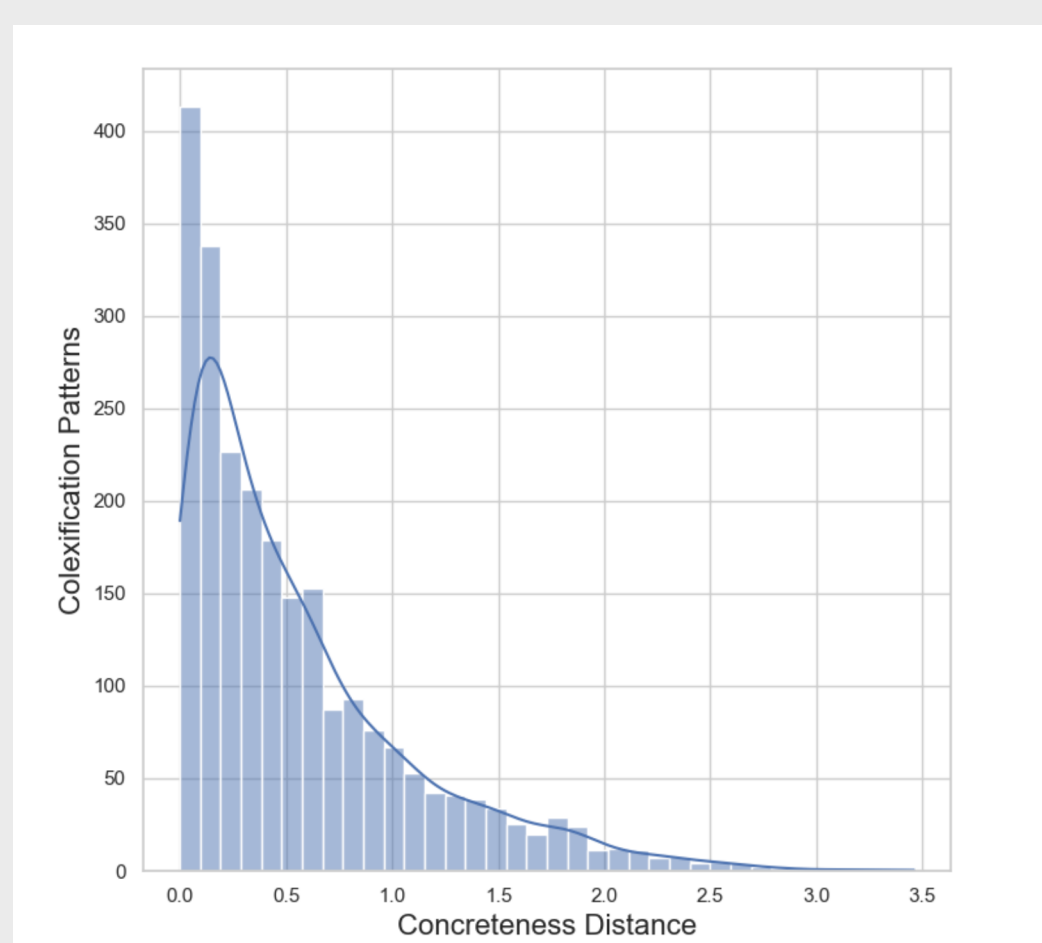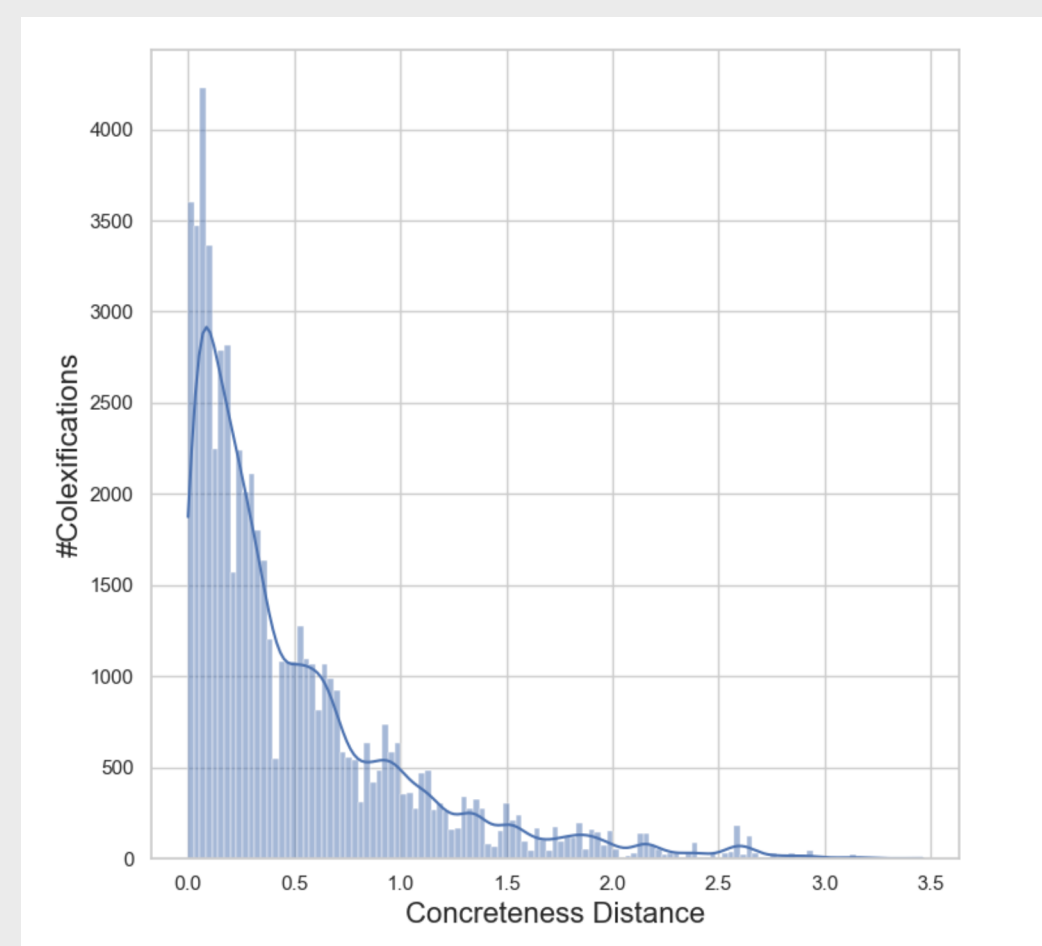
Histograms of Number of Colexifications against Concreteness Distances of the Colexified Pair of Concepts.
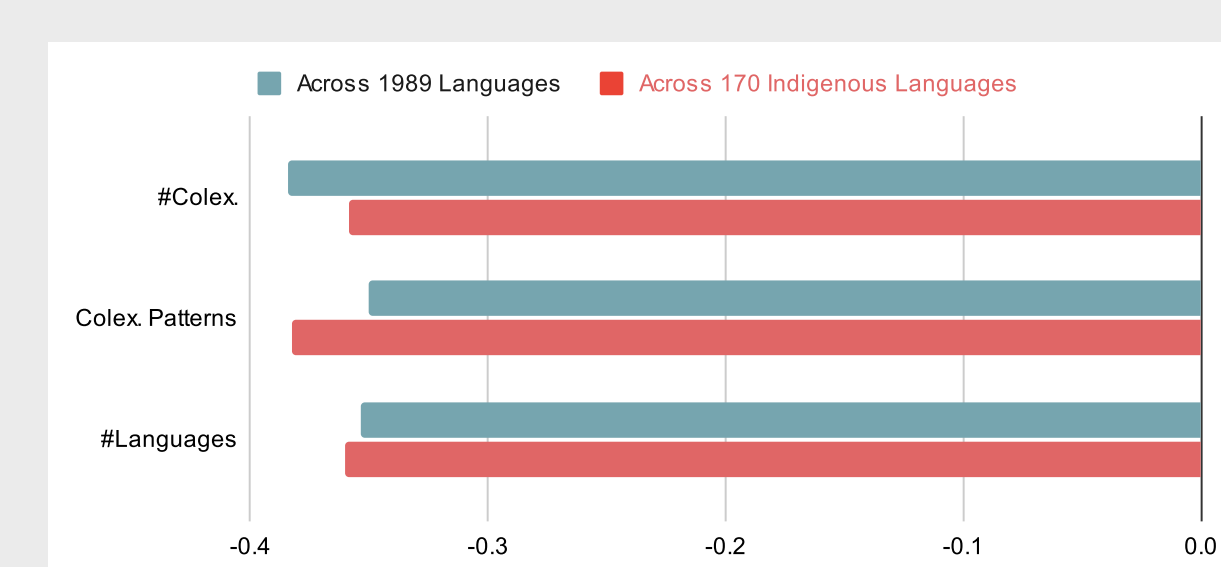
Pearson Coefficients between the Number of Colexifications and Concreteness Distances of the Colexified Pair of Concepts.

- Histograms reveal a notable disparity in the frequency of colexifications between the lower and higher ranges of concreteness distances for colexified concepts.
- Additionally, statistical analysis demonstrates a significant negative correlation between the number of colexifications across all dimensions and the concreteness distances.

Map of Languages in the curated dataset of colexifications from CLICS[3], color coded by macroareas: Eurasia, Afria, Australia, Papunesia, North America, South America. Each dot represents a language, and the sizes of the dots are proportional to the number of colexifications in the dataset.

3. Dataset Curation

| Language	Macroarea	Word(s)
Aikanã	South America	ya, yazune'i
Barasana-Eduria	South America	bũhíhũ
Cubeo	South America	awiã
Upper Chehalis	North America	luʼkʷaɬ
Cha'palaa	South America	'pahta
Mundurukú	South America	ka-ŝi
Naduhup	South America	wædhɔ´

CLICS[3]

Abstract (language based) — Concrete (experience based)
1 2 3 4 5
Brysbaert et al. (2014)

YEAR Conc=3.25
1.58
SUN Conc=4.83
0.07
MOON Conc=4.9
CLOCK Conc=5
0.17
UNKNOWN
DAY (24 HOURS) Conc=UNKNOWN

The numbers on the edges are the concreteness distances between colexified concepts, i.e. $|Conc_i - Conc_j|$.

5. Conclusions and Future Work

- Our findings challenge previous theories and findings on colexification and metaphoricity correlations, supported by diverse data from various concepts and languages.
- In future research, we aim to assess a stronger hypothesis: colexifications encode closeness in concreteness through the use of language/language family weights across concepts. This approach will contribute to the study of human conceptualizations, particularly aiding the development of large-scale lexical databases for interdisciplinary investigations, including psycholinguistics and multilingual natural language processing.

References

- Brysbaert, M., Warriner, A.B. & Kuperman, V. Concreteness ratings for 40 thousand generally known English word lemmas. Behav Res 46, 904–911 (2014). https://doi.org/10.3758/s13428-013-0403-5
- Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. Cognition, 201, 104280.
- Lakoff, George, and Mark Johnson. Metaphors we live by. University of Chicago press, 2008.
- Gardent, Claire, and Christian Retoré. "IWCS 2017—12th International Conference on Computational Semantics—Short papers." IWCS 2017—12th International Conference on Computational Semantics—Short papers. 2017.
- Naumann, Daniela, Diego Frassinelli, and Sabine Schulte im Walde. "Quantitative semantic variation in the contexts of concrete and abstract words." (2018).

GitHub: https://github.com/siebeniris/ColexPhon
yiyic@cs.aau.dk