Recursive numeral systems optimize the trade-off between lexicon size and average morphosyntactic complexity*

Milica Denić[1] and Jakub Szymanik[2]

[1] Tel Aviv University, Tel Aviv, Israel
milica.z.denic@gmail.com
[2] University of Trento, Trento, Italy
jakub.szymanik@gmail.com

Abstract

Human languages vary in terms of which meanings they lexicalize, but there are important constraints on this variation. It has been argued that languages are under the pressure to be simple (e.g., to have a small lexicon size) and to allow for an informative (i.e., precise) communication with their lexical items, and that which meanings get lexicalized may be explained by languages finding a good way to trade off between these two pressures ([12] and much subsequent work). However, in certain semantic domains, it is possible to reach very high levels of informativeness even if very few meanings from that domain are lexicalized. This is due to productive morphosyntax, which may allow for construction of meanings which are not lexicalized. Consider the semantic domain of natural numbers: many languages lexicalize few natural number meanings as monomorphemic expressions, but can precisely convey any natural number meaning using morphosyntactically complex numerals. In such semantic domains, lexicon size is not in direct competition with informativeness. What explains which meanings are lexicalized in such semantic domains? We will propose that in such cases, languages are (near-)optimal solutions to a different kind of trade-off problem: the trade-off between the pressure to lexicalize as few meanings as possible (i.e, to minimize lexicon size) and the pressure to produce as morphosyntactically simple utterances as possible (i.e, to minimize average morphosyntactic complexity of utterances).

1 Introduction

Human languages vary in terms of which meanings they lexicalize (as simple morphemes).[1] There are nonetheless important constraints on this variation: some meanings are frequently lexicalized, while others rarely. For instance, color terms label only convex regions of conceptual color space cross-linguistically [6, 11]. What explains such cross-linguistic generalizations?

A prominent hypothesis is that certain meanings are rarely or not lexicalized because lexicalizing them wouldn't improve communicative efficiency of a language, cf. [12] and much subsequent work. According to this hypothesis, languages are under two pressures: they should be simple (e.g., have simple lexicons), and they should be informative, i.e., allow for a precise communication. These two pressures are in competition. For instance, if a language

*For helpful discussions, we wish to thank Emmanuel Chemla and Roni Katzir, as well as the audiences of the CoSaQ closing workshop in Amsterdam, LINGUAE seminar at École Normale Supérieure in Paris, and Tel Aviv University Linguistics Colloquium. This research was supported by a postdoctoral fellowship awarded to MD by the Azrieli Foundation, and by the funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007- 2013) / ERC Grant Agreement n. STG 716230 CoSaQ.

[1] We consider that a meaning is lexicalized in a language if there is a simple morpheme with this meaning. Whether this morpheme can stand alone as a single word is not part of our criterion for lexicalization. For instance, English lexicalizes the number meaning 6 in a morpheme *six*, and plural meaning in a morpheme *-s*.

only has one word for colors, it will have a simple lexicon, but it won't allow for a very precise communication about colors. Adding more words for various colors would allow for a more precise communication, but at the cost of having a more complex lexicon. Natural languages have been argued to lexicalize the meanings which allow them to achieve a good compromise between these two pressures, i.e., *to optimize the simplicity/informativeness trade-off* [12, 17, 22, 21, 13, 23, 18, 19, 14, 4, 5, 24, 20].

However, in certain semantic domains, it is possible to reach very high levels of informativeness even if very few meanings from that semantic domain are lexicalized. This is due to productive morphosyntax, which may allow for construction of meanings which are not lexicalized. Consider a semantic domain of natural numbers in languages with a so-called *recursive numeral system*: these are languages in which any natural number meaning can be expressed precisely (e.g., English). Many such languages lexicalize few natural number meanings, but can precisely convey any natural number meaning using morphosyntactically complex numerals (e.g., *sixty-one* in English). In such semantic domains, lexicon size is not in direct competition with informativeness. What explains which meanings are lexicalized in these domains?

We will propose that, in semantic domains in which productive morphosyntax enables precise communication even with very few lexicalized meanings, languages are under the pressure to lexicalize as few meanings as possible (i.e, to minimize lexicon size) and the pressure to produce as morphosyntactically simple utterances as possible (i.e, to minimize average morphosyntactic complexity of utterances). Lexicon size and average morphosyntactic complexity of utterances are in competition in such domains: reducing average morphosyntactic complexity of utterances will often require lexicalizing more meanings, and reducing the size of the lexicon will often result in needing utterances of greater morphosyntactic complexity to communicate. *We thus propose that, in such domains, languages lexicalize those meanings which allow them to optimize the trade-off between lexicon size and average morphosyntactic complexity of utterances.*

We will evaluate this proposal within a case study on lexicalized number meanings in languages with a recursive numeral system, and present evidence that recursive numeral systems indeed optimize the trade-off between lexicon size and average morphosyntactic complexity of numerals. This conclusion is in tension with [21], who have instead argued that recursive numeral systems optimize the simplicity/informativeness trade-off — we briefly review their results and explain why we do not think they support the conclusion that recursive numeral systems optimize the simplicity/informativeness trade-off in Appendix B.

Once our results are in place, we will abstract away from numeral systems and discuss which notion of communicative efficiency may explain what meanings get lexicalized across semantic categories more generally, in light of this result and previous work within the simplicity/informativeness trade-off optimization approach [12].

Finally, we will turn to several other lines of work [25, 16, 8, 14, 1] which share with the present work the idea that speakers attempt to minimize complexity of their utterances, and explain in what way they are different from the proposal we are pursuing.

2 Case study on recursive numeral systems

We will present an experiment investigating how close natural languages' recursive numeral systems are to trading off optimally between lexicon size and average morphosyntactic complexity of numerals. We will evaluate this by comparing natural languages to artificially generated languages. The artificial languages represent the space of possibilities in terms of lexicon size/average morphosyntactic complexity of numerals trade-off in recursive numeral systems, and will reveal what the (approximately) optimal solutions to the trade-off problem are.

2.1 Natural languages

We collected cross-linguistic data on numerals denoting numbers 1-99 and their morphosyntactic components in recursive numeral systems from the sample of languages in the *Numeral bases* chapter in *The World Atlas of Language Structures (WALS)* [2]. Out of 172 recursive numeral systems in [2], 44 were excluded due to challenges with data collection or data interpretation. 128 languages were thus included in the analysis.[2]

These numeral systems differ in terms of which number meanings they lexicalize. Some recurring options are listed in (1). The generalization seems to be that they lexicalize the following numbers from the range 1-99: (i) the first n numbers, with n varying across languages, often the first five or the first 10, and (ii) a couple of additional numbers such as 10 and/or 20 as in (1-b) or (1-c).

(1) Lexicalized number concepts:
   a. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (75 languages)
   b. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 (19 languages)
   c. 1, 2, 3, 4, 5, 10 (9 languages)

For each of the 128 languages, for each numeral denoting a number in the range 1-99, its morphosyntactic components and their denotations were identified (cf. Table 1 for a few examples of numerals in Fulfulde). The morphosyntactic components of numerals include number-denoting morphemes, and morphemes for arithmetic operators (addition, multiplication and subtraction), cf. [9].[3] Both types of morphemes can in principle be phonetically overt or covert: in practice, number-denoting morhemes are very rarely covert, while morphemes denoting arithmetic operators often are.

Finding morphosyntactic components of numerals required studying the morphosyntactic patterns for numerals in each of the 128 languages. In most languages, the relevant morphosyntacic patterns were described in the descriptive grammars we consulted, or were otherwise easy to detect. There were however more difficult cases, even in languages we (the authors) are closely familiar with (for instance, is there a phonetic variant of the morpheme for number 2 in English *twelve*?). In such cases, we applied the following decision rule: if a numeral is an exception to an established morphosyntactic pattern in a language, but shows (phonetic or orthographic, depending on the available data) elements of morpheme(s) which should have been there if the morphosyntactic pattern was respected, we assume that the numeral follows the morphosyntactic pattern but with phonetic variants not seen elsewhere; otherwise we assume that the numeral is monomorphemic. For instance, English is a base-10 language and because of this we may expect that the numeral for 12 will be built from morphemes for 2 and 10; as the numeral *twelve* has elements of other phonetic realizations of the morpheme for the number 2 (e.g., *two*, *twe(n?)-* from *twenty*), we assume that English *twelve* incorporates the morpheme for number 2 (*twe-*?) and 10 (*-lve*?). Of course, one could argue for the application of an alternative decision rule (unless the morphosyntactic pattern is transparent, assume that the numeral is monomorphemic). It turns out that in practice applying this alternative decision wouldn't qualitatively alter the results; cf. footnote 4.

[2]Natural languages data, including source references for each language, and scripts used for the analysis are available at `https://github.com/milicaden/numerals_ac2022`.

[3]In the Danish numeral system, a morpheme denoting the fraction *half* is used to construct certain complex numerals. This suggests that there is a limited use of division too involved in the composition of numerals. As, to our knowledge, the extent to which division can be used productively in numeral systems is not well understood [9], we excluded Danish from our corpus.

Table 1: Fulfulde numerals for numbers 6, 30 and 62

Denotation (numeral)	**Morphosyntactic make-up**	**Number of morphemes**
6 (*jowe go'o*)	5 (*jowe*) + (∅) 1 (*go'o*)	3
30 (*chappande tatti*)	10 (*chappande*) · (∅) 3 (*tatti*)	3
62 (*chappande jowe go'o i didi*)	10 (*chappande*) · (∅) (5 (*jowe*) + (∅) 1 (*go'o*)) + (*i*) 2 (*didi*)	7

2.2 Lexicon size and average morphosyntactic complexity

We consider that lexicon size of a language is the number of lexicalized meanings — rather than the number of form-meaning pairs, e.g., we consider that English has one lexicon entry for 10 which can be phonetically realized as *ten*, *-teen*, *-ty*. This is important to keep in mind as it affects how our results should be interpreted: we will argue that languages optimize the trade-off between the number of lexicalized meanings and average morphosyntactic complexity of utterances, rather than the trade-off between the number of form-meaning pairs and average morphosyntactic complexity of utterances. Of course, an important question that remains open is why languages have different phonetic realizations of a single meaning.

Average morphosyntactic complexity of numerals in a language L is computed according to the formula in (2). In (2), $ms\_complexity(n, L)$ is the morphosyntactic complexity of the numeral (i.e., the number of morphemes in it) of the language L denoting the number n and $P(n)$ is the probability that the number n needs to be communicated. For instance, average morphosyntactic complexity of English is obtained as $P(1) \times$ the number of morphemes in 'one' + $P(2) \times$ the number of morphemes in 'two' +...+ $P(99) \times$ the number of morphemes in 'ninety-nine'. We assume that the probabilities that different numbers need to be communicated follow a power-law distribution as in (3) (cf. [3, 15, 21]). Qualitatively, this probability distribution captures that the larger the number n, the lower the need to talk about it.

(2)

$$average\_ms\_complexity(L) = \sum_{n \in [1,99]} P(n) \cdot ms\_complexity(n, L)$$

(3)

$$P(n) \propto n^{-2}$$

2.3 Results

In order to evaluate whether natural languages with a recursive numeral system optimize the trade-off between lexicon size and average morphosyntactic complexity of numerals, we need to establish how close they are to the optimal solutions to the lexicon size/average morphosyntactic complexity of numerals trade-off problem. The set of optimal solutions — called *the Pareto frontier* — is a set of (theoretically possible) languages for which there is no other (theoretically possible) language which is better on one of the two dimensions (lexicon size, average morphosyntactic complexity of numerals) without being worse on the other.

We estimated the Pareto frontier using an evolutionary algorithm (cf. [18, 19, 4, 5]), which involves generations of many artificial languages (i.e., artificial recursive numeral systems). We describe this procedure in Appendix A.

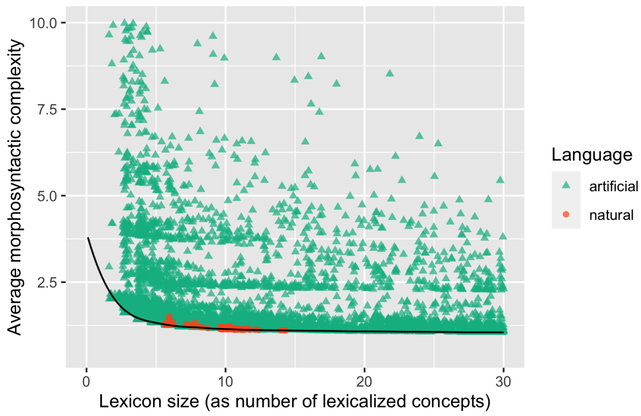

Figure 1: Experiment: Lexicon size and average morphosyntactic complexity of natural languages compared to artificial languages generated via an evolutionary algorithm. Natural languages lie at or very close to the Pareto frontier (black curve).

The natural languages and the artificial languages generated through all generations of the evolutionary algorithm are plotted in Figure 1. The estimated Pareto frontier is plotted as the black curve in Figure 1. Natural languages all lie along or very close to the Pareto frontier in Figure 1. This speaks in favor of natural languages' recursive numeral systems optimizing the trade-off between lexicon size and average morphosyntactic complexity of numerals.[4]

3 Discussion

Three pressures Previous work has provided evidence that in many semantic domains languages lexicalize those meanings which allow them to optimize the simplicity/informativeness trade-off [12, 17, 22, 21, 13, 23, 18, 19, 14, 4, 5, 24, 20].

The semantic domain of natural numbers in languages with a recursive numeral system is an example of a domain in which productive morphosyntax enables precise communication even with very few lexicalized meanings. These are domains in which simplicity of the lexicon is not in direct competition with informativeness. Our results about recursive numeral systems suggest that, in such domains, languages lexicalize those meanings which allow them to optimize the trade-off between lexicon size and average morphosyntactic complexity of utterances. Our work in combination with previous work thus evidences that there are (at least) three pressures shaping lexicons of natural languages: minimizing lexicon size, maximizing informativeness, and minimizing average morphosyntactic complexity of utterances. It is plausible that in many semantic domains the three pressures play a role. More specifically, we expect the three

[4] Recall that, when the number of morphemes in a numeral was difficult to establish, we applied the following decision rule: if a numeral is an exception to an established morphosyntactic pattern in a language, but shows (phonetic or orthographic, depending on the available data) elements of morpheme(s) which should have been there if the morphosyntactic pattern was respected, we assume that the numeral follows the morphosyntactic pattern but with phonetic variants not seen elsewhere; otherwise we assume that the numeral is monomorphemic. The alternative would have been to treat all difficult cases as monomorphemic. Importantly, however, applying the alternative decision rule (treat all difficult cases as monomorphemic) wouldn't alter the results qualitatively. With such an alternative decision rule, natural languages would have slightly larger lexicon sizes and slightly lower average morphosyntactic complexity. In other words, they would be shifted slightly down-right-wards in Figure 1. It is easy to see that they would remain close to the Pareto frontier under such a shift.

pressures to be shaping languages' lexicons in domains in which morphosyntactically complex expressions can improve how precisely meanings from that domain can be communicated, but nonetheless cannot ensure that all the meanings can be communicated precisely.

Previous work on utterance complexity Multiple studies have found that meanings that are conveyed more frequently tend to be associated to shorter forms (with length often operationalized as the number of phonemes) [25, 16, 8, 14]. This is an instance of Zipf's principle of least effort [25], according to which humans are prone to spending the least amount of effort to accomplish a task. The pressure to minimize average morphosyntactic complexity of utterances can be viewed as another instance of the same principle. The novelty of our proposal is in showing that this pressure plays a role in determining which meanings get lexicalized (and not only how long forms are associated with lexicalized meanings). According to our proposal, this pressure on its own doesn't suffice to explain which meanings are lexicalized across languages: we have shown that languages lexicalize meanings which allow them to optimize the trade-off between this pressure and the pressure to minimize lexicon size.

Another recent proposal [1] gives a central role to utterance complexity in explaining typological patterns in the semantic domain of Boolean connectives (*and, or, nor...*). [1] propose that the systems of Boolean connectives are optimal solutions to the trade-off problem between two pressures. The first pressure is the desire to minimize production effort when expressing observations (i.e., *average morphosyntactic complexity*). The second pressure, labeled *conceptual complexity* in [1], relates to how sentences with connectives are used to update the information shared by conversational participants (= *context*). [1] propose that languages are under the pressure to minimize the total conceptual complexity of contextual updates of the connectives they lexicalize. Our proposal shares with [1] the idea that languages are under the pressure to minimize average morphosyntactic complexity. Where the two proposals differ is what the competing pressure is: in our case, the pressure to minimize how many meanings are lexicalized, while in [1] the competing pressure relates to another aspect of language use, namely, how conversational participants update contextual information when they hear an expression.

4 Conclusion

In this paper, we ask what explains which meanings get lexicalized across languages. We pursue the explanation according to which languages lexicalize meanings which allow them to support efficient communication. We however argue for a refinement of what it means for a language to support efficient communication. In particular, the standard approach to communicative efficiency — the simplicity/informativeness trade-off optimization approach — cannot explain which meanings get lexicalized in semantic domains in which lexicon size and informativeness are not in direct competition. We have argued that in such domains languages optimize the lexicon size/average morphosyntactic complexity of utterances trade-off.

Our work in combination with previous work thus evidences that there are (at least) three pressures shaping lexicons of natural languages: minimizing lexicon size, maximizing informativeness, and minimizing average morphosyntactic complexity of utterances. A more general proposal for how supporting efficient communication shapes which meanings get lexicalized across languages may thus be that languages lexicalize those meanings which allow them to solve the trade-off problem between these three pressures in a (nearly) optimal way.

References

[1] Fausto Carcassi and Giorgio Sbardolini. Assertion, denial, and the evolution of boolean operators. *Mind & Language*, 2022.

[2] Bernard Comrie. Numeral bases. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013.

[3] Stanislas Dehaene and Jacques Mehler. Cross-linguistic regularities in the frequency of number words. *Cognition*, 43(1):1–29, 1992.

[4] Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. Complexity/informativeness trade-off in the domain of indefinite pronouns. In *Semantics and linguistic theory*, volume 30, pages 166–184, 2021.

[5] Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. Indefinite pronouns optimize the simplicity/informativeness trade-off. *Cognitive Science*, 2022.

[6] Peter Gärdenfors. *The Geometry of Meaning*. The MIT Press, 2014.

[7] Harald Hammarström. *Rarities in numeral systems*. De Gruyter Mouton, 2010.

[8] Martin Haspelmath. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics*, 57(3):605–633, 2021.

[9] James R Hurford. *The linguistic theory of numerals*, volume 16. Cambridge University Press, 1975.

[10] James R Hurford. A performed practice explains a linguistic universal: Counting gives the packing strategy. *Lingua*, 117(5):773–783, 2007.

[11] Gerhard Jäger. Natural Color Categories Are Convex Sets. In Maria Aloni, Harald Bastiaanse, Tikitu de Jager, and Katrin Schulz, editors, *Logic, Language, and Meaning: Amsterdam Colloquium 2009*, pages 11–20. 2010.

[12] Charles Kemp and Terry Regier. Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054, 2012.

[13] Charles Kemp, Yang Xu, and Terry Regier. Semantic typology and efficient communication. *Annual Review of Linguistics*, 4:109–128, 2018.

[14] Francis Mollica, Geoff Bacon, Noga Zaslavsky, Yang Xu, Terry Regier, and Charles Kemp. The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, 118(49), 2021.

[15] Steven T Piantadosi. A rational analysis of the approximate number system. *Psychonomic bulletin & review*, 23(3):877–886, 2016.

[16] Steven T Piantadosi, Harry Tily, and Edward Gibson. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529, 2011.

[17] Terry Regier, Charles Kemp, and Paul Kay. Word meanings across languages support efficient communication. In Brian MacWhinney and William O'Grady, editors, *The Handbook of Language Emergence*, volume 87, pages 237–263. Wiley-Blackwell, Hoboken, NJ, 2015.

[18] Shane Steinert-Threlkeld. Quantifiers in natural language optimize the simplicity/informativeness trade-off. In *Amsterdam Colloquium 2019*, 2019.

[19] Shane Steinert-Threlkeld. Quantifiers in natural language: Efficient communication and degrees of semantic universals. *Entropy*, 23(10):1335, 2021.

[20] Wataru Uegaki. The informativeness/complexity trade-off in the domain of boolean connectives. *Linguistic Inquiry*, pages 1–39, 2022.

[21] Yang Xu, Emmy Liu, and Terry Regier. Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind*, 4:57–70, 2020.

[22] Yang Xu, Terry Regier, and Barbara C. Malt. Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, 40(8):2081–2094, 2016.

[23] Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942, 2018.

[24] Noga Zaslavsky, Mora Maldonado, and Jennifer Culbertson. Let's talk (efficiently) about us: Person systems achieve near-optimal compression. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*, 2021.

[25] George Kingsley Zipf. *Human Behavior and the Principle of Least Effort.* Addison-Wesley, Cambridge, MA, 1949.

A Evolutionary algorithm

[9, 10] in his seminal work proposes that number-denoting morphemes across languages are of the syntactic category Digits (D) or Multipliers (M). Roughly, morphemes denoting numbers which are 'bases' as 10 or 20 in 'base-10' or 'base-20' languages are of the syntactic category M; other morphemes are of the syntactic category D. He proposes that, across languages, numerals are constructed according to the grammar in (4). We will use this grammar to generate numerals of artificial languages.

(4) NUMBER ⟶ D | PHRASE | PHRASE + NUMBER | PHRASE − NUMBER
PHRASE ⟶ M | NUMBER · M

For our purposes, an artificial language consists of (i) a lexicon of number-denoting morphemes, which are of category D or M (cf. the grammar in (4)), and (ii) a set of numerals for numbers 1-99 generated according to the morphosyntactic rules in (4) from the lexicon, where each numeral is the shortest expression (or one of them, in case of a tie) for a given number. For example, an artificial language L may have in its lexicon morphemes for 1, 2, 4, 6, 8, 10, such that 1, 2 and 4 are of category D and 6, 8 and 10 are of category M. The numeral for, for instance, number 16 in L would be randomly selected from one of the three morphosyntactically simplest options: $10 + 6$, $2 \cdot 8$, $8 + 8$ ($4 \cdot 4$ is not an option because it cannot be generated by the grammar in (4) (because 4 is of category D in L).

Due to computational constraints, two restrictions are imposed on the search for the shortest expression for a given number in a language. (i) Expressions of depth x (i.e., expressions with at most x number-denoting morphemes and $x - 1$ arithmetic operators) are incrementally constructed from expressions of lower depths (e.g., expressions of depth 2 are constructed by combining expressions of depth 1). However, at all depths, the meaning of expressions is restricted to be a natural number in $[1, 200]$. (ii) If no expression for a number meaning is found with a depth of at most 7, the search is abandoned, and the language is discarded.

The evolutionary algorithm works as follows. First, the generation 0 is created, which consists of 2000 artificial languages. The lexicons of these artificial languages are generated by drawing two random samples of numbers between 1-99; these stand for morphemes of category D and M respectively. As natural languages tend to have very few morphemes of category M (often only 1 or 2), we restrict the size of the random sample for the category M to at most 5 (no such restriction is imposed on the category D). The numerals for numbers 1-99 of these languages are generated as explained above. The dominant languages of a generation (those for which there is no language which is better on one dimension of lexicon size and average morphosyntactic complexity without being worse on the other) each give rise to an equal number of offspring languages, which are obtained via a small number of mutations (between 1 and 3; these mutations included removing a number morpheme from D or M, adding a number morpheme to D or M, and interchanging a number morpheme in D or M, making sure that

M is no larger than 5) from dominant languages. The dominant languages from generation 0 together with their offspring languages constitute generation 1, whose size is limited to 2000 languages. This process is repeated for 100 generations. Finally, the dominant languages are selected from the union of the last generation and the natural languages. Each of these dominant languages is a point in a two-dimensional (lexicon size and average morphosyntactic complexity of utterances) space; we do spline interpolation of these points to form the Pareto frontier.

B Simplicity/informativeness trade-off in numeral systems [21]

[21] analyze the simplicity/informativeness trade-off in 24 *restricted* and 6 *recursive* numeral systems. Restricted numeral systems have numerals for only the first few numbers, and use a quantifier such as *many* for any higher number. For instance, the language Krenak only has numerals for numbers 1-3 [7]. Recursive numeral systems are considered by [21] to be maximally informative when it comes to communicating about number meanings, while restricted numeral systems have lower degrees of informativeness. On the other hand, according to their approach to measuring complexity, the six studied recursive numeral systems are more complex than most of the 24 studied restricted numeral systems. Simplifying somewhat, this is because they assume that the complexity measure of a language should incorporate both the complexity of the lexicon and the complexity of morphosyntactic rules. As recursive numeral systems always have morphosyntactic rules for building numerals but restricted numeral systems often don't have any, recursive numeral systems tend to have a greater measure of complexity than restricted numeral systems. Recursive numeral systems are thus more complex and more informative than restricted numeral systems in [21]. [21] conclude that natural languages' numeral systems optimize the simplicity/informativeness trade-off, without making a distinction between restricted and recursive languages.

Their conclusion may seem to be in tension with our findings. However, a careful examination of the results in [21] reveals that, while restricted numeral systems are indeed close to being Pareto-optimal in trading off simplicity and informativeness, recursive numeral systems are not. Given that recursive numeral systems are maximally informative, if they were optimizing the simplicity/informativeness trade-off, we would expect to find them in the proximity of the minimally complex numeral system which is maximally informative. According to the results of [21] (cf. Figure 4b in [21]), the maximally informative system with minimal complexity has complexity ≈ 50. Strikingly, the six recursive numeral systems examined by [21] have much higher complexity than that, ranging between ≈ 75 and ≈ 150. The results of [21] thus also do not support the hypothesis that recursive numeral systems optimize the simplicity/informativeness trade-off — in other words, there is no tension between the results of [21] and our findings. Importantly, however, conclusions of [21] still hold for restricted numeral systems which are much closer to being Pareto-optimal than recursive numeral systems.