The Unnatural Language ToolKit (ULTK): a software library for research in computational semantic typology

Nathaniel Imel
New York University
n.imel@nyu.edu

Christopher Haberland
University of Washington
haberc@uw.edu

Shane Steinert-Threlkeld
University of Washington
shanest@uw.edu

1 Introduction

This paper introduces the **U**nnatural **L**anguage **T**ool**k**it (ULTK), an open-source Python library for computational semantic typology research (`https://clmbr.shane.st/ultk/`). ULTK's key features include unifying data structures, algorithms for generating artificial languages, and data analysis tools for related computational experiments. The `language` module organizes the basic data structures for constructing meaning spaces, expressions, and languages. A `grammar` submodule contains methods for building and enumerating expressions from custom Language of Thought (Fodor, 1975, 2008; Quilty-Dunn et al., 2022) grammars, which allows for straightforward computation of minimum length descriptions for symbolically expressible semantic representations. This approach has been used successfully in many investigations of concept learning (Feldman, 2000; Goodman et al., 2015). The second main module of ULTK, `effcomm`, organizes efficient communication analyses, which have become popular styles of explanation in recent functionalist accounts of semantic universals (Kemp et al., 2018). This module contains functions for defining informativity based on literal and pragmatic communicative agents and algorithms for exploring the space of artificial languages.

After first elaborating on the structure of these two modules, we then provide two case studies, illustrating two major styles of explanation in computational semantic typology research: (1) an efficient communication analysis of modal semantic typology, and (2) an analysis of the relative ease of learning of monotone versus non-monotone quantifiers. ULTK's accessible design, documentation, and open-source nature are intended to reduce barriers for researchers when implementing computational linguistic typological experiments.

2 Language module

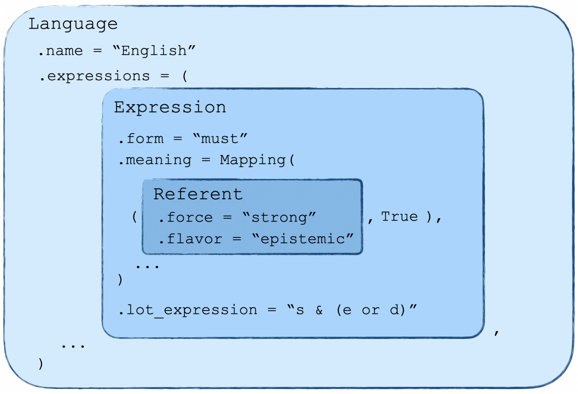

Figure 1: Structure of a `ultk.language.Language`, using the English modal vocabulary as an example.

In ULTK, a `Language` (Figure 1) is a collection of `Expressions`; an `Expression` is a mapping between a surface form and a `Meaning`; a `Meaning` maps a `Universe`'s `Referents` to an object of arbitrary type (e.g., `bool` if the meaning is boolean). A `Referent` is a wrapper for any hashable Python object, which could be as simple as an index or as complex as a model-theoretic structure. A `Universe` is a collection of `Referents`. In this way, a `bool` `Meaning` corresponds to the characteristic function of a set. To capture probabilistic meanings, it is natural to use `float` meanings.

Grammar This submodule contains classes and functions for building `Grammars` and generating expressions. These are often used for semantic representations: at its core, this module enables composing functions to arbitrary depth according to their input and output types. A `Grammar` is made up of arbitrary `Rules`, with `GrammaticalExpressions` formed by combining rules with licensed input and output types (Piantadosi, 2014). A `Rule` minimally consists of a name, a left-hand side (output type) a right-hand side (sequence of input types), a function to apply, and optionally a weight (for defining

Name	LHS	RHS Types	Function
and	`bool`	`bool` , `bool`	`λ p1, p2: p1 and p2`
or	`bool`	`bool` , `bool`	`λ p1, p2: p1 or p2`
not	`bool`	`bool`	`λ p: not p`
weak	`bool`	`Referent`	`λ m: m.force == "weak"`
strong	`bool`	`Referent`	`λ m: m.force == "strong"`
epistemic	`bool`	`Referent`	`λ m: m.flavor == "epistemic"`
deontic	`bool`	`Referent`	`λ m: m.flavor == "deontic"`
circumstantial	`bool`	`Referent`	`λ m: m.flavor == "circumstantial"`

Name	LHS	RHS Types	Function
union	`frozenset`	`frozenset` , `frozenset`	`λ s1, s2: s1 | s2`
intersection	`frozenset`	`frozenset` , `frozenset`	`λ s1, s2: s1 & s2`
cardinality	`int`	`frozenset`	`λ s: len(s)`
subset_eq	`bool`	`frozenset` , `frozenset`	`λ s1, s2: s1 < s2`
diff	`bool`	`frozenset` , `frozenset`	`λ s1, s2: s1 − s2`
empty	`bool`	`frozenset`	`λ s: len(s) == 0`
nonempty	`bool`	`frozenset`	`λ s: len(s) > 0`

Figure 2: The ULTK LoT grammars in our case studies, modals (top) and quantifiers (bottom, snippet).

probabilistic grammars). A `Grammar` can be initialized by loading a Python module with arbitrary functions parsed as `Rules`, or by loading a YAML file (Figure 2). The `grammar` submodule can be used to generate minimum length descriptions for `Meanings` in order to quantify their representational complexity for computational experiments; this can be done by depth-bounded enumeration (with user-specified uniqueness criteria) or by approximate Bayesian inference over PCFGs.

3 Effcomm module

The `effcomm` module provides tools for analyzing the communicative efficiency of languages. The `agent` and `informativity` submodules implement Rational Speech Act-style agents and enable the computation of literal and pragmatic informativity of languages (Frank and Goodman, 2012; Degen, 2023). These tools for measuring informative communication, together with tools from `language.grammar` for measuring the complexity of languages, can be combined to study how languages balance, or trade off, various pressures efficiently. The `effcomm` module also includes submodules for generating hypothetical languages through various sampling strategies (`sampling`), approximating Pareto-optimal solutions to efficiency trade-offs via an evolutionary optimization algorithm (`optimization`), and evaluating the languages' communicative properties (`tradeoff`). The `analysis` submodule provides utilities for visualizing language distributions in trade-off space. These components are designed to work together to support end-to-end efficient communication analyses of artificial or natural languages.

4 Case study 1: efficient communication for modals

Efficient communication has been proposed as an explanation for variation in semantic typology (Kemp et al., 2018). Using ULTK, we replicate Imel et al. (2024) by applying this analysis to modals. To do this, we (1) convert attested modal vocabularies to ULTK `Languages`, (2) generate artificial vocabularies, and (3) measure efficiency and a notion of *naturalness*. For the latter , we consider the degree to which a language satisfies the Independence of Force and Flavor (IFF) semantic universal (Steinert-Threlkeld et al., 2023). We define a modal `Universe` of (force, flavor) `Referents` and construct `Languages` as sets of `Expressions` mapping these referents to truth values (Fig. 1).

Languages Natural vocabularies are derived from a public database (Guo et al., 2022), while artificial ones are generated via ULTK's `language.sampling` and `effcomm.optimization` modules. The former samples meanings randomly while controlling IFF satisfaction, and the latter uses an evolutionary algorithm to approximate the Pareto frontier for the complexity/communicative cost trade-off. This step uses some convenience methods that ULTK provides for turning data in fieldwork-natural formats into its natural internal data structures that are needed for an efficient communication analysis; this helps lower the barrier-of-entry to conducting such analyses.

Efficient communication Complexity is measured as minimum description length in a boolean Language-of-Thought (LoT) (Kemp and Regier, 2012), using `language.grammar` to enumerate and cache shortest expressions. Communicative cost is measured in `effcomm.informativity`, which models literal communication and uses communicative need priors estimated from English news data. While the results we present here use literal speakers and listeners, ULTK offers convenience methods for iterating pragmatic agents to arbitrary depth from a given language (Frank and Goodman, 2012; Degen, 2023).

Results Figure 3 plots complexity vs. communicative cost, with artificial languages colored by

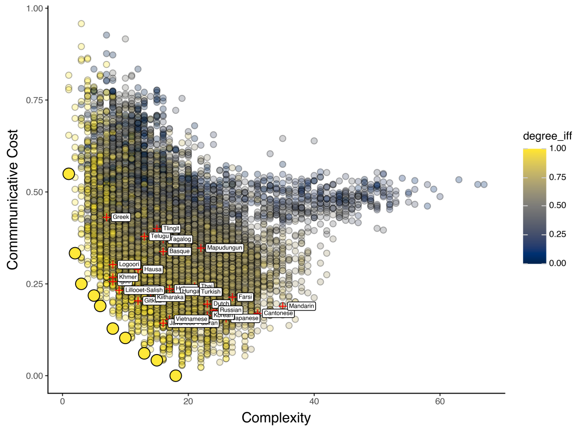

Figure 3: Replication of Imel et al. (2024) via ULTK; see text for details. Full demo available at `https://github.com/CLMBRs/ultk/tree/main/src/examples/modals`.

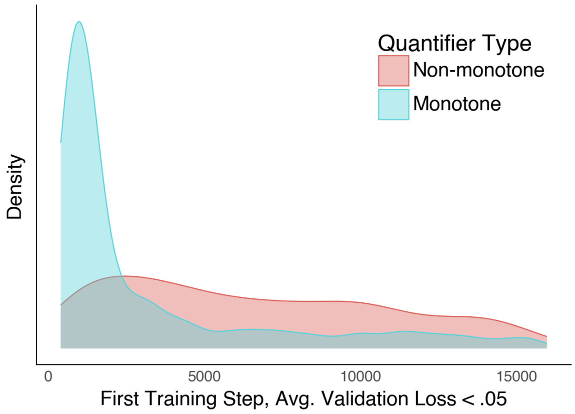

Figure 4: Replication of Haberland and Steinert-Threlkeld (2025) via ULTK; see text for details. Full demo available at `https://github.com/CLMBRs/ultk/tree/main/src/examples/learn_quant`.

naturalness and natural ones marked in red. Natural languages cluster closer to the Pareto frontier (large points) than chance ($t(15536) = 46$, $p \approx 0$), and naturalness negatively correlates with Pareto distance ($\rho = -0.38$, $p \approx 0$). Similar replications of efficient communication analyses (e.g., kinship (Kemp and Regier, 2012), indefinite pronouns (Denić et al., 2022), quantifiers (Steinert-Threlkeld, 2021), connectives (Uegaki, 2021)) are under development. ULTK provides abstractions and utilities that allow for relatively simple replication of these existing analyses and, therefore, makes it easy to conduct new ones as well.

5 Case study 2: ease of learning for (monotone) quantifiers

Semantic *universals* constrain natural linguistic meanings (Croft, 2003). For example, all simple determiners in natural languages are argued to be monotonic (Barwise and Cooper, 1981). A possible explanation is that monotone quantifiers are *easier to learn* (Steinert-Threlkeld and Szymanik, 2019; Chemla et al., 2019). Using ULTK, we replicate one of the results contained in Haberland and Steinert-Threlkeld (2025) (which contains full experimental details), showing that monotone quantifiers are easier to learn than non-monotone quantifiers. We generate a large number of quantifier expressions composed from a LoT grammar (see Figure 2). We measure ease of learning as the number of steps required by a neural model to learn to correctly judge the truth-value of a quantifier. We generate 2000 quantifiers from the LoT grammar and measure the speed at which both LSTM and Transformer models learn to verify expressions that are both monotonic and non-monotonic. We find that monotone quantifiers are typically learned much faster than non-monotone ones (Figure 4). This suggests that ease of learning may be a factor shaping the lexical semantic typology of the world's languages, at least in this domain (see (Steinert-Threlkeld, 2020; Steinert-Threlkeld and Szymanik, 2020; Maldonado et al., 2022; Maldonado and Culbertson, 2019; Strohmaier and Wimmer, 2022) for other case studies in other domains).

This case study demonstrates the potential of using ULTK to answer questions about the relation between semantic universals and ease of learning. In addition to the LoT grammar, the library provides basic tools for structuring `Meaning`s and other objects in a way that is consumable by external machine learning libraries.

6 Conclusion

The **U**nnatural **L**anguage **T**ool**k**it (ULTK) is an open-source library, enabling linguists to execute computational typological research. Our intention is to lower the barrier-to-entry to conducting efficient communication and ease-of-learning analyses of typological phenomena. In the limit, typologists and fieldworkers will be able to input data structured in natural ways, and the library will facilitate analyses in these these domains. The two case studies presented here demonstrate the possibility of this division of labor and the utility of the ULTK library. Future work will expand both coverage of methods and improve the ease of use to continue making this dream a reality. We welcome submissions of contributions, questions, and suggestions to our code repository (`https://github.com/CLMBRs/ultk`).

References

Jon Barwise and Robin Cooper. 1981. Generalized quantifiers and natural language. In *Philosophy, language, and artificial intelligence: Resources for processing natural language*, pages 241–301. Springer.

Emmanuel Chemla, Brian Buccola, and Isabelle Dautriche. 2019. Connecting Content and Logical Words. *Journal of Semantics*, 36(3):531–547.

William Croft. 2003. *Typology and universals*. Cambridge University Press.

Judith Degen. 2023. The Rational Speech Act Framework. *Annual Review of Linguistics*, 9(1):519–540.

Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. 2022. Indefinite Pronouns Optimize the Simplicity/Informativeness Trade-Off. *Cognitive Science*, 46(5):e13142.

Jacob Feldman. 2000. Minimization of Boolean complexity in human concept learning. *Nature*, 407(6804):630–633. Publisher: Nature Publishing Group.

Jerry A Fodor. 1975. *The language of thought*, volume 5. Harvard university press.

Jerry A Fodor. 2008. *LOT 2: The Language of Thought Revisited*, 1 edition. Oxford University Press, Oxford.

Michael C. Frank and Noah D. Goodman. 2012. Predicting Pragmatic Reasoning in Language Games. *Science*, 336(6084):998–998.

Noah D. Goodman, Joshua B. Tenenbaum, and Tobias Gerstenberg. 2015. Concepts in a Probabilistic Language of Thought. In *The Conceptual Mind: New Directions in the Study of Concepts*. The MIT Press. _eprint: https://direct.mit.edu/book/chapter-pdf/2271061/9780262326865_cav.pdf.

Qingxia Guo, Nathaniel Imel, and Shane Steinert-Threlkeld. 2022. A Database for Modal Semantic Typology. In *Proceedings of the 4th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 42–51, Seattle, Washington. Association for Computational Linguistics.

Christopher Haberland and Shane Steinert-Threlkeld. 2025. Quantifiers that are more monotone are easier to learn. In *Proceedings of Semantics and Linguistic Theory (SALT 35)*, volume 35.

Nathaniel Imel, Qingxia Guo, and Shane Steinert-Threlkeld. 2024. An Efficient Communication Analysis of Modal Typology. LingBuzz Published In:.

Charles Kemp and Terry Regier. 2012. Kinship categories across languages reflect general communicative principles. *Science (New York, N.Y.)*, 336(6084):1049–1054.

Charles Kemp, Yang Xu, and Terry Regier. 2018. Semantic typology and efficient communication. *Annual Review of Linguistics*, pages 1–23.

Mora Maldonado and Jennifer Culbertson. 2019. Something about \emph{us}: Learning first person pronoun systems. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci 2019)*.

Mora Maldonado, Jennifer Culbertson, and Wataru Uegaki. 2022. Learnability and constraints on the semantics of clause-embedding predicates. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44).

Steven T. Piantadosi. 2014. LOTlib: Learning and Inference in the Language of Thought. Published: available from https://github.com/piantado/LOTlib.

Jake Quilty-Dunn, Nicolas Porot, and Eric Mandelbaum. 2022. The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences. *Behavioral and Brain Sciences*, pages 1–55.

Shane Steinert-Threlkeld. 2020. An Explanation of the Veridical Uniformity Universal. *Journal of Semantics*, 37(1):129–144.

Shane Steinert-Threlkeld. 2021. Quantifiers in Natural Language: Efficient Communication and Degrees of Semantic Universals. *Entropy. An International and Interdisciplinary Journal of Entropy and Information Studies*, 23(10):1335.

Shane Steinert-Threlkeld, Nathaniel Imel, and Qingxia Guo. 2023. A semantic universal for modality. *Semantics and Pragmatics*, 16:1:1–20.

Shane Steinert-Threlkeld and Jakub Szymanik. 2019. Learnability and semantic universals. *Semantics and Pragmatics*, 12:4:1–39.

Shane Steinert-Threlkeld and Jakub Szymanik. 2020. Ease of learning explains semantic universals. *Cognition*, 195:104076.

David Strohmaier and Simon Wimmer. 2022. Contrafactives and Learnability. *Proceedings of the Amsterdam Colloquium*, pages 305–312.

Wataru Uegaki. 2021. The informativeness / complexity trade-off in the domain of Boolean connectives. *Linguistic Inquiry*.