Indefinite pronouns optimize the simplicity/informativeness trade-off[*]

Milica Denić
Universiteit van Amsterdam

Shane Steinert-Threlkeld
University of Washington at Seattle

Jakub Szymanik
Universiteit van Amsterdam

Abstract

The vocabulary of human languages has been argued to support efficient communication by optimizing the trade-off between simplicity and informativeness (Kemp and Regier 2012). The argument has been based on cross-linguistic analyses of vocabulary in semantic domains of *content words* such as kinship, color, and number terms. The present work extends this analysis to a category of *function words*: indefinite pronouns (e.g. *someone*, *anyone*, *no-one*, cf. Haspelmath 2001). We build on previous work to establish the meaning space and featural make-up for indefinite pronouns, and show that indefinite pronoun systems across languages optimize the simplicity/informativeness trade-off. This demonstrates that pressures for efficient communication shape both content and function word categories, thus tying in with the conclusions of recent work on quantifiers by Steinert-Threlkeld (2019). Furthermore, we argue that the trade-off may explain some of the universal properties of indefinite pronouns, thus reducing the explanatory load for linguistic theories.

1 Introduction

The vocabulary of human languages has been argued to support efficient communication by optimizing the trade-off between simplicity and informativeness (Kemp and Regier 2012). In informal terms, the simplicity of a system is a measure of how easy it is to mentally represent that system. The informativeness of a system is a measure of how precisely the system allows us to communicate intended meanings. These two properties of a system trade-off against each other (Zipf 1949, Ferrer i Cancho and Solé 2003, Rosch 1978, Kemp and Regier 2012). The reason is that, in general, the fewer expressions a language has, the fewer semantic distinctions it is able to make, but the easier it will be to mentally represent. In other words, simplifying the language often entails sacrificing

---

[*] We wish to thank Lisa Bylinina, Fausto Carcassi, Gennaro Chierchia, Mora Maldonado, Terry Regier, Matteo Seminaroti and Benjamin Spector, as well as the audiences of the MLC seminar at the University of Amsterdam, the LINGUAE seminar at École Normale Supérieure in Paris, the CLE seminar at the University of Edinburgh, CogSci 2020 and SALT30 for helpful discussions. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007- 2013) / ERC Grant Agreement n. STG 716230 CoSaQ. An earlier version of this work was published in the proceedings of *Semantics and Linguistic Theory (SALT) 30.*

informativeness, and improving informativeness often requires added complexity. That the category systems of natural languages are (near-)optimal solutions to trading off these two measures has been argued based on cross-linguistic data and generalizations coming from various semantic domains: kinship terms, color terms, container terms, number terms, and more recently quantifiers and connectives (Kemp and Regier 2012, Regier, Kemp, and Kay 2015, Xu and Regier 2014, Xu, Regier, and Malt 2016, Steinert-Threlkeld 2019, Uegaki 2020, Xu, Liu, and Regier 2020).

The present work extends this analysis to *indefinite pronouns*, a domain of *function words* whose syntactic, semantic and typological properties have been extensively studied by comparative linguists (Haspelmath 2001). Examples of indefinite pronouns in English are expressions such as *someone, something, anyone, anything, no-one, nothing.*[1]

There are at least two reasons to pursue the extension of this framework to indefinite pronouns. First, it would strengthen the case that the simplicity/informativeness trade-off shapes both content and function word categories in language. Recent work has provided some evidence in this direction in the case of quantifiers and connectives (Steinert-Threlkeld 2019, Uegaki 2020), albeit alternative proposals have been put forward for the case of connectives (Enguehard and Spector 2021). In addition, the lack of an appropriate cross-linguistic dataset on quantifiers makes it impossible at the moment to categorically defend the claim that the category of quantifiers in natural languages is optimized for simplicity/informativeness trade-off. The case of indefinite pronoun systems helpfully differs because such a rich cross-linguistic dataset is available in this domain: Haspelmath's (2001) seminal work on indefinite pronouns includes a dataset on their meaning and distribution in 40 languages. We will rely heavily on this data in conducting the aforementioned efficiency analyses.

Second, this analysis would help make progress on two related questions: (i) what explains the variation among natural languages, and (ii) what causes *linguistic universals*, which are properties of language common to all (or nearly all) human languages. More concretely, in relation to (i), a striking finding of Haspelmath's typological research is that no two languages in his corpus have the same systems of indefinite pronouns: in other words, it is not possible to establish a one-to-one mapping between any two languages' indefinite pronouns' meaning and distribution. Strikingly, and in relation to (ii), this diversity is constrained in important ways, which Haspelmath formulates as *implicational linguistic universals*. In this work, we ask whether natural languages are different solutions to the simplicity/informativeness trade-off problem, which would explain some of the variation observed among natural languages indefinite pronouns systems, as well as whether the simplicity/informativeness trade-off is a cause of linguistic universals in the domain of indefinite pronouns.

The paper is organized as follows. First, we set the background for our analysis by (i) describing the meaning space of indefinite pronouns, (ii) detailing how Haspelmath's (2001) dataset is used in our research, and (iii) explaining how simplicity and informativeness of languages are measured. We then report the results of two computational experiments. The first experiment demonstrates that natural languages' indefinite pro-

---

[1] We note that the term *indefinite pronoun* is not a standard term for these expressions in the formal semantics literature. However, as Haspelmath's (2001) work on these expressions is well known and widely cited, we have opted for keeping the term. Roughly, in more standard terminology, the expressions categorized as indefinite pronouns by Haspelmath (2001) include "vanilla" existential indefinites, negative polarity indefinites, epistemic indefinites, free choice indefinites, and negative indefinites (including negative concord indefinites and negative quantifiers).

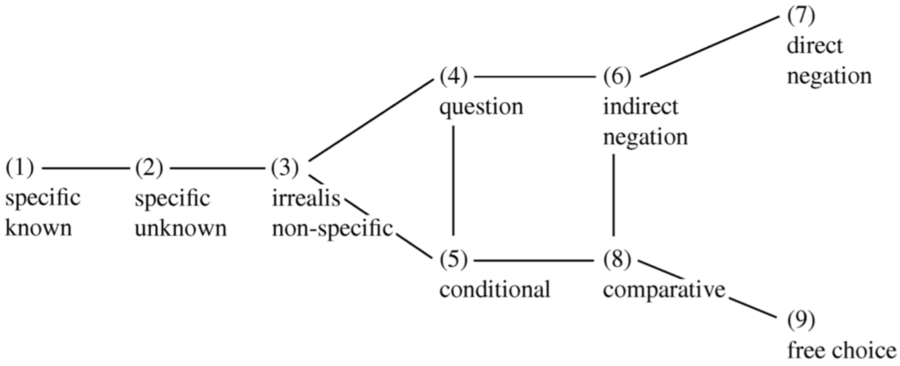


Figure 1: Haspelmath's map of functions of indefinites.

noun systems are (near-)optimal in how they trade off simplicity and informativeness. The second experiment demonstrates that Haspelmath's implicational universals play a role in the simplicity/informativeness trade-off optimization. We proceed to replicate the two experiments in three additional settings pertaining to how simplicity and informativeness are measured. Finally, we discuss the implications of the results.

2 Meaning space of indefinite pronouns

We first explain the space of possible meanings expressed by indefinites: what are the meanings that interlocutors may want to communicate by an indefinite pronoun?

Haspelmath (2001) describes each indefinite pronoun in each language in his dataset in terms of which *functions* it can take. These functions are depicted on a "map" in Figure 1. Some of them are meaning-driven (functions 1, 2, 3, 9 in Figure 1, i.e. specific known, specific unknown, non-specific, free choice), and others are driven by syntactic distribution (functions 4, 5, 6, 7, 8, i.e. question, conditional, comparative, indirect negation, direct negation in Figure 1).

Haspelmath's functions thus mix meaning and syntactic distribution, and as such cannot be (all) taken as components of the meaning space of indefinite pronouns. To focus purely on meaning, we introduce semantic flavors and translate Haspelmath's syntactic functions into them. More specifically, Haspelmath's syntactic functions relate to two well-studied categories of indefinite pronouns: *negative polarity indefinites* and *negative indefinites* (Fauconnier 1975, Ladusaw 1979, Haspelmath 2001, Penka 2011, Bernini and Ramat 1996). In the present work, we abstract away from certain differences in the syntactic distribution of different negative polarity indefinites, and likewise, of different negative indefinites. We thus assume that the meaning space that the system of indefinite pronouns has to cover consists of the following six *'semantic flavors'*, which are described informally and illustrated with an example in (1)-(6).

(1) ***Specific known flavor*** [the indefinite pronoun refers to a specific individual that the interlocutors can uniquely identify]:
Someone managed to mess this up — we all know who!

(2) ***Specific unknown flavor*** [the indefinite pronoun refers to a specific individual that the interlocutors cannot uniquely identify]:
I heard that *someone* failed, but I don't know who.

(3) ***Non-specific flavor*** [the indefinite pronoun is interpreted as an existential quantifier over some domain of possible referents, not referring to a specific individual]:
You should probably talk to *someone* else about this too.

(4) ***Negative polarity flavor*** [the indefinite pronoun is interpreted as an existential quantifier over a widened domain of possible referents]:
Less than three companies hired *anyone* this year.

(5) ***Free choice flavor*** [the indefinite pronoun is interpreted as a wide-scope universal quantifier over some domain of possible referents]:
You can hire almost *anyone* here: most of them great.

(6) ***Negative indefinite flavor*** [the indefinite pronoun is interpreted as a negated existential quantifier over some domain of possible referents]:
Who went to the party? *No one.*

We indexed each indefinite pronoun in Haspelmath's (2001) corpus with the flavors it can convey as follows.

- Functions 1, 2, 3, 9 correspond to flavors (1), (2), (3), (5) respectively.
- To decide whether an indefinite pronoun can convey flavor (6), we collected data on whether indefinite pronouns in languages of Haspelmath's corpus can be interpreted as negated existentials, relying in most cases on occurrence in negative fragment answers, which is a test for negative indefinites (Bernini and Ramat 1996).[2]
- If (a) an item can take at least one of the functions 4 and 6 but not function 3; or (b) it can take function 8 and at least one of the functions 4 and 6, or it can take function 8 but cannot take function 9; or (c) it can take function 7 but not function 3 and cannot be interpreted as a negated existential; or (d) it can take function 5, but not functions 3 and 9, then it has flavor (4).[3]

The data on indefinite pronouns presented in Haspelmath 2001 shows that languages differ greatly in how they cover the meaning space with their lexical items. Haspelmath

---

[2] These data are available in an online Appendix to the paper at https://github.com/milicaden/indefinite-pronouns-simplicity-informativeness.

[3] The motivation behind this complicated disjunctive criterion for negative polarity flavor is as follows. (a) Both indefinites with non-specific flavor and indefinites with negative polarity flavor can be used in questions and under the scope of negation, hence we can only conclude from functions 4 and 6 that an indefinite has a negative polarity flavor if we know independently that it cannot get non-specific flavor. (b) Indefinite pronouns with function 8 might have either negative polarity or free choice flavor: items with negative polarity but not free choice flavor such as English *ever* are acceptable in comparatives, but so are instances of *any* modified by *almost*, and modification by *almost* is commonly taken as evidence for the free choice interpretation of *any* (cf. Aloni and Roelofsen 2014, Heim 2006). If an item can take function 8 in combination with functions 4 and/or 6 which code for the negative polarity environments (*any* cannot be modified by *almost* in those environments which evidences that free choice flavor is not available in those environments), we may conclude that the indefinite pronoun can have the negative polarity flavor. Similarly, if an indefinite pronoun takes function 8 but not function 9, we may conclude that it cannot convey the free choice flavor and thus must be conveying the negative polarity flavor. (c) As all negative indefinites are indexed with function 7 by Haspelmath, we only rely on function 7 as revealing negative polarity flavor if the indefinite in question cannot be interpreted as negated existential (and if it cannot get non-specific flavor more generally, for the same reasons as above). (d) Finally, in the antecedent of conditionals we may find indefinites with negative polarity, non-specific, and free choice flavor; we can thus conclude from function 5 that an indefinite has a negative polarity flavor if we know independently that it cannot get non-specific or free choice flavor.

however established that this variation is constrained in the following way: any indefinite pronoun in any language can only take functions which form a connected area on the map in Figure 1. For instance, if an indefinite pronoun can take functions 2 and 4, it is also able to take function 3. It is relatively straightforward to 'translate' Haspelmath's universals from functions to flavors in the same principled way as established above. Here are two examples of such 'translations' of universals. (i) If an item can convey the specific unknown and the negative polarity flavor, it can convey the non-specific flavor. (ii) If an item can convey the specific known and the non-specific flavor, it can also convey the specific unknown flavor.

3 Measuring simplicity and informativeness

Before we introduce the relevant measures, a terminological note is in order. We will henceforth, in line with the previous work, talk about *complexity* as the opposite of simplicity, and of *communicative cost* as the opposite of informativeness. One may thus equivalently speak of languages striving to maximize both simplicity and informativeness, or striving to minimize complexity and communicative cost, that is, of simplicity/informativeness trade-off or complexity/communicative cost trade-off. For present purposes, we will define measures of complexity, informativeness and communicative cost, with communicative cost being a decreasing function of informativeness, and analyze the trade-off between complexity and communicative cost.

3.1 Complexity

Our measure of complexity relies on featural make-up of indefinite pronouns. For this measure, we will again build on Haspelmath's work. Haspelmath (2001) (Chapter 5) proposes that there are 5 binary features indefinite items can carry: known to the speaker (K), specific (S), scalar endpoint (SE), scale reversal (R), and in the scope of negation (N). Haspelmath further assumes that the feature R ($+$ or $-$) requires the indefinite pronoun to carry $SE+$ feature.

Let us review briefly what these features stand for in Haspelmath's work. The features K and S are relatively transparent. S relates to the semantic notion of specificity, i.e. to whether the speaker has a specific referent in mind for the indefinite pronoun. K relates to whether or not the referent is known to the speaker. As for SE, Haspelmath (2001) motivates it from Fauconnier's work on negative polarity and free choice indefinites: in short, negative polarity and free choice indefinites evoke a pragmatic scale of alternatives ordered by likelihood, and they associate with its lowest (least likely) endpoint (Fauconnier 1975). Scale reversing contexts reverse the order of alternatives on the pragmatic scale; these are essentially downward-entailing contexts. The R feature reflects that negative polarity indefinites should associate with the lowest point on the scale in the scale reversing contexts ($R+$), while free choice indefinites associate with the lowest point on the scale in the non-scale reversing contexts ($R-$). Finally, the N feature relates to whether the indefinite pronoun necessarily appears in the scope of negation.

Haspelmath assumes that each of these five binary features characterizes a subset of functions of indefinite pronouns (cf. Haspelmath (2001) for details). In the continuation, we assume the five binary features to characterize sets of semantic flavors as follows:[4]

---

[4] The features $SE+$ and $R+$ are considered to characterize the negative indefinite flavor for two reasons.

- $K+$: specific known
  $K-$: specific unknown, non-specific, negative polarity, free choice, negative indefinite
- $S+$: specific known, specific unknown
  $S-$: non-specific, negative polarity, free choice, negative indefinite
- $SE+$: negative polarity, free choice, negative indefinite
  $SE-$: specific known, specific unknown, non-specific
- $R+$: negative polarity, negative indefinite
  $R-$: free choice
- $N+$: negative indefinite
  $N-$: specific known, specific unknown, non-specific, negative polarity, free choice

Haspelmath does not provide a general recipe for how to generate the featural make-up of an indefinite from the combination of functions that it might be able to take. We provide here one such recipe which is simple yet general enough to enable us to define any item in terms of its feature content based on the semantic flavors it can convey. Let us treat features as sets of flavors that they characterize, and the combination of features to correspond to set-theoretic operations of intersection or union. We then define the featural make-up of the indefinite pronoun to be *the shortest formula*—in a language whose primitives are the five binary features listed above and the set-theoretic operations of union and intersection—that would correspond exactly to the flavor(s) an item can convey. For instance, in this language, the two formulae '$K+$' and '$K + \cap S+$' amount to the same set of flavors, i.e. {specific known}. However, because '$K+$' is a shorter formula than '$K + \cap S+$', we consider the featural make-up of an indefinite pronoun that conveys only the specific known flavor to be '$K+$' and not '$K + \cap S+$'.[5]

We measure the *complexity of an item c(i)* as the number of features in its featural make-up (repetitions of the same features are counted as well). For instance, the featural make-up of an item which can convey only the specific unknown flavor would be '$S + \cap K-$', and its complexity would thus be 2. This measure relates to the length of the formula in terms of the number of primitives: as we are assuming that the features can be combined by binary set-theoretic operations, if there are n features in a formula, there will be $n - 1$ set-theoretic operators.

Our measure of *complexity of a language Comp(L)* is defined as the sum of complexity measures of each item in the language (cf. (7)).

First, Haspelmath considers these two features to characterize the direct negation function, and as discussed, all indefinite pronouns with the negative indefinite flavor are indexed by Haspelmath as having the direct negation function. Second, these two features characterize the negative polarity flavor, and at least certain negative indefinites (the so-called negative concord items or N-words) have been argued to simply be a special type of negative polarity indefinites (i.e. negative polarity items which can induce the presence of negation); see for instance Chierchia 2013.

[5] It is worthwhile pointing out that the features Haspelmath proposes for indefinite pronouns are clearly applicable to items from various other domains. For instance, the concept of scale has been argued to play a role in the semantics of scalar particle *even* (Karttunen and Peters 1979), and linguists have in fact drawn connections between scalar particles and negative polarity indefinites (e.g. Lee and Horn 1994).

(7) **Complexity of a language:**

$$Comp(L) = \sum_{i \in L} c(i)$$

3.2 Informativeness and communicative cost

Our measure of informativeness is rooted in the notion of successful communication. The communication unfolds as follows. A speaker has a semantic flavor f from the set of flavors F in mind that they want to communicate to a listener by using an indefinite pronoun i from their language L; the listener tries to guess upon hearing i which flavor the speaker intended (Skyrms 2010, Kemp, Xu, and Regier 2018, Steinert-Threlkeld 2019). We assume a simplified communicative scenario between a literal speaker and a literal listener who do not engage in pragmatic reasoning about each other's communicative intentions: these agents' production and interpretation of indefinite pronouns follows solely from their semantics.[6] The literal speaker's probability to use i to communicate f, $P_{S-lit}(i|f)$ is defined in (8). The literal listener's probability to guess f upon hearing i, $P_{L-lit}(f|i)$ is defined in (9). $[\![\cdot]\!]$ is the interpretation function; $[\![i]\!](f) = 1$ iff $f \in i$ (iff f is a possible interpretation of i). The literal listener is closely related to L_0 communicative agent from the basic rational speech act (RSA) model Frank and Goodman (2012).[7]

(8) **Literal speaker:**

$$P_{S-lit}(i|f) = \frac{[\![i]\!](f)}{\sum_{i' \in L} [\![i']\!](f)}$$

(9) **Literal listener:**

$$P_{L-lit}(f|i) = \frac{[\![i]\!](f)}{\sum_{f' \in F} [\![i]\!](f')}$$

The informativeness of a language is defined in (10). It corresponds to the probability that the communication will be successful given the need to communicate different flavors, reflected in the prior over flavors from the set of flavors F, as well's as the speaker's and listener's probability.

(10) **Informativeness of a language:**

$$I(L) = \sum_{f \in F} \sum_{i \in L} P(f) P_{L-lit}(f|i) P_{S-lit}(i|f)$$

The prior over flavors is estimated from the corpus in Beekhuizen, Watson, and Stevenson 2017 in which indefinite pronouns are annotated for functions. Going from functions to semantic flavors,[8] estimated priors over flavors are provided in Table 1.

---

[6] We consider pragmatic speakers and listeners later in the paper.

[7] L_0 from the basic RSA model as in Frank and Goodman (2012) is more pragmatically sophisticated than the literal listener defined in (9) in that L_0 integrates the prior over flavors when interpreting indefinite pronouns.

[8] This was done using the English version of the corpus in Beekhuizen et al. 2017 as follows. If an indefinite pronoun was *anyone* or *anybody* and it was annotated to have one of the functions *question, indirect negation, direct negation*, we indexed it as having the *negative polarity flavor*. If an indefinite pronoun was *anyone* or *anybody* and it was annotated to have one of the functions *conditional, comparative*, we indexed it as having the *negative polarity flavor* half of the time, and as having the *free choice flavor* half of the time (this is because these environments allow for both *negative polarity* and *free choice*

Semantic flavor	Prior probability
specific known	0.08
specific unknown	0.08
non-specific	0.26
negative polarity	0.33
free choice	0.1
negative indefinite	0.15

Table 1: Prior probability distribution over flavors, as estimated from the corpus in Beekhuizen et al. 2017.

The *communicative cost* of a language should be a decreasing function of the informativeness of the language; we define the communicative cost of a language L in (11). This means that maximizing informativeness of a language is equivalent to minimizing communicative cost.

(11) **Communicative cost of a language:**

$$Cost(L) = \frac{1}{I(L)}$$

4 Experiment 1

In Experiment 1,[9] we address the question of whether natural languages optimize the simplicity/informativeness trade-off in the domain of indefinite pronouns.

For our purposes, a language is a set of indefinite pronouns for the ontological category PERSON (e.g. *someone, anyone, no one*).[10] We evaluate the optimality of natural languages by measuring how distant they are from the Pareto frontier in comparison to artificially generated languages. The Pareto frontier consists of Pareto optimal languages; a language is (Pareto) optimal if there is no other language that has both lower complexity and lower communicative cost. The idea is that, if natural languages optimize simplicity/informativeness trade-off, they should be more optimal than artificially generated languages, and being more optimal means being closer to the set of optimal languages (the Pareto frontier).

---

flavor; we have no data on the rate of each of these flavors in these two environments and thus assume them to be equally frequent). If an indefinite pronoun was *someone, somebody* and it was annotated to have one of the functions *question, indirect negation, direct negation, conditional, comparative*, we indexed it as having the *non-specific flavor*. If an indefinite pronoun was *no one, nobody*, we indexed it as having the *negative indefinite flavor*. Indefinite pronouns annotated with functions *non-specific* and *free choice* were indexed as having corresponding semantic flavors. Beekhuizen et al. (2017) did not distinguish between *specific known* and *specific unknown* functions, and annotated the indefinite pronouns which had one of those two functions with *specific* only. For lack of evidence to the contrary, we assume that among the indefinite pronouns annotated with *specific* by Beekhuizen et al. (2017), those with *specific known* and those with *specific unknown* flavor are equally frequent.

[9] The scripts and data files used for the experiments reported in this paper can be found at https://github.com/milicaden/indefinite-pronouns-simplicity-informativeness.

[10] Comparable experiments may be conducted for other ontological categories, such as THING, TIME, PLACE, etc. What may change in these additional experiments are the prior probability distributions over flavors and the inventory of indefinite pronouns for natural languages, as natural languages may have gaps in their indefinite pronoun paradigms for certain ontological categories.

To evaluate this, we artificially generated 10000 languages, which could have between 1 and 7 indefinite pronouns (7 is the maximum number of indefinite pronouns that any natural language has in Haspelmath's corpus). Each indefinite pronoun in each artificial language was randomly assigned one of the 63 logically possible combinations of flavors ($2^6 - 1$ combinations, excluding an indefinite pronoun that doesn't convey any of the 6 flavors).

Random assignment of the 63 logically possible combinations of flavors to items of artificial languages resulted in artificial languages having more items with overlapping meanings and more semantic gaps than natural languages. These two properties put artificial languages at a disadvantage compared to natural languages with respect to simplicity/informativeness trade-off for an uninteresting reason. Because of this, before we compare natural and artificial languages with respect to simplicity/informativeness trade-off, we control for this difference as follows.

The artificial languages were matched to natural languages for *the degree of overlap*, defined as in (12). The degree of overlap captures how many different indefinite pronouns in a language can be used to express a flavor: if the indefinite pronouns in the language have more overlapping meanings, the degree will be higher. Matching ensured that the average degree of overlap among natural languages (call it $\overline{nat\_overlap}$) is similar to the average degree of overlap among artificial languages.

(12) **Degree of overlap:**

$$Syn(L) = \sum_{f \in F} |\{i \in L : f \in i\}| - 1$$

Furthermore, the artificial languages were matched to natural languages for *the degree of coverage*, defined as in (13). The degree of coverage measures how many flavors can be expressed by indefinite pronouns in a language. As before, matching ensured that the average degree of coverage among natural languages (call it $\overline{nat\_coverage}$) is similar to the average degree of coverage among artificial languages.

(13) **Degree of coverage:**

$$Cov(L) = |\{f \in F : \exists i \in L.\ f \in i\}|$$

The matching procedure was solving the optimization problem in (14), which is an integer programming problem (Schrijver 1998). The objective is to find a maximal subset of the 10000 artificially generated languages under the following constraints: (i) their average degree of overlap is within $\overline{nat\_overlap} \cdot (1 + / - \epsilon)$; and (ii) their average degree of coverage is within $\overline{nat\_coverage} \cdot (1 + / - \epsilon)$, with $\epsilon = 0.01$. The procedure was implemented in R (R Core Team 2020) using *ompr* package for modeling mixed integer linear programs (Schumacher 2017) and *glpk* solver (Makhorin 2008). Due to computational constraints, the time to run the algorithm was limited to 24h, after which the best solution at that time was collected (note that this means that the collected solution is not necessarily the optimal solution, i.e. there may exist a larger subset of artificial languages which falls within constraints (i) and (ii), but which was not found within the 24h time limit).

(14)

$$\text{maximize} \sum_{k=1}^{10000} x_k$$

subject to

$$\sum_{k=1}^{10000} x_k \cdot \text{overlap}_k \leqslant \overline{\textit{nat\_overlap}} \cdot (1+\epsilon) \cdot \sum_{k=1}^{10000} x_k$$

$$\sum_{k=1}^{10000} x_k \cdot \text{overlap}_k \geqslant \overline{\textit{nat\_overlap}} \cdot (1-\epsilon) \cdot \sum_{k=1}^{10000} x_k$$

$$\sum_{k=1}^{10000} x_k \cdot \text{coverage}_k \leqslant \overline{\textit{nat\_coverage}} \cdot (1+\epsilon) \cdot \sum_{k=1}^{10000} x_k$$

$$\sum_{k=1}^{10000} x_k \cdot \text{coverage}_k \geqslant \overline{\textit{nat\_coverage}} \cdot (1-\epsilon) \cdot \sum_{k=1}^{10000} x_k$$

where

$$\forall k \in \{1, \ldots, 10000\},\ x_k = \begin{cases} 1 \text{ if the artificial language } k \text{ is in the matched set} \\ 0 \text{ otherwise} \end{cases}$$

and

$$\text{overlap}_k \text{ is the degree of overlap of the artificial language } k,$$

$$\text{coverage}_k \text{ is the degree of coverage of the artificial language } k,$$

$$\overline{\textit{nat\_overlap}} \text{ is the average degree of overlap among natural languages,}$$

$$\overline{\textit{nat\_coverage}} \text{ is the average degree of coverage among natural languages,}$$

$$\epsilon = 0.01$$

After matching for overlap and coverage, 479 artificial languages remained for comparison to natural languages. After matching, mean degree of overlap of both natural and artificial is 0.68 (that of artificial was 7.1 pre-matching), and mean degree of coverage of natural languages is 5.58, and that of artificial languages is 5.52 (that of artificial was 5.2 pre-matching).

Controlling for overlap makes sure that, if natural languages are superior to artificial languages in terms of the simplicity/informativeness trade-off, this is not simply due to artificial languages having more words whose meanings overlap. In addition to this, controlling for coverage makes sure that, if natural languages are superior to artificial languages in terms of the simplicity/informativeness trade-off, this is not simply due to artificial languages having more semantic gaps.

The artificially generated languages serve to map the space of possibilities for indefinite pronoun systems whose degrees of overlap and coverage are comparable to those of natural languages. We compute complexity and communicative cost of natural and (matched) artificial languages of Experiment 1, and plot them in Figure 2.

We follow Steinert-Threlkeld (2019) in using an evolutionary algorithm to estimate the Pareto frontier. The algorithm works as follows. First, the generation 0 is generated, which consists of 2000 randomly generated languages. The dominant languages (those for which there is no language which has both lower complexity and lower communicative cost) 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 an item,

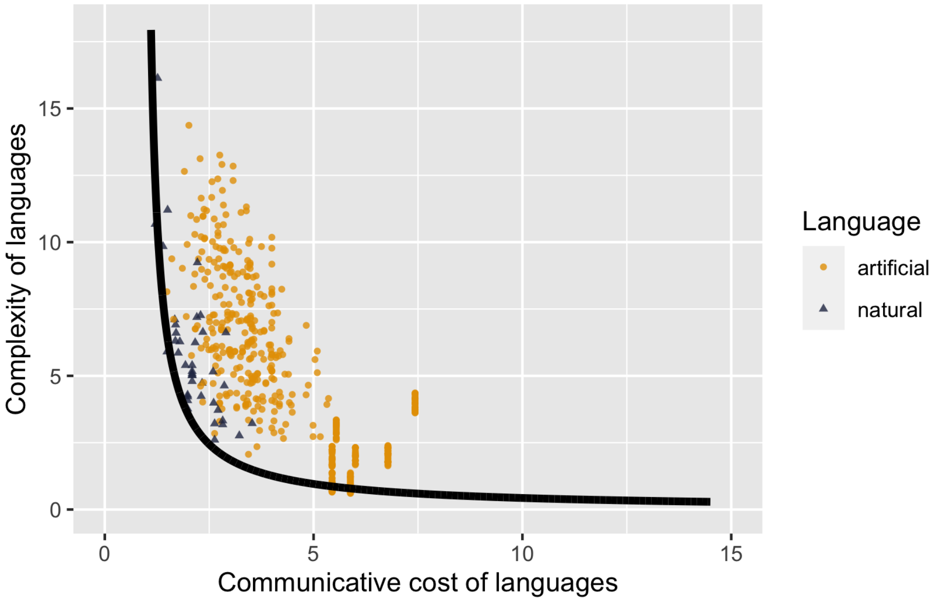


Figure 2: Experiment 1: Complexity and communicative cost of 40 natural and 479 artificial languages (natural and artificial languages matched for the degree of overlap and coverage).

adding an item, and interchanging an item) from dominant languages. The dominant languages from generation 0 together with their offspring languages constitute generation 1. This process is repeated for 100 generations. Finally, the dominant languages are selected from the union of the last generation, the 40 natural languages, and the matched artificial languages from Experiment 1. A curve in (15) is then fitted to the dominant languages to estimate the Pareto frontier. The estimated Pareto frontier is plotted as the black curve in Figure 2.

(15)

$$Comp(L) \sim \frac{a}{b + Cost(L)}$$

As can be seen in Figure 2, most of the 40 natural languages lie near this Pareto frontier, suggesting that they are optimizing the simplicity/informativeness trade-off. This is further supported by the comparison of distances from the Pareto frontier of natural and artificial languages. For each language, we compute its distance from the Pareto frontier as the minimum Euclidean distance between the language and a point on the Pareto frontier. We find that natural languages are significantly closer to the Pareto frontier than artificial languages ($M_1 = 0.48$, $M_2 = 1.55$, $t(82.8) = -16$, $p < .001$). Interestingly, focusing on where natural languages lie with respect to the Pareto frontier in Figure 2, we see that while most natural languages lie in the lower left corner of the plot, with low communicative cost and low complexity, there is quite some variability in terms of what the closest point on the Pareto frontier for each of the natural languages would be. This suggests that some of the diversity that can be observed in the indefinite pronoun systems across languages is due to languages approaching different optimal solutions to the trade-off problem (cf. also Kemp et al. (2018)).

5 Experiment 2

Experiment 1 demonstrates that natural languages optimize the simplicity/informativeness trade-off of their indefinite pronoun systems. The trade-off optimization is thus likely to explain some of the variation between indefinite pronoun systems among natural languages, as well as some of their universal properties. Does the trade-off explain some of Haspelmath's universals? To answer this question, in Experiment 2, we compare 5000 artificial languages which satisfy Haspelmath's universals, as originally stated, in terms of functions rather than flavors,[11] to 5000 artificial languages which do not (henceforth *Haspel-ok* and *Not Haspel-ok* languages respectively). We do this comparison for the following reason: if the reason why natural languages satisfy Haspelmath's universals is because these help optimize simplicity/informativeness trade-off, then artificial languages which satisfy Haspelmath's universals should be more optimal than artificial languages which do not satisfy them. Languages of both groups had between 1 and 7 items, and *Not Haspel-ok* languages were matched to *Haspel-ok* languages for their degree of overlap and coverage. After matching, 2881 *Not Haspel-ok* languages remain, with mean degree of overlap of *Haspel-ok* languages 5.66 and that of *Not Haspel-ok* 5.61 (that of *Not Haspel-ok* was 7.4 pre-matching), and with mean degree of coverage of *Haspel-ok* languages 4.58, and that of *Not Haspel-ok* 4.63 (that of *Not Haspel-ok* was 5.2 pre-matching).[12] Each item in each *Haspel-ok* language is sampled from a pool of all logically possible items which satisfy Haspelmath's universals (in terms of which of Haspelmath's functions they can take). On the other hand, each item in each *Not Haspel-ok* language is sampled from a pool of all logically possible items (in terms of which of Haspelmath's functions they can take): items of *Not Haspel-ok* languages thus may or may not conform to Haspelmath's universals. The Pareto frontier is estimated in the same way as in Experiment 1.

In Figure 3, we plot the complexity and communicative cost of *Haspel-ok* and of (matched) *Not Haspel-ok* languages, as well as the estimated Pareto frontier. Using the same measure of distance from Pareto frontier as in Experiment 1, we find that the *Haspel-ok* languages are significantly closer to Pareto frontier than the *Not Haspel-ok* languages ($M_1 = 1.61$, $M_2 = 2.26$, $t(4819.4) = -34.3$, $p < .001$). This demonstrates that languages which satisfy Haspelmath's universals are indeed better at trading simplicity and informativeness than languages which do not. Importantly, the results of Experiment 2 demonstrate that this holds in general, and not only for the 40 natural languages from Haspelmath's corpus.

---

[11] Recall that all items which convey the negative indefinite flavor, such as English *no one*, are indexed as having *direct negation function* by Haspelmath (but they are not the only items to be indexed as having *direct negation function*). As we relied on acceptability in negative fragment answers to be the distinguishing test for the negative indefinite flavor, we consider the *negative fragment answer* to be the 10th function, and add an extra universal to Haspelmath's list: if an indefinite pronoun can have the *negative fragment answer* function, it can also have the *direct negation function*.

[12] The direction of the matching is not expected to affect our conclusions; indeed, the results of Experiment 2 remain qualitatively the same if *Haspel-ok* languages are matched to *Not Haspel-ok* languages.

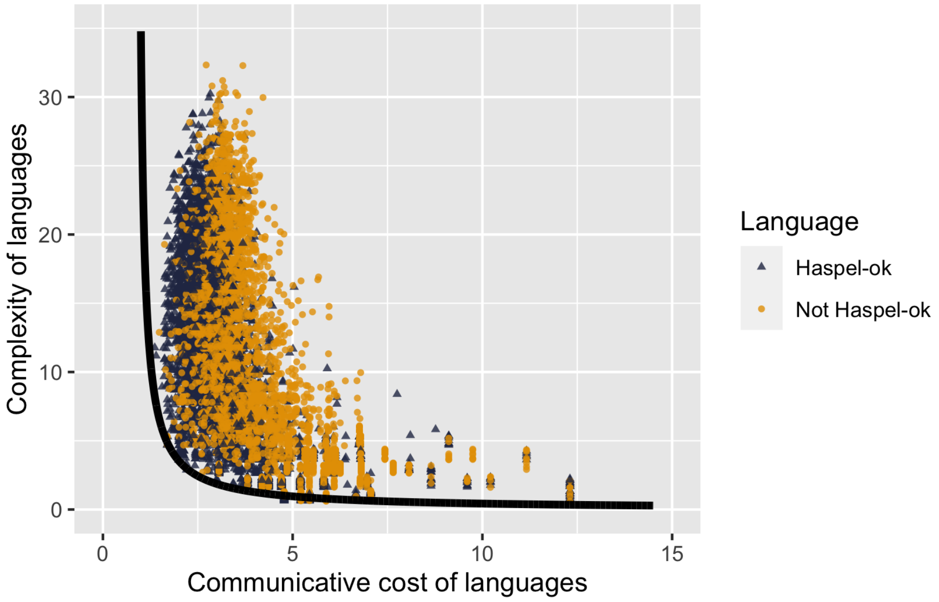


Figure 3: Experiment 2: Complexity and communicative cost of 5000 artificial languages which satisfy Haspelmath's universals and 2881 artificial languages which do not (*Haspel-ok* and *Not Haspel-ok* languages matched for the degree of overlap and coverage).

6 Alternative measures of simplicity and informativeness

The simplicity/informativeness trade-off analyses depend heavily on the way simplicity and informativeness are measured (equivalently, complexity and communicative cost). We will now consider one alternative way of measuring complexity, and one alternative way of measuring informativeness and communicative cost, motivated below.

6.1 Alternative measure of complexity: Adding the set difference

5 binary features, as well as flavors they are compatible with, were introduced in Section 3.1. The featural make-up of an indefinite pronoun was defined to be the shortest formula in a language whose primitives are the five binary features and the set-theoretic operations of union and intersection.

For completeness, we also consider a language which, in addition to these primitives, also includes the operation of set difference. In other words, we will consider an alternative definition of the featural make-up of an indefinite pronoun as the shortest formula in a language whose primitives are the five binary features and the set-theoretic operations of union, intersection and set difference. Complexity of an item and complexity of a language are defined as in Section 3.1, modulo the new approach to the featural make-up of the indefinite pronouns.

6.2 Alternative measure of informativeness and communicative cost: Pragmatic speakers and listeners

Our original measure of informativeness was rooted in the notion of successful communication between a literal speaker and a literal listener, who do not reason about each other's communicative intentions in producing or interpreting indefinite pronouns. It is however plausible that languages are optimizing for successful communication between more sophisticated communicative agents: henceforth, a pragmatic speaker and a pragmatic listener. The pragmatic speaker's probability to use i to communicate f, $P_{S-prag}(i|f)$ is defined in (16). What (16) models is the following: the pragmatic speaker reasons about how the literal listener will interpret their utterances, and chooses their utterance based on how well the utterance communicates the intended meaning to the literal listener. The pragmatic listener's probability to guess f upon hearing i, $P_{L-prag}(f|i)$ is defined in ((17)). What (17) models is the following: the pragmatic listener interprets their utterance based on the probability that the pragmatic speaker would use that utterance to communicate various meanings, together with the prior probability that the meanings in question need to be communicated. The pragmatic speaker and pragmatic listener defined in (16) and (17) respectively are closely related to S_1 and L_1 communicative agents from the basic RSA model (Frank and Goodman 2012).[13] The temperature parameter α in the definition of the pragmatic speaker in (16), which controls how rational the speaker's decision-making is, is assumed to be 1: this is a typical choice when no strong assumptions about the speaker's rationality are being made.

(16) **Pragmatic speaker:**

$$P_{S-prag}(i|f) = \frac{exp(\alpha \cdot log\ P_{L-lit}(f|i))}{\sum_{i' \in L} exp(\alpha \cdot log\ P_{L-lit}(f|i'))}$$

(17) **Pragmatic listener:**

$$P_{L-prag}(i|f) = \frac{P_{S-prag}(f|i) \cdot P(f)}{\sum_{f' \in F} P_{S-prag}(f|i) \cdot P(f)}$$

With this, the revised definition of informativeness is as follows:

(18) **Informativeness of a language (pragmatic version):**

$$I(L) = \sum_{f \in F} \sum_{i \in L} P(f) P_{L-prag}(f|i) P_{S-prag}(i|f)$$

The communicative cost of a language is defined as in Section 3.2, modulo the revised definition of informativeness.

6.3 Results

Experiments 1 and 2 were repeated with these new measures of complexity and communicative cost. We report here the results of these replications with the following settings: (i) the original measure of communicative cost (which we will refer to as *literal cost*) in combination with the new measure of complexity (which we will refer to as *3-operators*

---

[13] The only difference between the S_1 and L_1 from the basic RSA model and the pragmatic speaker and listener as defined in (16) and (17) is that the pragmatic speaker and listener do not reason about how complex various indefinite pronouns are: in other words, the communicative scenario being modelled is that of interlocutors whose language production and comprehension is not influenced by the complexity of different expressions. This allows to evaluate informativeness of the language independently of complexity.

complexity); (ii) the new measure of communicative cost (which we will refer to as *pragmatic cost*) in combination with the original measure of complexity (which we will refer to as *2-operators complexity*); and (iii) pragmatic cost in combination with 3-operators complexity.

The results replicate entirely in these three alternative settings.

Experiment 1-i and Experiment 2-i: Literal cost and 3-operators complexity The procedure for conducting Experiments 1-i and 2-i was identical to those for Experiments 1 and 2, respectively, modulo the following: while Experiments 1 and 2 employed 2-operators complexity, Experiments 1-i and 2-i employed 3-operators complexity.

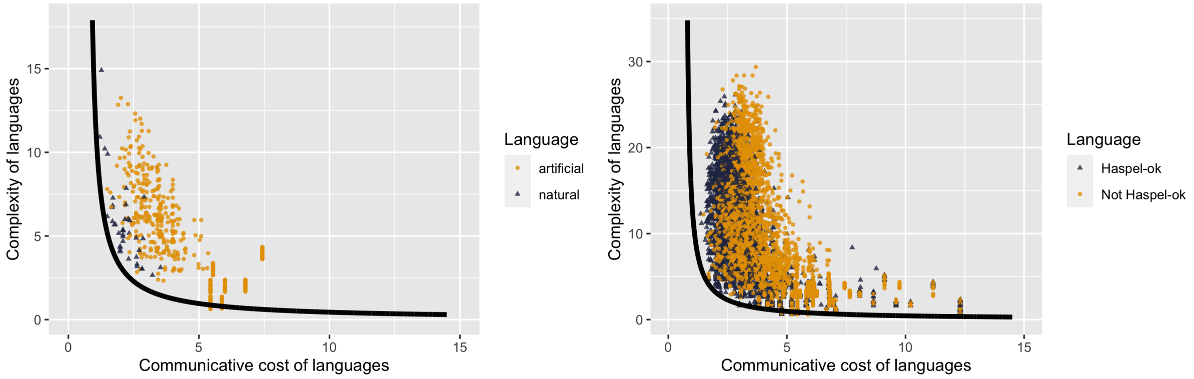
In Figures 4a and 4b, we plot the results of Experiments 1-i and 2-i respectively. In Experiment 1-i we find that the *natural* languages are significantly closer to Pareto frontier than the *artificial* languages ($M_1 = 0.6$, $M_2 = 1.6$, $t(85.1) = -15.1$, $p < .001$). In Experiment 2-i, we find that the *Haspel-ok* languages are significantly closer to Pareto frontier than the *Not Haspel-ok* languages ($M_1 = 1.7$, $M_2 = 2.3$, $t(4820.2) = -33.9$, $p < .001$).

Experiment 1-ii and Experiment 2-ii: Pragmatic cost and 2-operators complexity The procedure for conducting Experiments 1-ii and 2-ii was identical to those for Experiments 1 and 2, respectively, modulo the following: while Experiments 1 and 2 employed literal cost, Experiments 1-ii and 2-ii employed pragmatic cost.

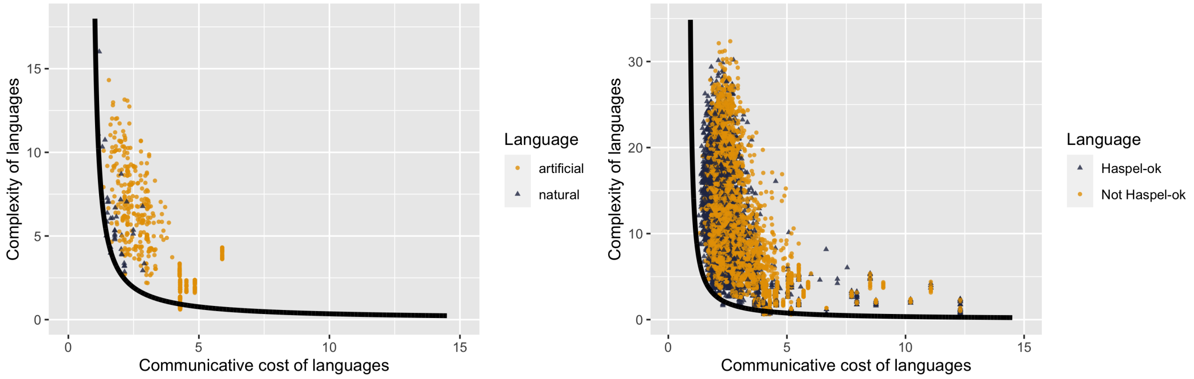
In Figures 4c and 4d, we plot the results of Experiments 1-ii and 2-ii respectively. In Experiment 1-ii we find that the *natural* languages are significantly closer to Pareto frontier than the *artificial* languages ($M_1 = 0.4$, $M_2 = 1.2$, $t(73.9) = -12$, $p < .001$). In Experiment 2-ii, we find that the *Haspel-ok* languages are significantly closer to Pareto frontier than the *Not Haspel-ok* languages ($M_1 = 1.2$, $M_2 = 1.6$, $t(4797.4) = -24.6$, $p < .001$).

Experiment 1-iii and Experiment 2-iii: Pragmatic cost and 3-operators complexity The procedure for conducting Experiments 1-iii and 2-iii was identical to those for Experiments 1 and 2, respectively, modulo the following: while Experiments 1 and 2 employed literal cost, Experiments 1-iii and 2-iii employed pragmatic cost; while Experiments 1 and 2 employed 2-operators complexity, Experiments 1-iii and 2-iii employed 3-operators complexity.

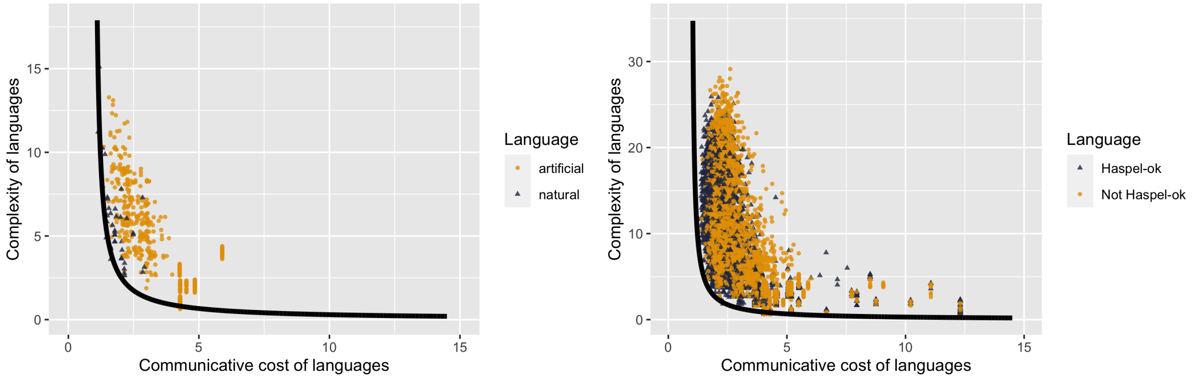
In Figures 4e and 4f, we plot the results of Experiments 1-iii and 2-iii respectively. In Experiment 1-iii we find that the *natural* languages are significantly closer to Pareto frontier than the *artificial* languages ($M_1 = 0.4$, $M_2 = 1.2$, $t(74.5) = -12.5$, $p < .001$). In Experiment 2-iii, we find that the *Haspel-ok* languages are significantly closer to Pareto frontier than the *Not Haspel-ok* languages ($M_1 = 1.1$, $M_2 = 1.5$, $t(4731.3) = -25.8$, $p < .001$).



(a) Experiment 1-i (b) Experiment 2-i



(c) Experiment 1-ii (d) Experiment 2-ii



(e) Experiment 1-iii (f) Experiment 2-iii

Figure 4: *Row 1*: Replications of Experiment 1 and 2 with settings: literal cost, 3-operators complexity. *Row 2*: Replications of Experiment 1 and 2 with settings: pragmatic cost, 2-operators complexity. *Row 3*: Replications of Experiment 1 and 2 with settings: pragmatic cost, 3-operators complexity

7 Discussion

In this work, we have argued that the simplicity/informativeness trade-off can explain how a language organizes its vocabulary to cover the meaning space expressed by indefinite pronouns. In general, such work relies on three fundamental assumptions: (1) that we are indeed dealing with a proper category (i.e. that all and only the expressions we are considering form a (sub-)system of a language); (2) that we have a good understanding of the meaning space that the system is covering; (3) that we have a reasonable way to estimate the simplicity and informativeness of the system. These assumptions are sometimes left implicit in related work; we discuss them explicitly in relation to indefinite pronouns, and point out where they may be questioned and may evolve. We believe that the points we raise in this discussion may be of relevance for any work belonging to this paradigm.

7.1 Are indefinite pronouns a well-defined category?

The question of whether the system of indefinite pronouns is optimized with respect to the simplicity/informativeness trade-off presupposes that the expressions that fall under the label ‘indefinite pronoun’ indeed form a category across languages. One way to establish that something is a category is to provide a set of criteria that would separate members from non-members. Haspelmath (2001) defines indefinite pronouns as expressions (i) which are *grammatical*, i.e. function words, that are syntactically mainly noun phrases, adverbial phrases, or adjectival phrases, and (ii) whose main semantic function is to express *indefinite reference*. Haspelmath however acknowledges that (ii) is too narrow to qualify as indefinite pronouns all and only the expressions that he lists under the term: for instance, the indefinite pronoun *nobody* arguably does not express indefinite reference, but rather conveys the non-existence of a referent. A more general semantic criterion is thus needed for the category of indefinite pronouns to be well-defined.

While it is beyond the scope of this work to attempt to identify such a more general semantic criterion, let us point out that much semantic work since Haspelmath has indeed argued in favor of close semantic connections between expressions that Haspelmath considers to be ‘indefinite pronouns’. For instance, Chierchia (2013) argues that expressions such as *someone, something* and *anyone, anything* and their cross-linguistic counterparts have a common semantic core; much work assumes that expressions such as *no-one, nothing* are underlyingly negation merged with an indefinite expression such as *someone, something* (cf. Jacobs 1980 and much subsequent work). Furthermore, the expressions that Haspelmath subsumes under the term ‘indefinite pronouns’ are often diachronically related across languages, suggesting that their meanings are closely related (Chierchia 2013, Roberts and Roussou 2003, Jäger 2010). While we can thus have some confidence in the reality of the category of indefinite pronouns, the task of justifying the categories posited must be addressed in any work of this kind. This task becomes more pressing in domains of functional vocabulary. In content word domains (e.g. color terms), one can use criteria about the kinds of entities referred to by the expression (e.g. colors) to demarcate the sub-system; but functional domains have more abstract meaning spaces, which makes these criteria harder to apply in practice.

7.2 Meaning space of indefinite pronouns

Investigating how efficiently systems of indefinite pronouns cover a certain meaning space presupposes that we have a good understanding of what the meaning space consists of.

We have described in Section 2 the six semantic flavors assumed to constitute the meaning space of indefinite pronouns. As indicated there, however, in doing so we have abstracted away from a number of subtle differences in syntactic distribution—especially in the domain of items with negative polarity and negative indefinite flavor—which may very well reflect subtle meaning differences. For instance, among negative polarity indefinites, there is a bewildering diversity in syntactic and semantic behavior of various expressions, as evidenced by the literature on strong, weak, and Bagel problem negative polarity items (Krifka (1994), Pereltsvaig (2004), a.o.). The situation is similar in the domain of free choice and negative indefinites (Chierchia 2013, Zeijlstra 2004). It is possible that future work on semantic correlates of these various types of negative polarity, free choice and negative indefinites will lead to a more fine-grained meaning space for indefinite pronouns.

7.3 Measures of complexity and communicative cost of languages

The results reported in this paper are robust across four settings: {literal cost, pragmatic cost} × {2-operators complexity, 3-operators complexity}, as evidenced by the results of additional experiments reported in Section 6. Other settings may however be explored in future work, which may prove to be a more accurate way of measuring communicative cost and complexity.

For instance, to construct a complexity measure for the systems of indefinite pronouns, we have built on the theory of features of indefinite pronouns put forward in Haspelmath 2001. To our knowledge, Haspelmath's is the only feature-based theory intended to account for all of the semantic flavors of indefinite pronouns discussed here. There are, however, more recent approaches to the feature content of some of the sub-categories of indefinite pronouns, such as negative polarity and free choice indefinites (see for instance Chierchia 2013). These recent proposals may be developed further to construct alternative measures of complexity. In addition to this, one can conceive of measures of complexity not based on features, but for instance, solely on the number of items in the system.

Similarly, one may consider alternative measures of informativeness. In this respect, it is important to point out a complication in the present approach to measuring communicative cost. The measure of communicative cost incorporates the need to communicate different flavors; in the present, and much other work (e.g. Kemp and Regier (2012)), the communicative need for different flavors is inferred from the frequency of use of different items which denote those flavors in corpora. However, the frequency of use of different items may depend not only on the communicative need of the flavors they can express, but also on the items' complexity. If this is indeed the case, there is a bias in the trade-off analyses such as the one developed in Kemp and Regier (2012) and much related work, including the present one: more complex items in languages would tend to denote flavors which are less commonly expressed, contributing to the overall better simplicity/informativeness trade-off in such languages (cf. a related discussion in Enguehard and Spector (2021)). At present, it is not clear how to circumvent this issue.

8 Conclusions

We find that natural languages optimize the simplicity/informativeness trade-off in how they organize their indefinite pronoun systems. These results represent an extension of efficiency analyses to a system of function words, thus tying in with Steinert-Threlkeld 2019 and Uegaki 2020 in concluding that similar communication pressures are shaping both content and function word categories across languages.

Furthermore, we find that there is quite some variability in terms of what the closest point on the Pareto frontier would be for natural languages' indefinite pronoun systems in Experiment 1. Going back to the question of what causes variation between natural languages, this result suggests that some of the variation is due to languages 'finding' different solutions to the problem of simplicity/informativeness trade-off optimization. Finally, in Experiment 2, we find that artificial languages satisfying Haspelmath's universals are better at trading simplicity and informativeness than languages which do not. In other words, Haspelmath's universals contribute to the simplicity/informativeness trade-off optimization. Going back to the question of what causes linguistic universals, this result suggests that simplicity/informativeness trade-off optimization is one of their causes. A question that remains for future work is to find out which of Haspelmath's universals help with the optimization: it is conceivable that only a proper subset of them do, and that the rest need a different explanation.

Finally, we have discussed the assumptions required at each step of this efficiency analysis in the case of indefinite pronouns. These considerations must be made in any such analysis, and a thorough discussion of the choice points can help illuminate future work in this and related domains.

References

Maria Aloni and Floris Roelofsen. Indefinites in comparatives. *Natural Language Semantics*, 22(2):145–167, 2014. doi: https://doi.org/10.1007/s11050-013-9103-z.

Barend Beekhuizen, Julia Watson, and Suzanne Stevenson. Semantic typology and parallel corpora: Something about indefinite pronouns. In *39th Annual Conference of the Cognitive Science Society (CogSci)*, pages 112–117, 2017.

Giuliano Bernini and Paolo Ramat. *Negative Sentences in the Languages of Europe: A Typological Approach.* De Gruyter Mouton, 1996. doi: https://doi.org/10.1515/9783110819748.

Gennaro Chierchia. *Logic in Grammar: Polarity, Free choice, and Intervention.* Oxford University Press, 2013. doi: 10.1093/acprof:oso/9780199697977.001.0001.

Émile Enguehard and Benjamin Spector. Explaining gaps in the logical lexicon of natural languages: A decision-theoretic perspective on the square of aristotle. *Semantics and Pragmatics*, 14:5, 2021.

Gilles Fauconnier. Pragmatic scales and logical structure. *Linguistic Inquiry*, 6(3):353–375, 1975.

Ramon Ferrer i Cancho and Ricard V. Solé. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences (PNAS)*, 100(3): 788–791, 2003. doi: https://doi.org/10.1073/pnas.0335980100.

Michael C. Frank and Noah D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012. doi: 10.1126/science.1218633.

Martin Haspelmath. *Indefinite Pronouns.* Oxford University Press, 2001. doi: 10.1093/oso/9780198235606.001.0001.

Irene Heim. Remarks on comparative clauses as generalized quantifiers. Unpublished manuscript, MIT, 2006.

Joachim Jacobs. Lexical decomposition in Montague-grammar. *Theoretical Linguistics*, 7(1-3):121–136, 1980. doi: 10.1515/thli.1980.7.1-3.121.

Agnes Jäger. Anything is nothing is something. *Natural Language & Linguistic Theory*, 28(4):787–822, 2010. doi: 10.1007/s11049-010-9113-1.

Lauri Karttunen and Stanley Peters. Conventional lmplicature. In Choon-Kyu Oh and David A. Dinneen, editors, *Syntax and Semantics 11: Presupposition*, pages 1–56. Academic Press, New York, 1979. doi: 10.1163/9789004368880_002.

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*, 4:109–128, 2018. doi: 10.1146/annurev-linguistics-011817-045406.

Manfred Krifka. The semantics and pragmatics of weak and strong polarity items in assertions. In *Semantics and Linguistic Theory (SALT)*, volume 4, pages 195–219, 1994. doi: 10.3765/salt.v4i0.2462.

William A Ladusaw. *Polarity Sensitivity as Inherent Scope Relations.* PhD thesis, University of Texas, Austin, 1979.

Young-Suk Lee and Laurence Horn. Any as indefinite plus even. Unpublished manuscript, Yale University, 1994.

Andrew Makhorin. Glpk (gnu linear programming kit). *http://www. gnu. org/s/glpk/glpk. html*, 2008.

Doris Penka. *Negative Indefinites.* Oxford University Press, 2011. doi: 10.1093/acprof:oso/9780199567263.001.0001.

Asya Pereltsvaig. Negative polarity items in Russian and the 'Bagel problem'. In Sue Brown and Adam Przepiorkowski, editors, *Negation in Slavic*, pages 1–28. Slavica Publishers, Bloomington, 2004.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

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. doi: 10.1002/9781118346136.ch11.

Ian Roberts and Anna Roussou. *Syntactic Change: A Minimalist Approach to Grammaticalization.* Cambridge University Press, 2003. doi: 10.1017/cbo9780511486326.

Eleanor Rosch. Principles of categorization. In Eleanor Rosch and Barbara Bloom Lloyd, editors, *Cognition and categorization.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.

Alexander Schrijver. *Theory of linear and integer programming.* John Wiley & Sons, 1998.

Dirk Schumacher. ompr: an alternative way to model mixed-integer linear programs. In *The R User Conference, useR! 2017 July 4-7 2017 Brussels, Belgium*, page 59, 2017.

Brian Skyrms. *Signals: Evolution, learning, and information.* Oxford University Press,

2010. doi: 10.1093/acprof:oso/9780199580828.001.0001.

Shane Steinert-Threlkeld. Quantifiers in natural language optimize the simplicity/informativeness trade-off. In *Amsterdam Colloquium 2019*, 2019. URL `http://events.illc.uva.nl/AC/AC2019/uploaded_files/inlineitem/Steinert-Threlkeld_Quantifiers_in_natural_language_.pdf`.

Wataru Uegaki. NAND and the communicative efficiency model. Unpublished manuscript, University of Edinburgh, 2020.

Yang Xu and Terry Regier. Numeral systems across languages support efficient communication: From approximate numerosity to recursion. In *36th Annual Meeting of the Cognitive Science Society (CogSci)*, pages 1802–1807, 2014.

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. doi: 10.1111/cogs.12312.

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. doi: 10.1162/opmi_a_00034.

Hedde Zeijlstra. *Sentential Negation and Negative Concord.* PhD thesis, University of Amsterdam, 2004.

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