Quantifiers in natural language optimize the simplicity/informativeness trade-off[*]

Shane Steinert-Threlkeld

Department of Linguistics
University of Washington
shanest@uw.edu

Abstract

While the languages of the world vary greatly, linguists have discovered many restrictions on possible variation. Semantic universals are restrictions on the range of variation in meaning across languages. Recently, in several domains—e.g. kinship terms, color terms—such universals have been argued to arise from a trade-off between simplicity and informativeness.

In this paper, we apply this method to a prominent domain of functions words, showing that the *quantifiers* in natural language also appear to be optimized for this trade-off. We do this by using an evolutionary algorithm to estimate the optimal languages, systematically manipulating the *degree of naturalness* of languages, and showing that languages become closer to optimal as they become more natural.

Our results suggest that very general communicative and cognitive pressures may shape the lexica of natural languages across both content and function words.

1   Introduction

While the languages of the world vary greatly, linguists have discovered many restrictions on possible variation (Croft, 1990; Hyman, 2008; von Fintel and Matthewson, 2008). *Semantic universals* are restrictions on the range of variation in meaning across languages. Recently, in several domains—kinship terms, color terms—such universals have been argued to arise from a trade-off between simplicity and informativeness (Kemp and Regier, 2012; Kemp et al., 2018). Roughly: a language cannot be both maximally simple (in terms of, e.g. cognitive load) and at the same time maximally informative. Intuitively, a maximally simple language would have a single term, which could not be used to convey significant information. A maximally informative language, on the other hand, would contain individual expressions for every possible thought to be expressed; such a language is highly complex, relying on significant memorization. The general claim: the semantic systems of the world's languages optimally balance these two competing pressures.

While the aforementioned case studies apply to domains of content words, the historically most prominent domain of semantic universals has been from a domain of function words, namely determiners (Barwise and Cooper, 1981; Peters and Westerståhl, 2006). In particular, the quantifiers expressed by determiners have been argued to have properties like monotonicity, quantitativeness, and conservativity. Recent work has offered a different explanation for these

universals in quantifiers, namely that they arise from a pressure of learnability: quantifiers satisfying the universals are easier to learn than those that do not, and therefore get lexicalized (Steinert-Threlkeld and Szymanik, 2019, 2020). This argument, however, does not rule out the possibility of explaining these universals in terms of the aforementioned trade-off.

In this paper, we argue that the semantic universals for quantifiers can in fact also be seen as arising from the trade-off between simplicity and informativeness. In the next section, we develop methods to make this argument, introducing measures for simplicity and informativeness that can apply to quantifier systems, a method for defining a degree of naturalness for artificial languages, and a novel technique for estimating the set of optimal languages (i.e. Pareto frontier) via an evolutionary algorithm. This new methodology allows us to perform statistical tests on the factors influencing the optimality of quantifier systems. Section 3 presents a regression showing that as languages become closer to being a natural language, they also become closer to being optimal. We then conclude with a discussion of the consequences of the main result and future directions.

2 Methods

In order to argue that more natural languages are more optimized for the trade-off, we need methods for generating artificial languages of varying degrees of naturalness and measuring how optimal they are. After presenting some preliminary background about quantifiers and precisely defining the measures of simplicity and informativeness, Sections 2.3 and 2.4 turn to those two tasks. Code for reproducing all of the results in this paper and carrying out further experiments may be found at `https://github.com/shanest/SimInf_Quantifiers`.

We are measuring a *degree* of naturalness, instead of just comparing actual natural languages to artifical languages, for two reasons. On the one hand, in the case of quantifiers, unlike the cases of kinship (Kemp and Regier, 2012) and color (Regier et al., 2015), there is not yet an existing catalog of the quantifier systems of any significant number of natural languages. On the other hand, even in the cases where such data does exist, finding a correlation between a degree of naturalness and optimality provides more information and may strengthen the argument by linking to language change/evolution. As languages have changed over time, they became 'closer' to our current natural languages; a correlation result of the kind we are developing shows that during this process, they also become closer to optimally trading off the two competing pressures.

2.1 Preliminaries

We represent quantifiers—the denotations of determiners like *all*, *some*, *most*, etc.—as sets of models $\langle M, A, B \rangle$ containing a domain of discourse and two distinguished subsets for the restrictor and nuclear scope (equivalently: a binary relation between those sets) (Barwise and Cooper, 1981; Peters and Westerståhl, 2006; Szymanik, 2016). For example:

$$[\![\text{every}]\!] = \{\langle M, A, B \rangle \colon A \subseteq B\}$$
$$[\![\text{at\_most\_3}]\!] = \{\langle M, A, B \rangle \colon |A \cap B| \leq 3\}$$
$$[\![\text{most}]\!] = \{\langle M, A, B \rangle \colon |A \cap B| > |A \setminus B|\}$$

In what follows, a *language* is a set of quantifiers. For computational reasons—namely because of the exponential growth of the space of possible models—we restrict ourselves to all models up to size 10 in the remainder of this paper.

2.2   Measuring Simplicity and Informativeness

Our measure of cognitive simplicity relies on representing quantifiers in a Language of Thought (Feldman, 2000; Piantadosi et al., 2016), i.e. using formulas in a logical language containing operations for set union, intersection, and complementation, as well as for measuring cardinalities and comparing, multiplying, and dividing them. Table 1 shows

Boolean	Set-Theoretic	Numeric
$\wedge, \vee, \neg$	$\cap, \cup, \subset, \lvert \cdot \rvert$	$/, +, -, >, =, \%$

Table 1: The operators in the grammar for generating quantifiers.

the entire set of operators used in this paper. The *complexity* of a quantifier is the length of the *shortest* formula in this language that denotes the quantifier. We found the shortest such formula by exhaustively enumerating all formulas with up to 12 operations and comparing the truth-values across all models up to size 10.[1,2] The complexity of a language is the sum of the complexities of the quantifiers in it. We specify an upper bound on the number of possible quantifiers in a language (10 in our experiments) and divide the sum by this number.

Our measure of informativeness stems from notions of communicative success: a speaker has an intended model that they want to communicate to a listener using the quantifiers in their language (Skyrms, 2010; Kemp et al., 2018). This is captured by the following:

$$I(L) := \sum_{\mathbb{M}} P(\mathbb{M}) \sum_{Q \in L} P(Q|\mathbb{M}) \sum_{\mathbb{M}' \in Q} P(\mathbb{M}'|Q) \cdot u(\mathbb{M}', \mathbb{M})$$

The prior over models, as well as the conditional distributions, are assumed to be uniform where defined (e.g. $P(Q|\mathbb{M}) = 1/n$ if $\mathbb{M} \in Q$, 0 otherwise, where $n = |\{Q \in L : \mathbb{M} \in Q\}|$ is the number of quantifiers in L containing $\mathbb{M}$).

This measure captures the following communicative scenario: a speaker has a model ($\mathbb{M}$) in mind, that it wishes to communicate to a listener. To do so, they can use the quantifiers in the language L. The speaker's behavior is captured by $P(Q|\mathbb{M})$. The listener then guesses a model ($\mathbb{M}'$) that the speaker has in mind, with probability $P(\mathbb{M}'|Q)$.

The utility $u(\mathbb{M}', \mathbb{M})$ measures how good it is for the listener to guess $\mathbb{M}'$ when the speaker had in mind $\mathbb{M}$. We base this on a measure of the distance between models, capturing the notion that non-exact matches can still be better or worse (Jäger, 2007; O'Connor, 2014). More precisely:

$$u(\mathbb{M}', \mathbb{M}) = \frac{1}{1 + d(\mathbb{M}', \mathbb{M})} \quad \text{where} \quad d(\mathbb{M}', \mathbb{M}) = \sum_{X \in A \setminus B, A \cap B, B \setminus A, M \setminus (A \cup B)} \max\{0, |X| - |X'|\}$$

Intuitively, this measure is inversely proportional to how many elements one has to move to transform the listener's guessed model into the sender's model (by summing this value across the four 'zones' in a model of the form $\langle M, A, B \rangle$).[3] For example, suppose $\mathbb{M}$ has 3, 4, 2, and 1 elements in $A \setminus B, A \cap B, B \setminus A, M \setminus (A \cup B)$ respectively, and $\mathbb{M}'$ has 2, 4, 3, and 1 elements

---

[1]For memory reasons, we collapse isomorphic models, representing a model $\langle M, A, B \rangle$ by the cardinalities of the four sets $A \cap B, A \setminus B, B \setminus A, M \setminus (A \cup B)$. This prevents us from capturing quantifiers like *the first three* which do not satisfy the universal known as QUANTITY (Steinert-Threlkeld and Szymanik, 2019). Future work will explore methods that relax this assumption while simultaneously addressing the resulting combinatorial explosion.

[2]Using length only is equivalent to using the probability of generating an expression with a PCFG that assigns equal weight to all productions from the same non-terminal.

[3]The addition of 1 in the denominator both prevents division by zero and makes distance-0 models have maximal utility of 1.

in the same zones. We then have that $d(\mathbb{M}, \mathbb{M}') = 1$, since moving one element from $A \setminus B$ to $B \setminus A$ will make the four zones have the same size in the two models.

2.3   Sampling Languages

To answer the question of whether natural languages optimize the trade-off between these two measures, we need to (i) define artificial languages, (ii) identify the natural languages, and (iii) compare how well each does at this optimization. For (i) and (ii), we systematically control 'how natural' a language is by biased sampling. A completely random language can be generated by randomly sampling a specified number of quantifiers from the space of all quantifiers generated by our grammar.

While there is no existing dataset of quantifiers across a large set of natural languages, a major cross-linguistic study (Keenan and Paperno, 2012; Paperno and Keenan, 2017) found that all natural language quantifiers belonged to three classes:

- Generalized existential: depending only on $|A \cap B|$.

  For example: $[\![\text{at least three}]\!] = \{\langle M, A, B \rangle \colon |A \cap B| \geq 3\}$.

- Generalized intersective: depending only on $|A \setminus B|$.

  For example: $[\![\text{every}]\!] = \{\langle M, A, B \rangle \colon |A \setminus B| = 0\}$.

- Proportional: comparing $|A \cap B|$ and $|A \setminus B|$.

  For example: $[\![\text{most}]\!] = \{\langle M, A, B \rangle \colon |A \cap B| > |A \setminus B|\}$.

We call a quantifier *quasi-natural* if it can be expressed in one of the three forms above. And a language will be considered natural if it contains only quasi-natural quantifiers.

Our complete sampling procedure, then, worked as follows: for each number of words between 1 and 10, we generated 8000 languages. Each language of size n was chosen to have $m \leq n$ quasi-natural quantifiers, with m chosen uniformly from $\{0, \ldots, n\}$. All remaining quantifiers were chosen randomly from the set of all quantifiers whose minimal formula has 12 or fewer operators. We refer to m/n as the *degree of naturalness* of a language. Thus, a language that has only quasi-natural quantifiers will have a degree of naturalness of 1 (and a language that has no quasi-natural quantifiers will have a degree of naturalness of 0).

2.4   Measuring Optimality

For (iii), we need a measure of optimality for a language, to see how it relates to the degree of naturalness. To do this, we measure how close a language is to the *Pareto frontier*, the set of languages which are not dominated (i.e. which have no language both simpler and more informative). The Pareto frontier contains the fully optimal languages: they cannot be made less complex or more informative without becoming worse on the other dimension. Writing P for the Pareto frontier, we define the optimality of a language as

$$\texttt{optimality}(L) := 1 - \min_{L' \in P} d(L, L')$$

where d is the Euclidean distance between points in the plane. This measure takes the *closest* point on the Pareto frontier to a given language. If a language is on the frontier, i.e. is optimal, that minimum distance will be 0, and so the degree of optimality will be 1. Because both communicative cost and complexity range from 0 to 1, the theoretically largest value for the

minimum distance is 1, and so `optimality` also ranges from 0 to 1. To summarize: the degree of optimality of a language increases as it gets closer to the Pareto frontier, the set of optimal languages.

A complication arises when trying to apply this measure: because the space of possible languages is enormous, we cannot exhaustively enumerate it and thereby uncover the *true* Pareto frontier. Moreover, our sampling procedures from the previous section are not guaranteed to uncover the Pareto frontier. Because of this, we need a method to estimate the Pareto frontier without being able to calculate it directly.

To estimate the true Pareto frontier, we used an evolutionary algorithm (Coello et al., 2007). Such algorithms take inspiration from evolutionary processes in that points in a space change over a sequence of generations, with 'children' arising via 'mutation' from previous points. More importantly, such algorithms are explicitly designed to solve *multi-objective optimization* problems. Since the Pareto frontier can be seen as the set of solutions to the problem of simultaneously optimizing multiple objectives (simplicity and informativeness), these algorithms are well-suited to estimating it.

Our algorithm—provided in full detail in Algorithm 1 in the Appendix—can be intuitively described as follows. We start with an initial seed of randomly generated languages. For some specified number of 'generations' we select the *dominant* languages among the current set of languages. Each language then has an equal number of 'children' languages (enough to maintain the size of the pool of languages). A child arises from a parent language by some small sequence of 'mutations'. In our case, this was between 1 and 3 mutations, where a mutation could be: (i) deleting a quantifier from the parent language, (ii) adding a quantifier to the parent language, or (iii) swapping a quantifier in the parent language (i.e. deleting one and adding a new one).

(1) Apply evolutionary algorithm to simultaneously optimize the two objectives.

(2) Find the dominating points among those from (1) and from the sampling procedure.

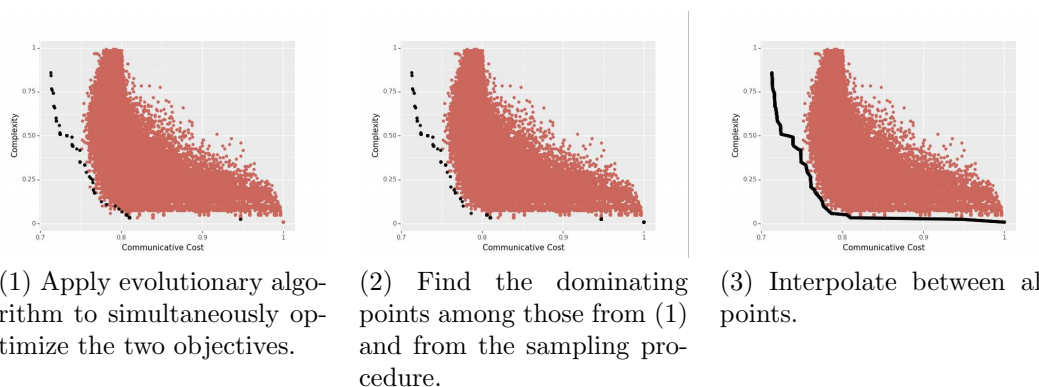(3) Interpolate between all points.

Figure 1: The overall Pareto frontier estimation algorithm, in three steps. The red points are the languages sampled as described in Section 2.3. The black points in panel (3) constitute the final estimate of the true Pareto frontier.

After running the above algorithm for some specified number of generations, we then take the dominant languages from the pool together with the languages we previously sampled, and then linearly interpolate between all of the points to form a smooth and dense frontier.[4] This entire process is depicted in Figure 1.

---

[4]More sophisticated evolutionary algorithms specify a convergence criterion. We leave the explorations of these refinements to future work.

3 Results

The main results can be seen in Figure 2. In this figure, the x-axis is communicative cost, which is $1 - I(L)$, and the y-axis is complexity. Each point represents a possible language, with the color of a point corresponding to degree of naturalness. The black line is the estimated Pareto frontier, i.e. the set of languages that optimally trade-off between these two factors.

A few things can be observed right away. All of the sampled points that were found to lie on the estimated Pareto frontier (i.e. which dominate all languages both sampled and discovered by the evolutionary algorithm) appear to have a very high degree of naturalness. These are the yellow points on the black frontier, where no brown or blue points (less natural languages) are to be found. Moreover, this seems to be a general trend: it appears that languages with a high degree of naturalness tend to be closer to the Pareto frontier than those with low degrees of naturalness.[5]
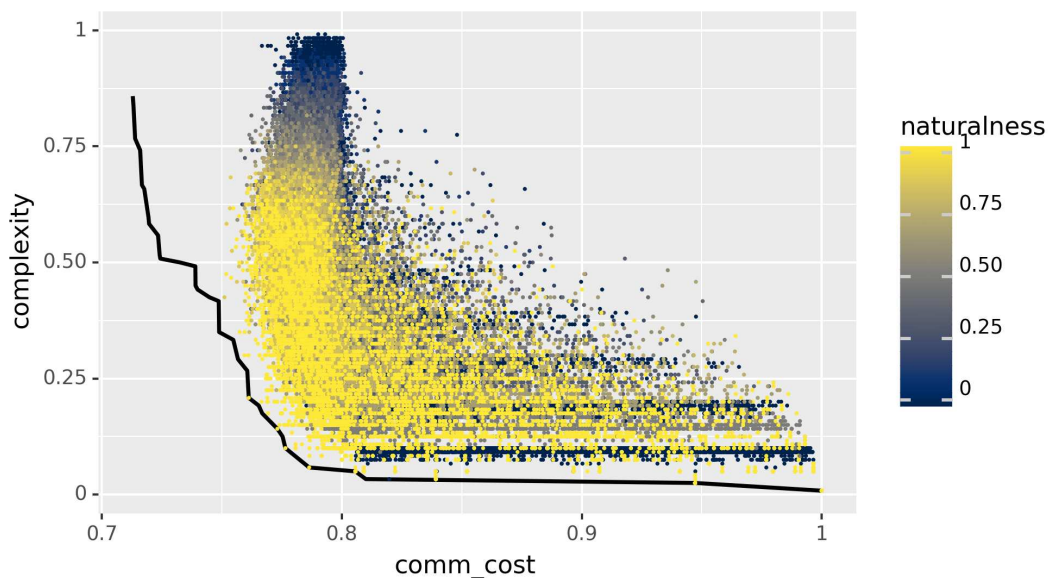

Figure 2: Languages in the space of communicative cost and complexity, colored by their degree of naturalness. Languages with more quasi-natural quantifiers appear to be closer to optimal, as measured by closeness to the (estimated) Pareto frontier, depicted in black.

In virtue of the methods described in the previous section, we can test this appearance statistically: a regression reveals that there is a significant positive correlation between the degree of optimality and degree of naturalness of a language ($\beta = 0.30$, $t = 88.95$, $p \approx 0$, 95% CI: $[0.293, 0.307]$). In other words, as languages become more similar to natural languages with respect to their quantifiers, they come closer to optimally trading off between the competing pressures of cost and complexity.

---

[5]A possible exception to the general trend lies in the bottom-right corner. Exploring the properties of those languages remains for future work. Thanks to an anonymous referee for this suggestion.

4   Discussion

The languages of the world do not express all logically possible meanings, but only a restricted subset thereof. The results of this paper show that in the domain of quantifiers, natural languages appear to be optimizing a trade-off between simplicity and informativeness. We demonstrated this by estimating the Pareto frontier and finding a significant correlation between the degree of naturalness of a language and its closeness to the frontier. This suggests that very general communicative and cognitive pressures may shape the lexica of natural languages across both content and function words (Chemla et al., 2019).

Much work remains to be done. Directly concerning the results in the present paper, a few natural directions emerge. Firstly, while we argued that the statistical analysis presented here provides more information than simply comparing natural and artificial languages, more detailed and systematic documentation of the quantifier systems of the actual languages of the world would provide a stronger empirical foundation for the theory. Secondly, at several junctures, we made concrete modeling choices (e.g. what operators to use in our grammar, the size of models and of languages, the parameters of the evolutionary algorithm). While on quick inspection our main result seems robust, a more detailed analysis of sensitivity to those parameters would be welcome. Thirdly, alternative methods for measuring naturalness can be developed. For instance, Carcassi et al. (2019) provide an information-theoretic measure of the degree of monotonicity of a quantifier, which extends naturally to other universals as well. This raises the question: are more monotone quantifiers closer to the Pareto frontier than less monotone ones? Finally, alternative methods of sampling languages should be explored. Figure 2 shows that our evolutionary algorithm explores a region of the space that the naive sampling procedures do not discover (e.g. in the top-left corner). Because exhaustive enumeration is not feasible, more sophisticated sampling methods that sufficiently explore low-density regions of the space[6] will help make the results more robust.

Most generally, this work raises a number of interesting questions about how to distinguish between alternative explanations for semantic universals. The present results contribute to a body of literature arguing that semantic variation can be explained in terms of the simplicity/informativeness trade-off, across a range of domains (Kemp and Regier, 2012; Regier et al., 2015; Kemp et al., 2018; Gibson et al., 2019). At the same time, a growing body of literature argues that semantic typology—often in the same domains (e.g. quantifiers, color terms)—reflects relative ease of learning (Steinert-Threlkeld and Szymanik, 2019, 2020; Steinert-Threlkeld, 2019; van de Pol et al., 2019; Saratsli et al., 2019; Maldonado and Culbertson, 2019). Future studies can and should be developed to probe whether these explanations are in conflict and, if so, which is to be preferred (while remaining open to the possibility that different factors may explain the typological facts in different domains). One promising avenue for this work comes from integration with explicit models of language change and evolution. If, for example, as languages change, they appear to be regularly and continuously optimiz*ing* one of the above factors but not the other, that would provide evidence in its favor.

A   Estimating the Pareto Frontier

The complete algorithm for estimating the Pareto frontier, described in Section 2.4, appears as Algorithm 1. It is a simplified version of the non-dominated genetic sorting algorithm

---

[6]For example: while executing the evolutionary algorithm, we can store all of the languages that are generated, in addition to the pool that constitutes the current generation.

Algorithm 1 Estimating the Pareto Frontier

Parameters: `num_generations`, `num_langs`
Inputs: set of languages L, Pareto dominance method FIND_DOMINANT, INTERPOLATE method

    function GENETIC_ESTIMATE(`num_generations`, `num_langs`)
        `languages` ← SAMPLE_RANDOM_LANGUAGES(`num_langs`)
        for $i = 1, \ldots, $ `num_generations` **do**
            `dominant_languages` ← FIND_DOMINANT(`languages`)
            `languages` ← SAMPLE_MUTATED(`dominant_languages`, `num_langs`)
        end for
        return `languages`
    end function

    function SAMPLE_MUTATED(`languages`, `amount`)
        `amount_per_lang`, `amount_random` ← `amount`$/|$`languages`$|$
        `mutated_languages` ← $[]$
        for `language` $\in$ `languages` **do**
            for $i = 1, \ldots, $ `amount_per_lang` **do**
                Add MUTATE(`language`) to `mutated_languages`
            end for
        end for
        for $i = 1, \ldots, $ `amount_random` **do**
            `language` ← RANDOM_CHOICE(`languages`)
            Add MUTATE(`language`) to `mutated_languages`
        end for
        return `mutated_languages`
    end function

    function MUTATE(`language`)
        `mutated_language` ← `language`
        `num_mutations` ← RANDOM_CHOICE($[1, 2, 3]$)
        for $i = 1, \ldots, $ `num_mutations` **do**
            `mutation` ← RANDOM_CHOICE(
                      {ADD_QUANTIFIER, REMOVE_QUANTIFIER, SWAP_QUANTIFIER})
            `mutated_language` ← MUTATION(`language`)
        end for
        return `mutated_language`
    end function

    `estimate` ← GENETIC_ESTIMATE(`num_generations`, `num_langs`)
    `pareto_frontier` ← FIND_DOMINANT(`estimate` $\cup L$)
    `pareto_frontier` ← INTERPOLATE(`pareto_frontier`)

(Srinivas and Deb, 1994). There are two main parameters: how many generations to run the algorithm for (num_generations), and how large a pool of languages to maintain at each generation (num_langs). For the experiments in this paper, we set num_generations to 100 and num_langs to 2000. The set of languages L is the full set that we sampled according to the procedures described in Section 2.3.

The three final lines in the algorithm correspond to the three steps described in Figure 1 above. The method GENETIC_ESTIMATE contains the basic loop over generations. For finding dominant languages, we used the `pygmo` library's `non_dominated_front_2d` method (Biscani and Izzo, 2019). SAMPLE_MUTATED generates the new population at each generation by giving each dominant language its offspring. The function MUTATE performs the mutation of a single language, by choosing a number of mutations to apply and then randomly choosing from the available mutations.

References

Jon Barwise and Robin Cooper. Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, 4(2):159–219, 1981. doi: 10.1007/BF00350139.

Francesco Biscani and Dario Izzo. pagmo/pygmo2. 2019. doi: 10.5281/zenodo.1045336. URL https://esa.github.io/pagmo2/.

Fausto Carcassi, Shane Steinert-Threlkeld, and Jakub Szymanik. The emergence of monotone quantifiers via iterated learning. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci 2019)*, pages 190–196, 2019. URL https://psyarxiv.com/8swtd.

Emmanuel Chemla, Brian Buccola, and Isabelle Dautriche. Connecting Content and Logical Words. *Journal of Semantics*, 2019. doi: 10.1093/jos/ffz001.

Carlos A Coello Coello, Gary B Lamont, and David A van Veldhuizen. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Springer, 2007. doi: 10.1007/978-0-387-36797-2.

William Croft. *Typology and Universals*. Cambridge University Press, Cambridge, 1990.

Jacob Feldman. Minimization of Boolean Complexity in Human Concept Learning. *Nature*, 407:630–633, 2000. doi: 10.1038/35036586.

Edward Gibson, Richard Futrell, Steven T Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, 2019. doi: 10.1016/j.tics.2019.02.003.

Larry M. Hyman. Universals in phonology. *The Linguistic Review*, 25(1-2):83–137, 2008. doi: 10.1515/TLIR.2008.003.

Gerhard Jäger. The evolution of convex categories. *Linguistics and Philosophy*, 30(5):551–564, 2007. doi: 10.1007/s10988-008-9024-3.

Edward L Keenan and Denis Paperno, editors. *Handbook of Quantifiers in Natural Language*, volume 90 of *Studies in Linguistics and Philosophy*. Springer, Dordrecht, 2012. doi: 10.1007/978-94-007-2681-9.

Charles Kemp and Terry Regier. Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054, 2012. doi: 10.1126/science.1218811.

Charles Kemp, Yang Xu, and Terry Regier. Semantic typology and efficient communication. *Annual Review of Linguistics*, pages 1–23, 2018. doi: 10.1146/annurev-linguistics-011817-045406.

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

Cailin O'Connor. The Evolution of Vagueness. *Erkenntnis*, pages 1–21, 2014. doi: 10.1007/s10670-013-9463-2.

Denis Paperno and Edward L Keenan, editors. *Handbook of Quantifiers in Natural Language: Volume II*, volume 97 of *Studies in Linguistics and Philosophy*. Springer, 2017. doi: 10.1007/978-3-319-44330-0.

Stanley Peters and Dag Westerståhl. *Quantifiers in Language and Logic*. Clarendon Press, Oxford, 2006.

Steven T Piantadosi, Joshua B Tenenbaum, and Noah D Goodman. The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review*, 123(4):392–424, 2016. doi: 10.1037/a0039980.

Terry Regier, Charles Kemp, and Paul Kay. Word meanings across languages support efficient communication Informativeness and simplicity as competing principles. In Brian MacWhinney and William O'Grady, editors, *The Handbook of Language Emergence*, pages 237–263. Wiley, 2015.

Dionysia Saratsli, Stefan Bartell, and Anna Papafragou. Are Cross-Linguistically Frequent Semantic Systems Easier to Learn? The Case of Evidentiality. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci 2019)*, 2019.

Brian Skyrms. *Signals: Evolution, Learning, and Information*. Oxford University Press, 2010.

N. Srinivas and Kalyanmoy Deb. Muiltiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation*, 2(3):221–248, 1994. doi: 10.1162/evco.1994.2.3.221.

Shane Steinert-Threlkeld. An Explanation of the Veridical Uniformity Universal. *Journal of Semantics*, 2019. doi: 10.1093/jos/ffz019. URL https://semanticsarchive.net/Archive/DI5ZTNmN/UniversalResponsiveVerbs.pdf.

Shane Steinert-Threlkeld and Jakub Szymanik. Learnability and Semantic Universals. *Semantics & Pragmatics*, 12(4), 2019. doi: 10.3765/sp.12.4.

Shane Steinert-Threlkeld and Jakub Szymanik. Ease of Learning Explains Semantic Universals. *Cognition*, 195, 2020. doi: 10.1016/j.cognition.2019.104076.

Jakub Szymanik. *Quantifiers and Cognition: Logical and Computational Perspectives*, volume 96 of *Studies in Linguistics and Philosophy*. Springer, 2016. doi: 10.1007/978-3-319-28749-2.

Iris van de Pol, Shane Steinert-Threlkeld, and Jakub Szymanik. Complexity and learnability in the explanation of semantic universals of quantifiers. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci 2019)*, pages 3015–3021, 2019. URL https://psyarxiv.com/f8dbp/.

Kai von Fintel and Lisa Matthewson. Universals in semantics. *The Linguistic Review*, 25:139–201, 2008. doi: 10.1515/TLIR.2008.004.