The influence of the simplicity/informativeness trade-off on the sematic typology of quantifiers

MSc Thesis *(Afstudeerscriptie)*

written by

Wouter Posdijk
(born 7 June 1996 in Zeist)

under the supervision of **Shane Steinert-Threlkeld**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
26 June 2019	Paul Dekker
	Shane Steinert-Threlkeld
	Jakub Szymanik
	Maria Aloni
	Jelle Zuidema

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

Previous research has shown that a trade-off between simplicity and informativeness can explain the semantic typology of various semantic domains of content words, such as color, kinship and folk biology (Kemp, Xu & Regier, 2018). In this thesis, I investigate whether this trade-off extends to quantifiers, a domain of function words with a well-established semantic typology. In particular, I develop measures of simplicity (based on logical description length) and informativeness (based on communicative use) of individual quantifiers and quantifier languages and I investigate the relation between properties of natural languages, such as monotonicity, to optimality with respect to the trade-off. Results show that languages of natural quantifiers perform better than random languages, and that monotonicity correlates with how optimal a language is.

Contents

Chapter 1

Introduction

Human language both reflects and shapes human culture and cognition (Tomasello, 2009).

A major field that endeavors to understand culture and cognition through language is that of *Semantic Typology*, the study of word meanings in the languages of the world (Evans, 2010). The meanings that are expressed in our languages are not random; rather, they may be the result of millennia of cultural evolution and cognitive influence, and may reflect constraints shaped by biological evolution (Scott-Phillips, 2014). Which of these may have caused our semantic typology, and to what extent, is exactly the main question asked in this field.

Kemp, Xu, and Regier (2018) have reviewed a recent focus on the influence of a trade-off between *simplicity* and *informativeness* on the typology of various semantic domains such as kinship, color and folk biology. The idea is that the lexicalization of meanings is governed by a pressure to, on the one hand, be as *simple* as possible—such that languages are easy to understand—and on the other hand, be as *informative* as possible—such that communication may be successful. This pressure could occur for various reasons: perhaps there is mental pressure, evolutionary pressure, or cultural pressure causing this optimization. But current research does not yet attempt to explain the source of this pressure. Instead, it focuses on evaluating whether or not natural languages are optimized with respect to the trade-off.

Kemp, Xu, and Regier (2018) argue that this trade-off has significant influence on the semantic typology of kinship, color and folk biology. In this thesis I evaluate whether this view extends to the semantic domain of *quantifiers*. Quantifiers are an interesting case study because they are *function words* as opposed to *content words*. This means, roughly, that they only become meaningful when placed in a sentence, since their meaning is abstract. For example, 'all' does not refer to concrete entities in the world, but 'All dogs bark' *is* an informative sentence. 'Brother', on the other hand, refers to concrete entities and is thus a content word. For more detail on categories of content words and function words, see

Baker (2003) and Ouhalla (2003) respectively. Quantifiers being function words makes the measuring of their complexity and informativeness more challenging, since their meaning is less clear-cut. However, the semantic typology of quantifiers is well-studied (Szymanik, 2016, Van Benthem, 1986), providing a clear direction for said measures.

Thus, Quantifiers are an interesting case study for two reasons: (i) Unlike the domains studied before, they are function words and (ii) they have a well-studied semantic typology.

On top of extending the simplicity/informativeness research to quantifiers, I provide two methodological additions. Firstly, I provide a single, numerical measure of optimality with respect to the trade-off. Secondly, I have performed hypothesis testing using said measure of optimality, providing a statistical answer to the research questions at hand. This is an improvement to the field; while e.g. Kemp and Regier (2012) have convincing graphs, they lack statistical analysis to back up their claims.

In Chapter 2, I go into the necessary background information: previous research on the simplicity/informativeness trade-off and the semantic typology of quantifiers. In Chapter 3, I evaluate the trade-off for individual quantifiers, comparing lexicalized English quantifiers to logically possible ones. In Chapter 4, I extend this approach to whole languages, testing hypotheses using a regression analysis on the optimality of natural versus logically possible languages. In Chapter 5, I provide some future directions and Chapter 6 concludes.

Chapter 2

Simplicity, Informativeness and Quantifiers

In this chapter, I will first explain the basis of evaluating the simplicity/informativeness trade-off, alongside some prominent examples. Then, I will go into the semantic typology of quantifiers; I will cover the main framework for abstractly representing the meaning of quantifiers and I will cover the knowledge that is currently available about quantifiers in various languages throughout the world. Finally, I will consider how I extend research on the simplicity/informativeness trade-off to the semantic domain of quantifiers.

2.1 Simplicity vs. Informativeness

Languages need to facilitate efficient communication. To facilitate efficient communication, they need to be *informative* on the one hand, allowing for conveying information with reasonable accuracy, and *simple* on the other, keeping the cognitive load within reasonable limits.

One can see quickly how these two principles trade off against one another. Consider the two following examples:

1. L_1 has one word for every possible piece of information one may want to convey. It has a word for “The red coffee cup on the table in the Master of Logic study room”, as well as a word for “The feeling of excitement I felt yesterday when giving a presentation”, and so on. Thus it allows those communicating using this language to communicate every situation perfectly using only a single word, but this comes at the cost of an infinite lexicon.

2. L_2 only contains the word ‘everything’. This language requires little to no cognitive effort, but does not allow its users to transmit any meaningful information.

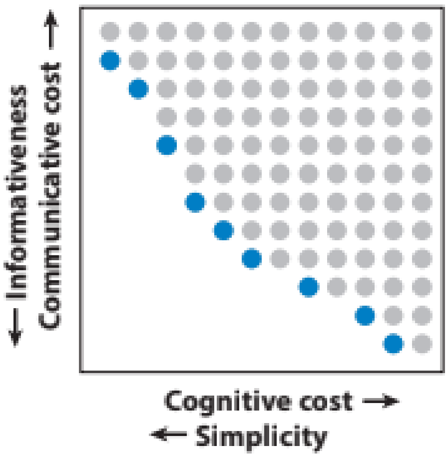


Figure 2.1: Possible solutions to a multi-objective optimization problem. The blue dots from the Pareto Front. From Kemp, Xu, and Regier (2018)

These two examples illustrate that these two objectives have the key properties of a trade-off: maximizing informativeness necessarily implies sub-optimal simplicity (as in L_1), and maximizing simplicity necessarily implies sub-optimal informativeness (as in L_2).

2.1.1 Multi-objective optimization

A problem where one tries to optimize for multiple objectives is a *multi-objective optimization problem* (Deb, 2014). If the two objectives are competing, there can be many possible solutions to such a problem. A solution is called *dominant* or *Pareto optimal* if there is no solution that performs better at *all* objectives. The set of such Pareto optimal solutions is called the Pareto front. Figure 2.1 gives an idea of what such a front looks like.

The simplicity/informativeness trade-off is thus a multi-objective optimization problem with simplicity and informativeness as its objectives. Research regarding the trade-off considers whether or not natural languages lie at or close to the Pareto front said problem. If so, then that implies that the trade-off is a factor in shaping language.

The large variety of solutions at the Pareto front—from very informative and not very simple to very simple but quite uninformative—is generally considered to account for the differences between natural languages. Cultures that place more emphasis on the semantic domain in question are expected have languages that lie more to the extreme ends of the Pareto front than those for cultures that lack such emphasis (Kemp, Xu, and Regier, 2018).

2.1.2 Measuring simplicity and informativeness

In this section, I discuss the general strategies used to measure simplicity (or complexity) and informativeness (or communicative cost).

Simplicity

Simplicity is generally measured in terms of the *mental representation* of the words in a language. If we know how the words are encoded mentally, we can make claims about the complexity thereof. Commonly, mental representations are built up as logical formulas in a predefined *Language of Thought* (Piantadosi and Jacobs, 2016), and language complexity can then be expressed as logical description length of some form. Kemp and Regier (2012) apply such a measure for kinship; I discuss how in the next subsection.

Alternatively, complexity can be measured using information-theoretic minimal codings (Carr et al., 2018); the length of these codings can point towards a complexity measure. These minimal codings are based on the extension, not on some separate mental representation. Thus the minimal codings *are* the mental representation used by this measure.

In the present research, I opt for the former approach, since it fits better with the current view of the semantic typology of quantifiers.

Informativeness

Informativeness is generally measured on the basis of probability of communicative success. This depends on the *extensions* of the words in the language; the things, be they abstract or not, that the words denote. Informativeness measures can be motivated from information theory (Carr et al., 2018) or from game-theoretic linguistics, drawing upon signalling games (Jäger, 2008, Huttegger and Zollman, 2011). The latter approach is adopted here, using the following *communication game*:

- $S =$ a set of possible states of the world
- $M = \{m_1, ..., m_n\}$, a set of messages
- $L : M \to \mathcal{P}(S)$ is the language, mapping messages to sets of possible states.

With the following rules:

1. A sender perceives a state $s \in S$ of the world (from some distribution $P(S)$).
2. The sender chooses m such that $s \in L(m)$.
3. A receiver perceives message m.

4. The receiver infers some s' such that $s' \in L(m)$.

The utility $u : S \times S \to [0,1]$ of this game can be 1 if and only if $s = s'$ for perfect accuracy, or can model similarity between s and s' to allow for more vagueness. If we define $\sigma(s) = \{m : s \in L(m)\}$, then the overall utility of the language is

$$u(L) = \sum_{s \in S} \sum_{m \in M} \sum_{s' \in S} P(s)P(m|s)P(s'|m) \cdot u(s, s')$$

This is the probability that the sender successfully uses the language to communicate an arbitrary state. This utility is the measure of informativeness.

2.1.3 Kinship

Kemp and Regier (2012) have shown that the trade-off between simplicity and informativeness is a major factor shaping languages in the semantic domain of kinship. Figure 2.2 (Kemp, Xu, and Regier, 2018) depicts a communicative scenario where ambiguity of a word reduces informativeness—'Brother' can mean both *Older brother* and *Younger brother*, so when it is uttered, the listener has incomplete information. Figure 2.3 shows the kinship category systems of English and Norther Paiute. There one can easily see that Northern Paiute has more categories, but also more specific categories, and indeed, the latter has higher informativeness but lower simplicity.

Informativeness was measured using information-theoretic communicative cost based on the extensions of the words in the family trees seen in figure 2.3. Simplicity was measured in the form of complexity, based on the minimum number of formulas required to describe every word in a language using a logical *Language of Thought* with predicates such as `PARENT(x,y)`, `OLDER(x,y)` and `MALE(x,y)`. Sampled, logically possible languages were compared to the kinship systems actually attested in natural languages. This yielded the results seen in Figure 2.4: natural languages (black) lie relatively close to the Pareto frontier, compared to logically possible languages (grey). This result implies that the trade-off is a major factor in shaping natural languages within the domain of kinship; natural languages are clearly relatively optimized with respect to the trade-off.

2.2 The Semantic Typology of Quantifiers

2.2.1 Generalized quantifier theory

Quantifiers are generally represented in terms of generalized quantifier theory (Szymanik, 2016). Within this theory, quantifiers talk about three sets: $A, B \subseteq M$, where M is the universe of objects. For example, in the case of 'All dogs bark', A is the set of dogs, B is the set of things that bark. In generalized quantifier

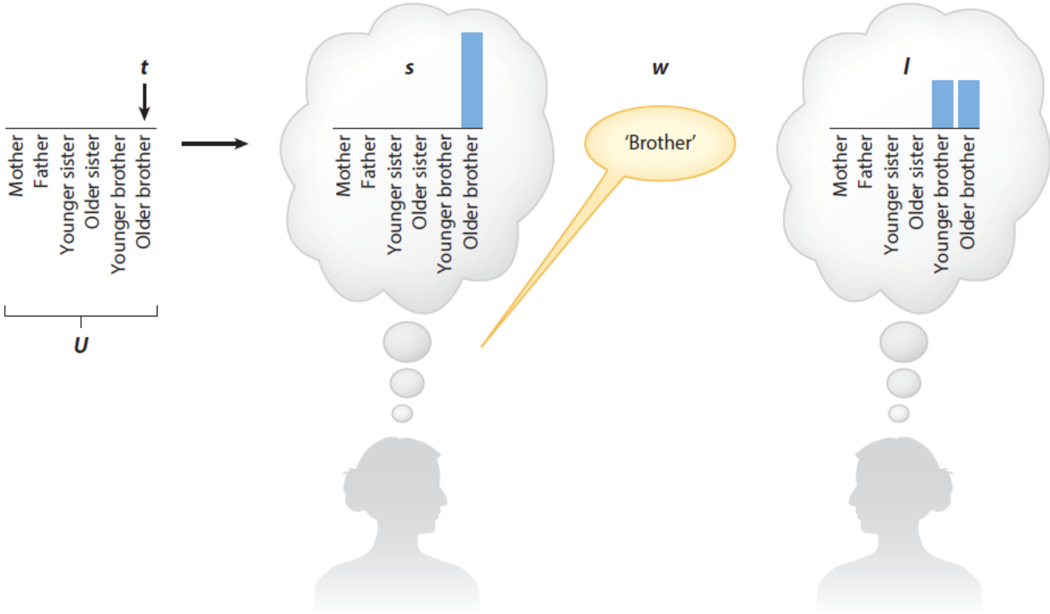


Figure 2.2: A communicative scenario where information is lost due to the fact that the word 'brother' is ambiguous. The speaker perceives the state 'older brother' and communicates said state using the word 'brother'. The receiver cannot retrieve the exact information state of the sender due to the ambiguity of 'brother'. From Kemp, Xu, and Regier (2018)

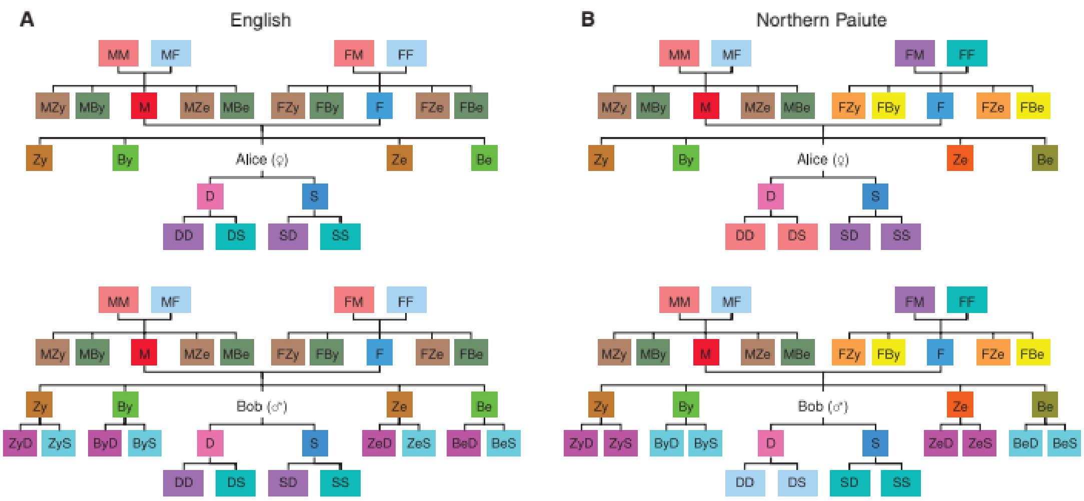


Figure 2.3: Category systems for kinship in English and Northern Paiute, from both a male and female speaker. The colors denote different categories. From Kemp, Xu, and Regier (2018).

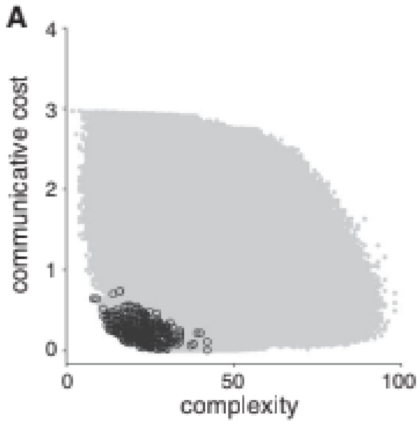


Figure 2.4: The complexity and communicative cost of natural languages (black) versus logically possible languages (grey). From Kemp and Regier (2012).

theory, 'all' is then equated with the set of all models (M, A, B) such that "All A's are B" holds. Such quantifiers can be described using logical formulas. In the current research, instead of equating quantifiers with the set of models where they are true, I equate them with such formulas. Thus,

$$\mathsf{all} := \lambda M.\lambda A.\lambda B.A \subseteq B$$

I omit the $\lambda M.\lambda A.\lambda B.$ throughout the rest of the thesis.

The set of models where this formula is true is then the *extension* of the quantifier:

$$\begin{aligned} [\![\phi]\!] &= \{(M, A, B) : \phi(M, A, B)\} \\ [\![\mathsf{all}]\!] &= \{(M, A, B) : A \subseteq B\} \end{aligned}$$

Analyzing quantifiers in these terms allows one to discern formal properties that may hold for some quantifiers but not for others. Some of these properties are relatively universal; that is to say, they are true for almost all quantifiers that occur in natural language. Four such properties discovered are the following (Barwise and Cooper, 1981, Van Benthem, 1986, Szymanik, 2016):

Extensionality A quantifier Q is *extensional* iff:

$$\text{If } M \subseteq M' \text{ and } (M, A, B) \in [\![Q]\!], \text{ then } (M', A, B) \in [\![Q]\!] \tag{EXT}$$

This means that the truth-value of Q depends only on A and B, and not on M. This excludes e.g. '$|M| > 5$'.

Isomorphism Invariance A quantifier Q is *isomorphism invariant* (or topic neutral) iff:

$$\text{If } (M, A, B) \cong (M', A', B'), \text{ then } (M, A, B) \in [\![Q]\!] \Leftrightarrow (M', A', B') \in [\![Q]\!] \tag{ISOM}$$

where isomorphism means that there is a one-to-one mapping between the models that preserves set cardinalities. This implies *topic neutrality*: instead of depending on the specific objects in the sets, isomorphism-invariant quantifiers depend only on the cardinalities $|A - B|, |A \cap B|, |B - A|$ and $|M - (A \cup B)|$ (Szymanik, 2016). This excludes, for example, $\mathsf{Jack} \in A \cap B$.

Conservativity A quantifier Q is *conservative* iff:

$$(M, A, B) \in [\![Q]\!] \Leftrightarrow (M, A, A \cap B) \in [\![Q]\!] \tag{CONS}$$

This property has the effect that the truth-value of the quantifier only depends on the elements of A. This excludes, for example, $|B - A| > 0$, since that depends on elements of $B - A$, which are not in A.

Monotonicity A quantifier Q is upward-right monotone iff:

$$\text{If } B \subseteq B' \subseteq M \text{ and } (M, A, B) \in [\![Q]\!], \text{ then } (M, A, B') \in [\![Q]\!] \tag{MON↑}$$

upward-left monotone iff:

$$\text{If } A \subseteq A' \subseteq M \text{ and } (M, A, B) \in [\![Q]\!], \text{ then } (M, A', B) \in [\![Q]\!] \tag{↑MON}$$

downward-right monotone iff:

$$\text{If } B \subseteq B' \subseteq M \text{ and } (M, A, B') \in [\![Q]\!], \text{ then } (M, A, B) \in [\![Q]\!] \tag{MON↓}$$

and downward-left monotone iff:

$$\text{If } A \subseteq A' \subseteq M \text{ and } (M, A', B) \in [\![Q]\!], \text{ then } (M, A, B) \in [\![Q]\!] \tag{↓MON}$$

Most natural language quantifiers are monotone in one way or another.

For example, ‘all’ is downward-left monotone:

All instruments produce music $\Rightarrow$ All guitars produce music

since INSTRUMENTS $\supseteq$ GUITARS. Simultaneously, ‘all’ is upward-right monotone:

All instruments produce music $\Rightarrow$ All instruments produce sound

since MUSIC_PRODUCERS $\subseteq$ SOUND_PRODUCERS.

The reader can verify that ‘no’ is downward-monotone in both arguments, and ‘most’ is upward-monotone in the right argument, but not monotone in the left. Clearly there is a large diversity with respect to monotonicity, but no lexicalized quantifiers are completely non-monotone. This universal rules out quantifiers like $\lambda M.\lambda A.\lambda B.|A \cap B| = 5$. That quantifier is only expressed in e.g. English as ‘exactly 5’, and is thus not *lexicalized* since it cannot be conveyed with a single word.

Quantifier	Pressuposition	Assertion
both	$\lvert A\rvert = 2$	$A \subseteq B$
the	$\lvert A\rvert = 1$	$A \subseteq B$
neither	$\lvert A\rvert = 2$	$\lvert A \cap B\rvert = 0$

Table 2.1: English quantifiers with presupposition

2.2.2 Presupposition

Piantadosi, Goodman, and Tenenbaum (2013) express some English quantifiers terms of logical formulas, noting that many quantifiers also have a presupposition (Beaver, 1997). These presuppositions can also be expressed as logical formulas. For some English examples, see Table 2.1.

2.2.3 Natural languages

To get a full view of the semantic typology of Quantifiers, cross-linguistic analysis is also required. The *Handbook of Quantifiers in Natural Language* (Keenan and Paperno, 2012), henceforth the *Handbook*, contains a survey of quantifiers in various natural languages. Keenan and Paperno identify three main cross-linguistic semantic categories of quantifiers, expressed in generalized quantifier theory:

1. Generalized Existential Quantifiers, such as 'no', which talk about $|A \cap B|$. This also includes numbers (e.g. $5 := |A \cap B| \geq 5$)

2. Generalized Intersective Quantifiers, which talk about $|A - B|$. An example is 'all' ($|A - B| = 0$)

3. Proportional quantifiers, talking about $|A - B|/|A|$, such as 'many'.

And they survey whether and how these types of quantifiers are expressible within the languages covered in the Handbook. Lexicalized quantifiers - those that are only a single word - are seemingly restricted to these three categories across languages. The main difference to be found is in the syntactic role of the quantifiers; but that is outside of the scope of the present research.

2.3 Simplicity vs. Informativeness for Quantifiers

Generalized quantifier theory provides us with both an extension and a mental representation of quantifiers in the form of a logical Language of Thought. The extension, when combined with communication games, can yield a measure of informativeness. The mental representation points to measures of logical complexity, such as total formula length.

In addition, the four semantic universals and universal categories from the *Handbook* provide properties of natural languages. I will investigate the relation between these properties and the simplicity/informativeness trade-off.

Chapter 3

Individual quantifiers

The simplicity/informativeness trade-off has been researched in a variety of semantic domains, but such research has mainly been focused on languages as a whole. However, individual words are also cognitive entities used for communication, and may therefore also be subject to the trade-off. In this chapter I investigate whether we can identify such a trade-off for individual quantifiers and, if it exists, whether English quantifiers are optimized with respect to said trade-off. For the sake of detail at the level of individual quantifiers, I consider both quantifiers with and without presupposition in this chapter.

The first section covers the measures that are used, along with their motivation. The second section covers the experiments: how quantifiers are generated, from which grammars, and how they perform compared to English quantifiers. The third section provides some points of discussion and the fourth concludes.

3.1 Measures

In this section, I cover the measures used for simplicity (or complexity) and informativeness (or communicative cost).

3.1.1 Complexity

The measure of complexity is based on the mental representation of quantifiers given by Generalized Quantifier Theory. For some ϕ expressed in logical language $\mathcal{L}$ (to be defined in the experiments):

$$\mathtt{comp}(\phi) = \begin{cases} \mathtt{basecomp}(f) + \sum_{i=1}^{k} \mathtt{comp}(\psi_i) & \text{if } \phi = f(\psi_1, ..., \psi_k), f \in \mathcal{O} \\ \mathtt{basecomp}(\phi) & \text{otherwise} \end{cases}$$

Where `basecomp`: $\mathcal{L} \to [0, 1]$ captures the complexity of the operators.

This measure can also be extended to quantifiers with a presupposition. To combine the the complexity of an assertion and a presupposition, I simply take the average, for a quantifier ϕ_ψ with presupposition ψ and expression ϕ:

$$\texttt{comp}(\phi_\psi) = \frac{\texttt{comp}(\phi) + \texttt{comp}(\psi)}{2}$$

In the experiments in this Chapter, all quantifiers, whether they have a presupposition or not, are measured using this measure (where $\texttt{comp}(\psi) = 0$ if the quantifier does not have a presupposition).

Choice of operators

What can be noted about quantifier representations using logical formulas, is that the choice of primitives heavily influences what the formulas look like. For example, 'all' may be written as $A \subseteq B$, $|A - B| = 0$ or $\texttt{is\_empty}(A - B)$, depending on what is allowed in the formulas.

Thus the measure of complexity has become heavily dependent on the operators that we use. This makes the choice of operators all the more important - a rather problematic fact, because choosing appropriate operators is not at all an easy task. There is as of yet no clear reason why we should, for example, include an $\texttt{is\_empty}$ predicate or simply introduce '$=$' and '$\emptyset$', or even introduce '$|\cdot|$', '$=$' and '0'.

To mitigate this problem somewhat, I use 3 different sets of operators in the experiments.

3.1.2 Informativeness

To measure informativeness, I draw upon the communication-based measure for languages introduced in Section 2.1, assuming $u(s, s') = 1$ if $s = s'$, 0 otherwise. This simplifies the utility:

$$U(L) = \sum_{s \in S} \sum_{m \in M} P(s)P(m|s)P(s|m)$$

What remains is to reduce this measure to a measure for individual words. A first attempt is to simply restrict the game to using only a single message m'. In that case, $P(m|s) = 1$ only if $m = m'$, and 0 otherwise. Also, s can only be conveyed if $s \in L(m')$. Thus we can reduce the utility to:

$$U(L) = \sum_{s \in L(m')} P(s)P(s|m')$$

Assuming that both the world and the language are uniformly distributed, this yields

$$U(L) = \sum_{s \in L(m')} \frac{1}{|S|} \frac{1}{|L(m')|} = \frac{1}{|S|}$$

Thus the utility of the language is completely invariant to the actual meaning of the word. This can be attributed to the fact that when using only a single word, the objectives of covering the meaning space and having specific meanings start to compete.

Therefore it seems that this simplification is not adequate. Another possibility is to consider the possible ways in which individual words can drive up the utility of the entire language regardless of other words. This mainly seems to be by keeping $P(s|m)$ large, and thus by keeping the meanings small. Assuming a uniform distribution, this yields us the following measure of informativeness:

$$\texttt{inf}(m') = \frac{1}{|L(m')|}$$

For the purposes of the experiment, I change it to a measure for communicative cost, normalizing with respect to the state space:

$$\texttt{communicative\_cost}(m') = \frac{|L(m')|}{|S|}$$

Matching the conception of word information in possible word semantics (Stalnaker, 1978). Thus, for any possible quantifier representation, if we know its extension, we can measure its informativeness.

Presupposition

Quantifiers that consist of both an assertion and a presupposition make this measure a little more difficult.

The 'successful communication' principle points towards a relative measure: I measure the specificity of the assertion relative to the presupposition. This matches the idea that presuppositions form common ground between speakers.

Thus, if a word m contains an assertion meaning $L(m)$ and a presupposition meaning $L_p(m)$, where $L(m) \subset L_p(m)$ then its communicative cost will be:

$$\texttt{communicative\_cost}(m) = \frac{|L(m)|}{|L_p(m)|}$$

Defining truth of a quantifier in model s as follows:

$$[\![\phi_\psi]\!]^s = \begin{cases} * & \text{if } [\![\psi]\!]^s = 0 \\ [\![\phi]\!]^s & \textit{otherwise} \end{cases}$$

where $[\![\phi]\!]_S = \{s \in S : [\![\phi]\!]^s = 1\}$, $[\![\phi]\!]^s = 1$ if ϕ is true in s, 0 otherwise, the final measure of communicative cost becomes:

$$\texttt{communicative\_cost}(\phi_\psi) = \frac{|\{s \in S | [\![\phi_\psi]\!]^s = 1\}|}{|[\![\psi]\!]_S|}$$

This is the measure used in the experiments.

Specification of the state space

In order to define the state space, I look towards Generalized Quantifier Theory.

A generalized quantifier model (M, A, B) consists of a universe M, and two subsets $A, B \subset M$. A generalized quantifier is then a mapping from the set of models $\mathcal{M}$ to truth-values $\{0, 1\}$.

For computational purposes I assume a single universe of objects M of 10 elements. By leaving M out of the grammar of the quantifiers generated, this does not become a problem, although it does build the semantic universal of *extensionality* into the model.

While the most straightforward option is to let the meaning space $S' = \{(M, A, B) | A, B \subset M\}$, further restriction is needed to allow for easier computation. I do not allow any direct object referencing (through e.g. $\in$) in the quantifiers, rendering them *isomorphism invariant*. Thus we can let $S = {}^{S'}\!/_{\cong}$, significantly decreasing the size of possible meanings.

I implement this quotient by considering models based only on the numerosities of A, B and $A \cap B$ (all other relevant numerosities can then be derived). Thus $S = \{(n_A, n_B, n_{A\cap B}) : (n_A + n_B - n_{A\cap B}) \leq |M|, n_{A\cap B} \leq n_A, n_{A\cap B} \leq n_B\}$

3.2 Experiments

In this section, I consider an *expression* to be a single logical formula, and I consider a *quantifier* to be a combination of an assertive expression and possibly a presupposition expression.

I run two experiments. Both consists of the following steps:

1. Generate possible expressions E up to length n given logical language $\mathcal{L}$.

2. Compose quantifiers from the set of expressions.

3. Calculate the communicative cost and complexity of both the generated and lexicalized quantifiers (as seen in Table 3.1)

The code for the experiments can be found at https://github.com/wouterposdijk/SimInf_Quantifiers.

In this section, I first cover which expressions are generated. Next, I explain how these expressions are combined into quantifiers. Then I provide a quick recap

Quantifier	Presupposition	Assertion
all	*none*	$A \subseteq B$
both	$\|A\| = 2$	$A \subseteq B$
the	$\|A\| = 1$	$A \subseteq B$
most	*none*	$\|A \cap B\| > \|A - B\|$
a	*none*	$\|A \cap B\| > 0$
many	*none*	$\|A \cap B\|/\|A\| > .25$
no	*none*	$\|A \cap B\| = 0$
neither	$\|A\| = 2$	$\|A \cap B\| = 0$

Table 3.1: The lexicalized quantifiers used in the experiment.

of the measures used, after which I discuss the parameter assignments of the two experiments. Finally, I discuss the results of the experiments.

3.2.1 Expression generation

The input parameters for expression generation are the set of operators $\mathcal{O}$, the meaning space S, and the largest expression length n.

The set of all expressions up to length n is

$$E = \{\phi \in \text{Form}(\mathcal{L}) | \texttt{len}(\phi) \leq n\}$$

where

$$\texttt{len}(\phi) = \begin{cases} 1 + \sum_{i=1}^{k} \texttt{len}(\psi_i) & \text{if } \phi = f(\psi_1, ..., \psi_k) \\ 1 & \text{otherwise} \end{cases}$$

and

$$\mathcal{L} = \{A, B\} \cup \{0, 2, ..., 10\} \cup \{0.1, 0.2, ..., 1.0\} \cup \{\top, \bot\} \cup \mathcal{O}$$

Then the desired set of expressions is the restriction of E to the lowest possible complexity:

$$E^* = \{\phi | \forall \psi \in \text{Form}(\mathcal{L}) : [\![\psi]\!]_S = [\![\phi]\!]_S \to \texttt{comp}(\phi) \leq \texttt{comp}(\psi)\}$$

This 'shortening' removes for example '$|A| - |A - B| = 0$' in favor of '$\texttt{is\_empty}(A \cap B)$' (depending on the allowed operators and their base complexity). This way, there are no expressions that are more complex than they need to be.

3.2.2 Quantifier generation

Quantifiers, consisting of a presupposition and and a primary expression, were generated simply by taking all possible combinations of boolean expressions (except $\bot$), and then removing any combination that has an extensionally equivalent simpler combination.

operator	**type**		
$\vert\cdot\vert$	SET	$\rightarrow$	INT
$\subset$	SET^2	$\rightarrow$	BOOL
$\cap$	SET^2	$\rightarrow$	SET
$\cup$	SET^2	$\rightarrow$	SET
$-_s$	SET^2	$\rightarrow$	SET
$-_n$	INT^2	$\rightarrow$	INT
$+$	INT^2	$\rightarrow$	INT
$>_n$	INT^2	$\rightarrow$	BOOL
$\geq_n$	INT^2	$\rightarrow$	BOOL
$>_q$	FLOAT^2	$\rightarrow$	BOOL
$=_n$	INT^2	$\rightarrow$	BOOL
$=_q$	FLOAT^2	$\rightarrow$	BOOL
$/$	INT^2	$\rightarrow$	FLOAT
$\wedge$	BOOL^2	$\rightarrow$	BOOL
$\vee$	BOOL^2	$\rightarrow$	BOOL
$\neg$	BOOL	$\rightarrow$	BOOL

Table 3.2: The basic operators used in the experiments.

3.2.3 Measuring

This subsection contains a quick recap of the measures covered above. The complexity and communicative cost of quantifiers ϕ_ψ are measured as follows:

$$\texttt{comp}(\phi_\psi) = \frac{\texttt{comp}(\phi) + \texttt{comp}(\psi)}{2}$$

$$\texttt{communicative\_cost}(\phi_\psi) = \frac{|\{s \in S | [\![\phi_\psi]\!]^s = 1\}|}{|[\![\psi]\!]_S|}$$

3.2.4 Parameters

Both experiments were run with universe size 10. The first experiment considers only quantifiers without presupposition, with an assertion up to length 12. The second considers quantifiers with assertions and presuppositions up to length 7.

Operators

Both experiments were ran for three sets of operators. See the basic operators used in Table 3.2, and the more complicated ones defined in their terms in Table 3.3.

The latter do not actually increase expressive power overall, but do allow for more expressive power at lower levels of complexity.

The operator sets used are the following:

operator	meaning
`is_empty`(X)	$\|X\| = 0$
`is_nonempty`(X)	$\|X\| > 0$
`proportion>=`(X, Y, q)	$\|X\|/\|Y\| \geq q$

Table 3.3: The more complex operators used in the experiments.

- $\mathcal{O}_{\text{numerical}} = \{|\cdot|, \subset, \cap, \cup, -_s, -_n, +, >_n, \geq_n, >_q, =_n, =_q, /\}$
- $\mathcal{O}_{\text{logical}} = \{|\cdot|, \subset, \cap, \cup, -_s, -_n, +, >_n, \geq_n, >_q, =_n, =_q, /, \wedge, \vee, \neg\}$
- $\mathcal{O}_{\text{extended}} = \{|\cdot|, \subset, \cap, \cup, -_s, -_n, +, >_n, \geq_n, >_q, =_n, =_q, /, \wedge, \vee, \neg,$ `is_empty`, `is_nonempty`, `proportion>=`$\}$

3.2.5 Results

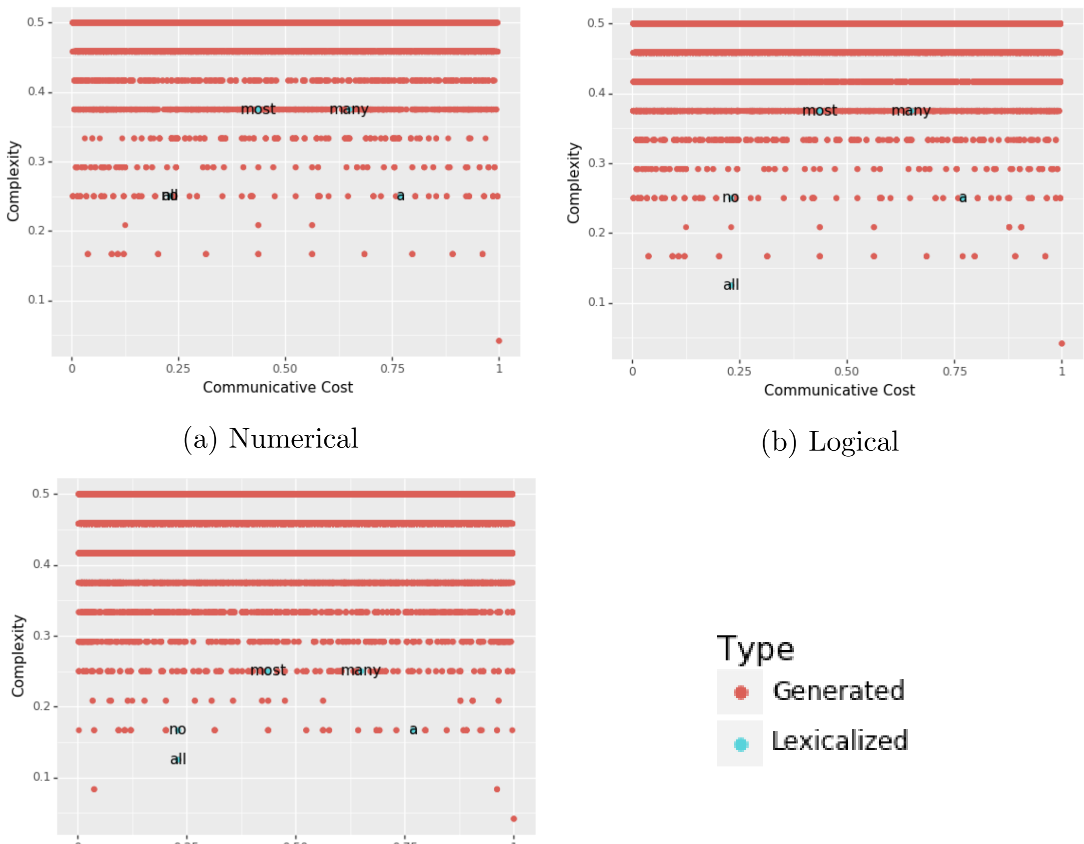
See the outcomes of the experiments in Figure 2.1 and 2.2. It is hard to detect any form of a Pareto front in the result graphs, both with and without presupposition. It seems that, especially as complexity ranks up, all levels of informativeness are possible. Naturally we see this phenomenon at lower complexities with the extended primitive list.

3.3 Discussion

The main implication of the results seems to be that there is no trade-off between simplicity and informativeness at the individual word level. Different primitive sets yield no tangible differences, and neither does adding presuppositions.

Reduction of informativeness measure The lack of a trade-off corresponds to the fact that we could not reduce the game-theoretic measure of informativeness for whole languages to individual words - the informativeness measure might simply not correspond strongly enough to the informativeness of whole languages. At the whole-language level a pressure to keep the amount of words down might keep the specificity of those words low - a factor that is not at all present when only considering individual words.

No optimization with respect to complexity Still, it is quite striking that the lexicalized individual words are not even remotely optimized with respect to complexity in logical terms. Perhaps there are factors at the whole language level that warrant this lack of specificity (such as there being too much overlap between the various simpler quantifiers), but another real possibility is that the measure of complexity is simply not very accurate; We have little knowledge of the ‘actual’ set of operators, let alone how complex they are in themselves.

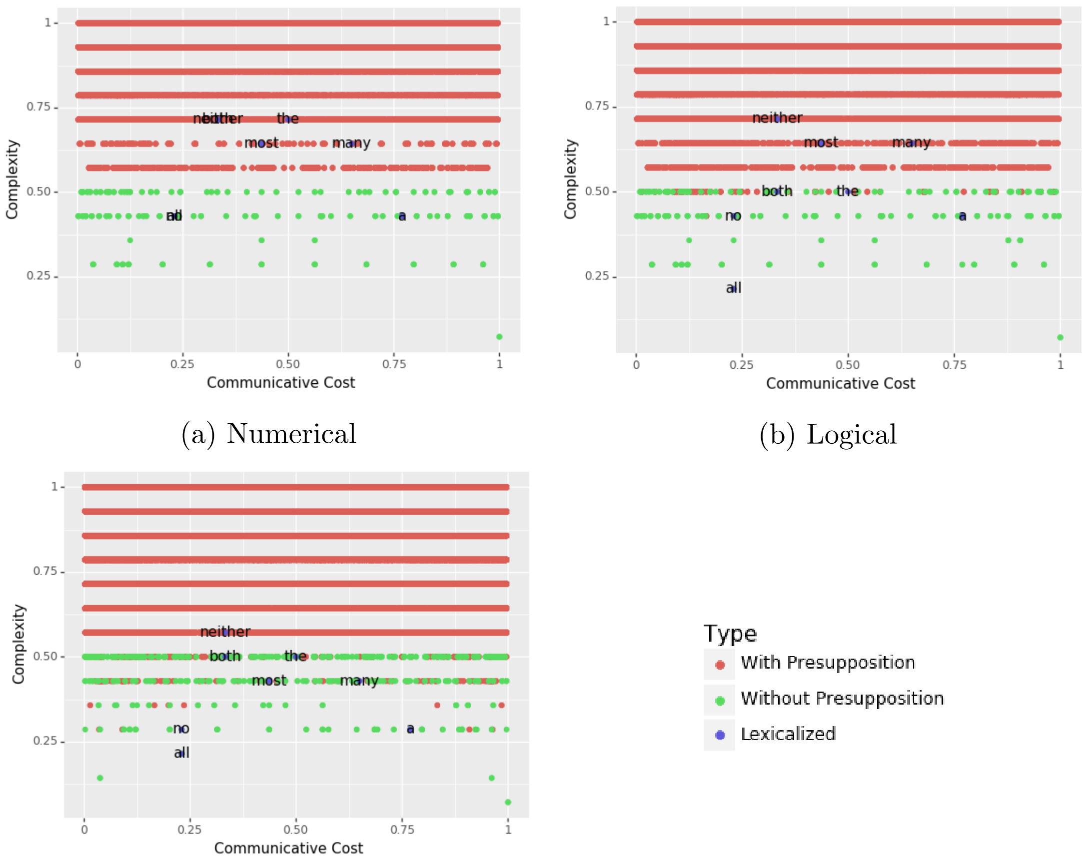


(a) Numerical

(b) Logical

(c) Extended

Figure 3.1: The results of the first experiment, generating quantifiers up to length 12 in a universe of size 10, and measuring them according to the measures specified.



(a) Numerical

(b) Logical

(c) Extended

Figure 3.2: The results of the second experiment, generating quantifiers up to length 7 and combining them into all possible combinations of assertion and presupposition.

Thus, for lack of a detailed complexity measure, the biggest contender for complexity at the language level ought to be *word count*; a factor that is obviously completely absent in the individual-word approach.

3.4 Conclusion

Even under varying circumstances, there is no Pareto front to be found for individual quantifiers. The way informativeness scales down to individual words combined with proportional and comparative quantifiers makes for a picture that tells us little about the semantic typology of quantifiers. It seems that, for individual words, there is no trade-off between simplicity and informativeness as interpreted here.

Chapter 4

Quantifier Languages

While a trade-off at the level of individual words seems unlikely, whole languages (considered in this thesis to be sets of expressions, not including quantifiers with presupposition) can still be expected to trade off. The example given in section 2.1—one word for everything versus one word for each thing—serves to show that having multiple words provides more ways to trade off simplicity and informativeness. This leads to the belief that, although individual words do not seem to follow any trade-off pattern, languages may still do so.

The key question answered in this chapter is thus the following: is there a trade-off between simplicity and informativeness at the level of whole languages, and if so, do natural languages optimize for this trade-off?

To answer this question, I first discuss measures for the simplicity and informativeness of quantifier languages. Secondly, I identify the trade-off at the level of whole languages. Then I introduce a measure of optimality with respect to the trade-off based on multi-objective optimization literature—a novel approach to evaluating the simplicity/informativeness trade-off. In the fourth and fifth section, I discuss my two approaches towards naturalness: sampling from natural categories of quantifiers and measuring the agreement of a language to semantic universals, respectively. In the sixth section I provide the setup of two experiments that investigate the relation between these measures of naturalness and optimality. The results are validated statistically, through a t-test and regression analysis, which is another novel contribution to the field. Afterwards, I provide some points of discussion and in the final section I conclude the chapter.

4.1 Simplicity and Informativeness for Languages

4.1.1 Simplicity

Like in the previous chapter, I define a complexity measure instead of a simplicity measure. This measure has two key desiderata:

1. More words means higher complexity

2. More complex words means higher complexity

These two desiderata are easily captured by a sum, normalized by the maximum amount of words:

$$\texttt{complexity}(L) = \sum_{\phi \in L} \texttt{complexity}(\phi)/\texttt{max\_words}$$

4.1.2 Informativeness

The measure for informativeness builds on the intuition introduced in the section 2.1—the idea that the informativeness of a language is the probability that said language can be used to communicate a random state of the world:

$$U(L) = \sum_{s \in S} \sum_{\phi \in L} \sum_{s' \in [\![\phi]\!]_S} P(s)P(\phi|s)P(s'|\phi) \cdot u(s, s')$$

The interpretation of u was assumed to be equality before, but in the context of whole languages that seems hardly accurate; very rarely do we communicate exact states of the world. More often are words used to transfer a meaning roughly. Jäger (2008) models this 'rough' communication in sim-max communication games (where sim-max stands for similarity-maximization). These games differ from 'exact' communication in that the utility is not just based on equality, but is inversely proportional to the distance (or dissimilarity) between the two states.

The interpretation of that inverse proportion used throughout this thesis is the following:

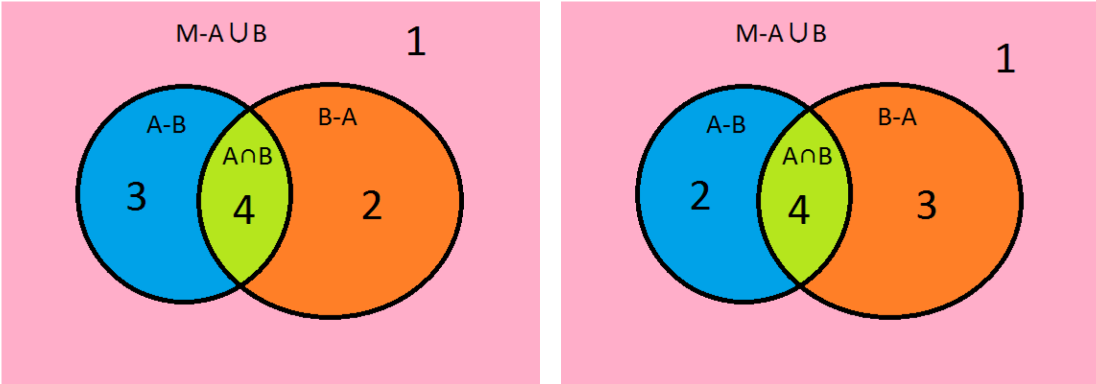
$$u(s, s') = \frac{1}{\texttt{dist}(s, s') + 1}$$

where

$$\texttt{dist}(s, s') = \sum_{X \in \{A-B, A\cap B, B-A, M-A\cup B\}} \max\{0, s_X - s'_X\}$$

Which can be interpreted as 'the amount of elements that have moved to a different set', measured as the amount that set numerosities have decreased (and not increased, since that would count every element twice).

As an example, take $s = (7, 6, 4)$ and $s' = (6, 7, 4)$ (recall that models are triplets $(n_A, n_B, n_{A\cap B})$). See these two models in Figure 4.1. Then $s_{A-B} = 7 - 4 = 3$, whereas $s'_{A-B} = 6 - 4 = 2$, so the "$A - B$" summand in the measure will be 1. So an element of $A - B$ has moved to one of the three other sets. In this case, $s_{B-A} - s'_{B-A} = -1$, so we can say that the element in question has 'moved' to $B - A$. However, we do not want to count it twice, so we only measure the



(a) s (b) s'

Figure 4.1: The models $s = (7, 6, 4)$ and $s' = (6, 7, 4)$ in terms of the four separate areas of relevance.

decrease in numerosities, not the increase. That way, $\texttt{dist}(s, s') = 1$, not 2 (the other two zones also contribute 0 to the measure, since they do not change).

The measure for informativeness that results captures three key aspects of informativeness:

1. Higher coverage of the meaning space increases informativeness.

2. Higher individual word specificity increases informativeness.

3. Overlapping words decrease increase informativeness (in part since the words would be more specific if they did not overlap).

4.2 Identifying the trade-off

To identify whether there is a trade-off, I sampled languages of 1–10 words, with 2000 samples per word amount. In Figure 4.2, see the complexity and informativeness of these languages. This figure implies that complexity and communicative cost trade off at the level of whole languages, forming a curved Pareto front as expected.

4.3 Measuring Optimality

In order to consider the influence of various factors on the optimality of languages, it is very useful to have a single measure of optimality. However, the very nature of trade-offs makes this a difficult task. Many measures exist (Tan, Lee, and Khor, 2002), but there is still no clear consensus as to the best possible measure.

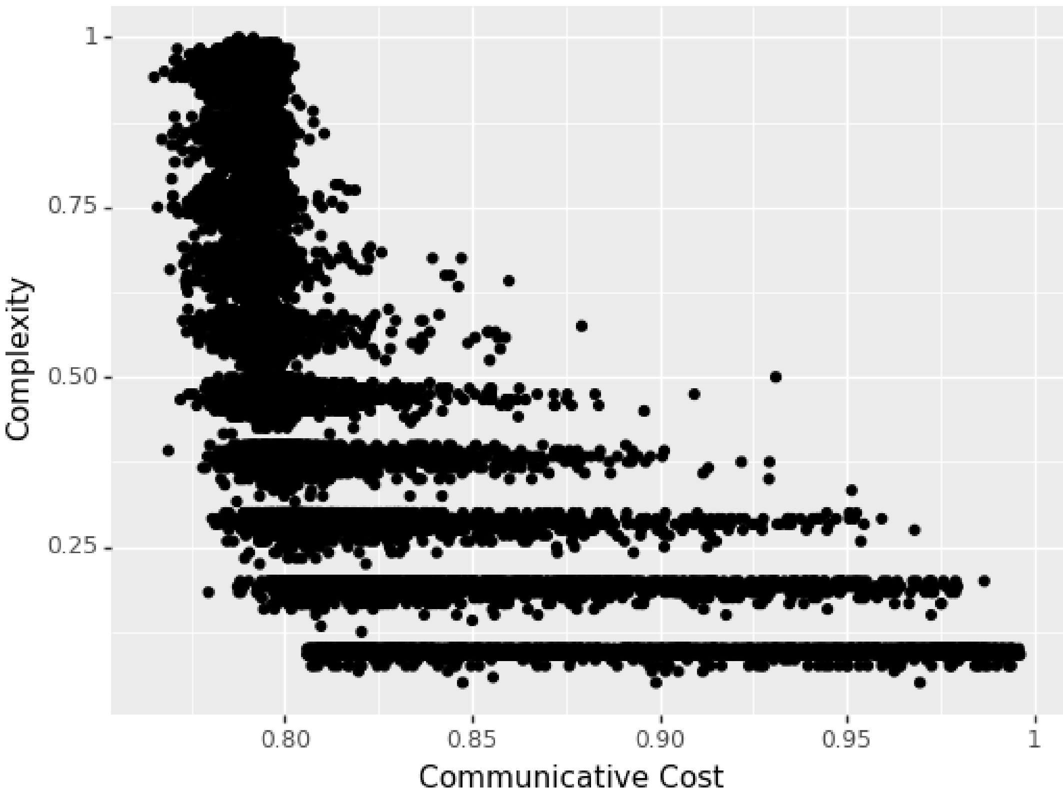


Figure 4.2: 2000 random languages for each word amount from 1 through 10.

The measure used here is one of the simplest, as mentioned by Deb and Jain (2002), the minimum Euclidian distance to the Pareto front:

$$\texttt{optimality(L)} = \min_{L^* \in P^*} \sqrt{\begin{array}{c}(\texttt{complexity}(L) - \texttt{complexity}(L^*))^2 + \\ (\texttt{communicative\_cost}(L) - \texttt{communicative\_cost}(L^*))^2\end{array}}$$

While many multi-objective optimization problems are concerned with both convergence (being as close to the frontier as possible) and diversity (covering as much of the frontier as possible), we are only concerned with the former. The minimum Euclidian distance is a simple and elegant way of capturing convergence, and operates on a single-point level, which allows for relation to other factors through statistical analysis.

4.3.1 Estimating the Pareto front

The measure mentioned requires knowledge of the *actual* Pareto front. Since the space of possible languages is too large to enumerate, it is impossible to know the exact Pareto front. Thus, we need an approximation of sorts. Thankfully, a slightly sub-optimal Pareto set can still function adequately for the optimality measure; being closer to the sub-optimal front also means being closer to the actual Pareto front.

I approximate the Pareto front as follows:

1. Find well-performing languages using a genetic algorithm.

2. Combine these languages with the sampled languages, and take the non-dominated ones.

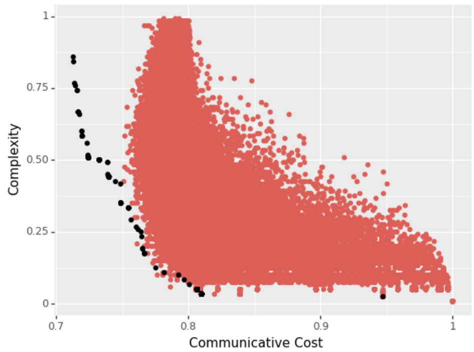
3. Fill up the gaps between the points on the frontier using linear interpolation.

Step 1: Find well-performing languages using a genetic algorithm Random sampling does not guarantee any kind of optimality among languages (especially if the density of languages close to the frontier is small). Instead, I use one of the standard ways to find good solutions to a multi-objective optimization problem: a genetic algorithm. Such algorithms emulate evolution in a sense; they generate an initial 'population' of solutions, pick the 'fittest' solutions, and let these fittest solutions 'reproduce', by generating various slightly changed versions of each solution. This process repeats for a predefined amount of generations or until some convergence criterion is reached. The specific algorithm used can be seen in Listing 4.1.

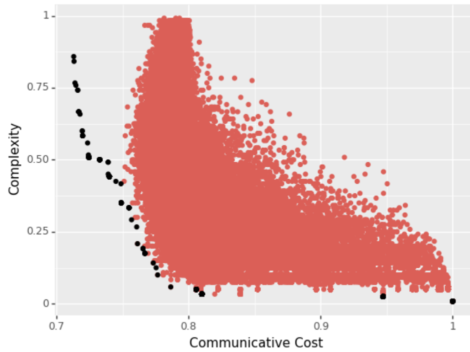
See the outcome of this algorithm compared to the results of random sampling in Figure 4.3a.

```
languages = sample_languages(2000)
for i in 1...num_generations:
    dominant_languages = find_dominant_languages(languages)
    offspring = sample_mutated(dominant_languages, 2000)
    languages = offspring

return languages

def mutate(language):
    mutated_language = language
    mutation_amount = choose_random([1,2,3])
    for i in 1...mutation_amount:
        mutation = choose_random(
                        [add_quantifier,
                        remove_quantifier,
                        interchange_quantifier])
        mutated_language = mutation(mutated_language)
    return mutated_language
```

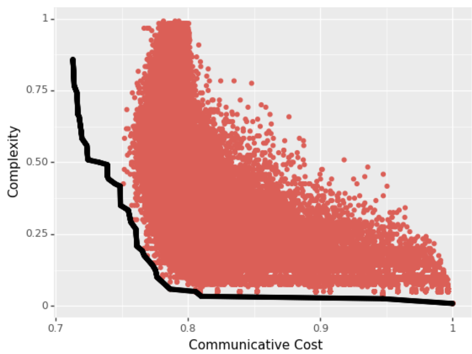
Listing 4.1: The evolutionary algorithm. Function `sample_mutated` generates 2000 mutated languages from the pool of dominant languages, giving each language an equal amount of offspring.



(a) Step 1: Find good languages by running an evolutionary algorithm.



(b) Step 2: Find the non-dominated points when combining the evolutionary outcome with the experiment outcome.



(c) Step 3: Apply linear interpolation.

Figure 4.3: The three steps in finding the approximate Pareto front. The red dots are the result of Experiment 2, as seen in Section 4.6. The black dots are the result of each respective step.

Step 2: combine with sampled languages and take the non-dominated points This results in the frontier seen in Figure 4.3b.

Step 3: Fill in the gaps using linear interpolation If the Pareto frontier consists of a small number of points with many gaps between them, a Euclidean distance measure seems hardly accurate. A point can then be very close to the frontier in theory, but simply because there is no point close to it on the frontier score badly on the metric. Thus, I interpolate the front linearly with a very small fixed distance of .001 (to approximate a line but avoid large computational overhead). This results in the Pareto front seen in figure 4.3c.

4.4 Naturalness of Languages

It is very difficult to find the right translation of natural languages to the representation language. Whereas research in other domains generally relies on big data sets which outline specifically which meanings have been lexicalized in which language (see e.g. Kemp and Regier, 2012), such data is not available when it concerns quantifiers. While the *Handbook* provides guidance, it does not provide the level of detail required for languages to be measured with respect to the trade-off.

As an alternative to translating languages, I use two approached: Firstly, as covered in this section, I define a broad set of quasi-natural quantifiers, and sample random quasi-natural languages with quantifiers from that set. In the next section I cover the other approach: measuring the degree to which logically possible languages agree to semantic universals.

As seen in section 2.2, the handbook identifies three main categories of quantifiers:

- Generalized Existential Quantifiers, such as 'no', which talk about $|A \cap B|$. This also includes numbers (e.g. $5 := |A \cap B| \geq 5$)
- Generalized Intersective Quantifiers, which talk about $|A - B|$. An example is 'all' ($|A - B| = 0$)
- Proportional quantifiers, talking about $|A - B|/|A|$, such as 'many'.

The set of quasi-natural quantifiers used is thus the following:

$$\mathcal{N} = \{|X \cap Y| * n, |X - Y| * n, |X \cap Y|/|X| * q$$

$$: X, Y \in \{A, B\}, X \neq Y, n \in \{0, 2, ..., 10\}, q \in \{0, .1, ..., 1\}, * \in \{=, >, <\}\}$$

All these quantifiers are shortened with respect to the grammar used (e.g. $|A - B| = 0$ is shortened to $A \subseteq B$ if '$\subseteq$' is in the grammar)

Using this set of quasi-natural quantifiers, I define the naturalness of a language as the proportion of words in the language that are quasi-natural:

$$\texttt{naturalness}(L) = \frac{|L \cap \mathcal{N}|}{|L|}$$

4.5 Semantic Universals

Along with naturalness as sampled from the quasi-natural set, we can also consider properties that natural languages tend to have—so-called *semantic universals*. (Szymanik, 2016) mentions four main semantic universals, as seen in Section 2.2. Since (ISOM) and (EXT) are already built into our model and grammar, there is only variance when it comes to (CONS) and (MON). Thus these are the two properties of languages that I measure. While these properties are boolean in principle, one can measure the degree to which a quantifier agrees to a semantic universal through the information-theoretic concept of *conditional entropy*, which measures how much knowledge one gains about a random variable given knowledge of another. This concept was first applied in research regarding quantifiers by Carcassi, Steinert-Threlkeld, and Szymanik (2019) to measure monotonicity. I extend this approach to the specific models used in the present research, and I extend it to conservativity.

4.5.1 Monotonicity

The monotoncity of a language is simply the mean of the monotoncities of the words in that language:

$$\texttt{monotonicity}(L) = \frac{\sum_{\phi \in L} \texttt{monotonicity}(\phi)}{|L|}$$

The monotonicity of each word is measured using an adjusted version of the gradual entropy-based measure introduced by Carcassi, Steinert-Threlkeld, and Szymanik (2019).

To define this measure, we first need a notion of submodel in terms of the model used in the present research. In terms of standard set-based models, the definition of B-submodel would be the following:

$$(M, A, B) \preceq_B (M, A', B') \text{ iff } A = A' \text{ and } B \subseteq B'$$

This translates to the present numerosity-based models as follows:

$$s \prec_B s' \text{ iff } s_A = s'_A \text{ and } s_{B-A} \leq s'_{B-A} \text{ and } s_{A\cap B} \leq s'_{A\cap B}$$

Resulting in a partial order. The definition of A-submodel is as expected:

$$s \preceq_A s' \text{ iff } s_B = s'_B \text{ and } s_{A-B} \leq s'_{A-B} \text{ and } s_{A\cap B} \leq s'_{A\cap B}$$

Using this definition of submodel, I measure the monotonicity of quantifiers based on how much information about the truth in a given model one gets based on whether there is truth in any submodel. This can be measured in information-theoretic terms using conditional entropy.

For e.g. right-upward monotonicity, this requires two random variables: 1_ϕ, the truth-value of ϕ given a model, and $1_\phi^{\preceq_B}$, whether or not a model has a B-submodel (or itself) where ϕ is true.

Monotonicity is then measured as follows, given the Shannon-entropy function H (Shannon, 1948):

$$\texttt{mon}{\uparrow}(\phi) = 1 - \frac{H(1_\phi | 1_\phi^{\preceq_B})}{H(1_\phi)}$$

$$\texttt{mon}{\downarrow}(\phi) = 1 - \frac{H(1_\phi | 1_\phi^{\succeq_B})}{H(1_\phi)}$$

$${\uparrow}\texttt{mon}(\phi) = 1 - \frac{H(1_\phi | 1_\phi^{\preceq_A})}{H(1_\phi)}$$

$${\downarrow}\texttt{mon}(\phi) = 1 - \frac{H(1_\phi | 1_\phi^{\succeq_A})}{H(1_\phi)}$$

These four measures are combined into one as follows:

$$\texttt{mon}_B(\phi) = \max(\texttt{mon}{\uparrow}(\phi), \texttt{mon}{\downarrow}(\phi))$$

$$\texttt{mon}_A(\phi) = \max({\uparrow}\texttt{mon}(\phi), {\downarrow}\texttt{mon}(\phi))$$

$$\texttt{monotonicity}(\phi) = \text{avg}(\texttt{mon}_A(\phi), \texttt{mon}_B(\phi))$$

4.5.2 Conservativity

I extend this line of thought to conservativity in much the same way, showcasing the flexibility of conditional entropy to gradually measure adherence to conditions such as the semantic universals. Instead of the partial order seen before, conservativity depends on a functional relation, the conservation of a model:

$$\text{con}_B(s) = (s_A, s_{A\cap B}, s_{A\cap B})$$

$$\text{con}_A(s) = (s_{A\cap B}, s_B, s_{A\cap B})$$

Note that conservativity is generalized here to work both in A and in B. The random variable $1_\phi^{\text{con}_B}$ models whether there is truth in the conservation of a model, and conservativity can be defined as:

$$\texttt{con}_B(\phi) = 1 - \frac{H(1_\phi | 1_\phi^{\text{con}_B})}{H(1_\phi)}$$

	Mean	SD
natural	0.0530	0.0276
random	0.0750	0.0262

Table 4.1: Mean and standard deviation of optimality for the two sampled sets.

$$\texttt{con}_A(\phi) = 1 - \frac{H(1_\phi | 1_\phi^{\text{con}_A})}{H(1_\phi)}$$

$$\texttt{conservativity}(\phi) = \max(\texttt{con}_A(\phi), \texttt{con}_B(\phi))$$

and the conservativity of a language is simply the mean of the conservativities of the words in the language:

$$\texttt{conservativity}(L) = \frac{\sum_{\phi \in L} \texttt{conservativity}(\phi)}{|L|}$$

4.6 Experiments

In this section, I will cover two experiments. The first concerns comparing random versus natural sampling, evaluating which leads to higher optimality. The second concerns sampling languages of various degrees of naturalness, and finding relations between optimality and the measures of naturalness, monotonicity and conservativity.

The code for the experiments can be found at https://github.com/wouterposdijk/SimInf_Quantifiers.

4.6.1 Experiment 1: Random vs. Natural sampling

In the first experiment, I have sampled 40000 languages, uniformly distributed over 1-10 words (4000 per word amount), 2000 per word amount from the set of quasi-natural quantifiers and 2000 per word amount from all generated quantifiers up to length 12, using the following operator set:

$$\mathcal{O} = \{|\cdot|, \subset, \cap, \cup, -_s, -_n, +, >_n, \geq_n, >_q, =_n, =_q, /, \wedge, \vee, \neg, \%\}$$

Figure 4.4 showcases the results. This figure gives the impression that the quasi-natural languages perform much better regarding the trade-off than the completely random languages. The mean optimality and standard deviation can be seen in Table 4.1.

Their means are significantly different ($p \approx 0.000, t(37271) = -79.82$). This implies that natural languages are, in general, better optimized with respect to the simplicity/informativeness trade-off than completely random languages.

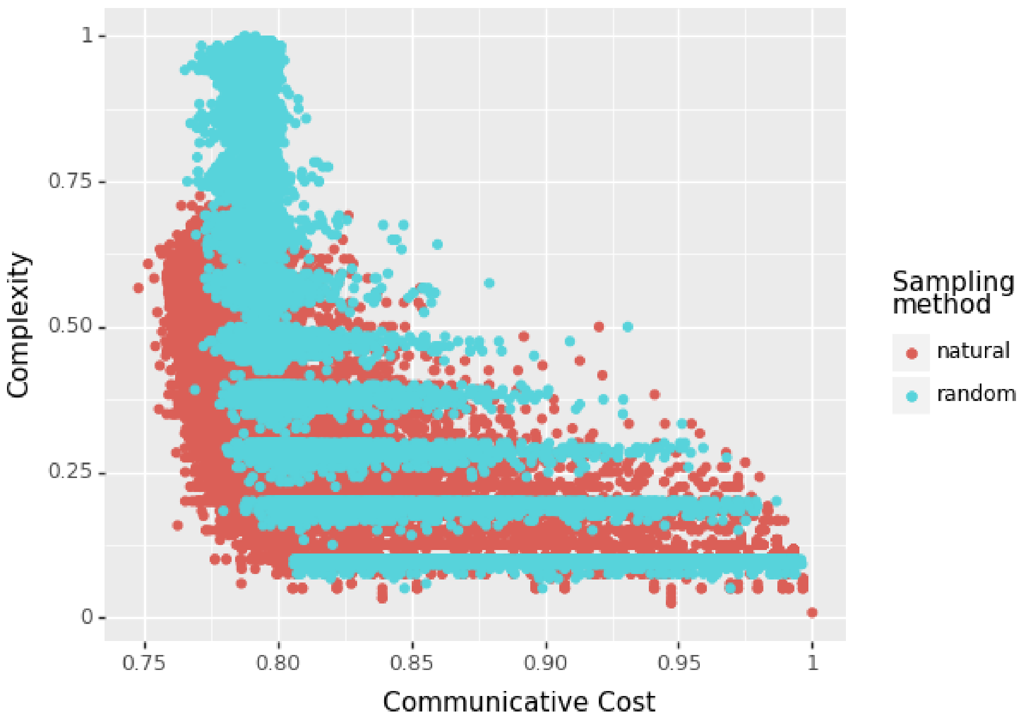


Figure 4.4: The communicative cost and complexity of the languages sampled in Experiment 1.

Language	**Complexity**	**Comm. cost**
$\lvert A\cap B\rvert/\lvert A\rvert > .4, \lvert A\rvert/\lvert B\rvert = .7, \lvert A\cap B\rvert/\lvert A\rvert < .7, \lvert A\cap B\rvert > \lvert B-A\rvert, \lvert A\cap B\rvert/\lvert A\rvert > .3, \lvert B-A\rvert > 2, \lvert A\cap B\rvert > 6, \lvert A-B\rvert < 10, \lvert A\cap B\rvert/\lvert A\rvert > .1, \lvert A\cap B\rvert = 6$	0.566	0.748
$\lvert A\cap B\rvert/\lvert B\rvert = .6, \lvert A\cap B\rvert/\lvert A\rvert > .7, \lvert A\cap B\rvert/\lvert A\rvert < .8, \lvert A-B\rvert = 4, \lvert B-A\rvert = 10, \lvert A\cap B\rvert = 6$	.35	.758
$\lvert A-B\rvert = 8, \lvert A-B\rvert < 8, \lvert A\cap B\rvert > 6$	.15	.773

Table 4.2: Some examples of dominating languages from the sampled quasi-natural languages.

Some examples of dominating languages from the sampled quasi-natural languages can be seen in Table 4.2.

4.6.2 Experiment 2: Gradual naturalness

In the second experiment, I have sampled 80000 languages, uniformly distributed over 1-10 words (8000 per word amount), with the amount of natural quantifiers sampled from a uniform distribution; that is to say, when sampling a language of e.g. 7 words, the amount of quantifiers that come from the quasi-natural set is sampled from a uniform distribution over 0 through 7.

I have measured all six properties mentioned in this chapter for all these languages:

- Complexity
- Communicative Cost
- Optimality
- Naturalness
- Monotonicity
- Conservativity

See the complexity and communicative cost of these languages in Figure 4.5.

In Figure 4.6, see the average naturalness, monotonicity and conservativity in various areas of the graph. The figures imply that monotonicity and naturalness are major factors pushing toward the frontier, and conservativity is not. Interestingly, high monotonicity mainly seems to be prevalent at the bottom front, whereas high naturalness seems prevalent both bottom and left.

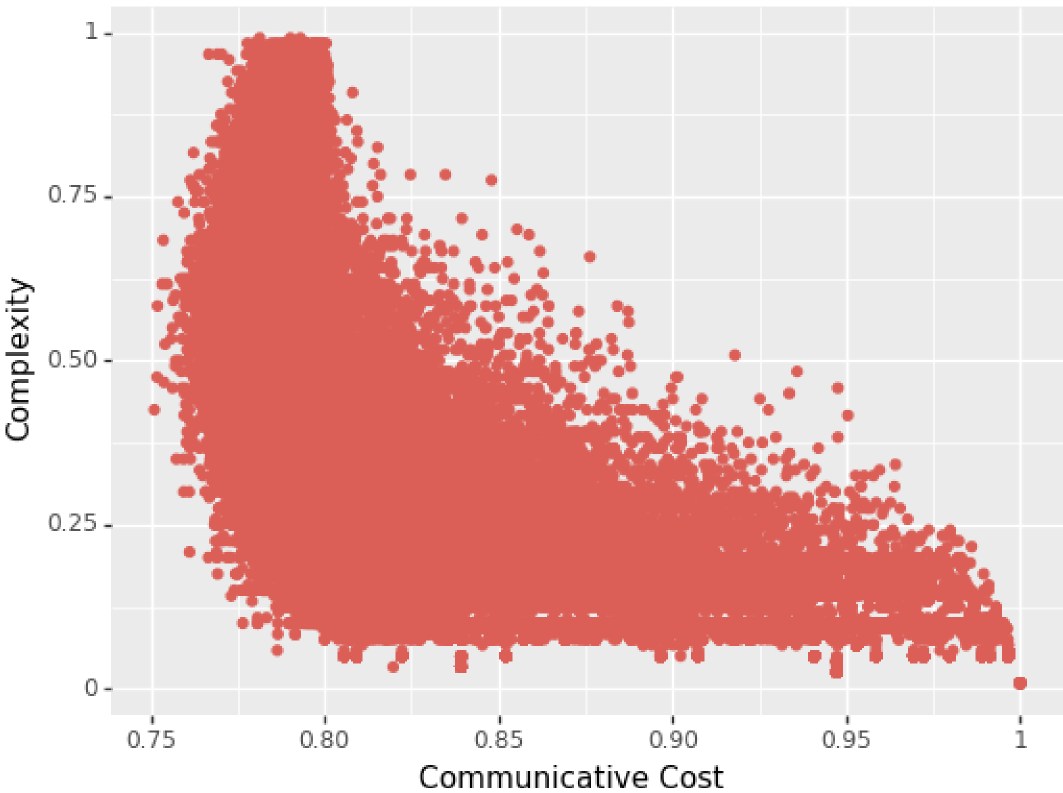
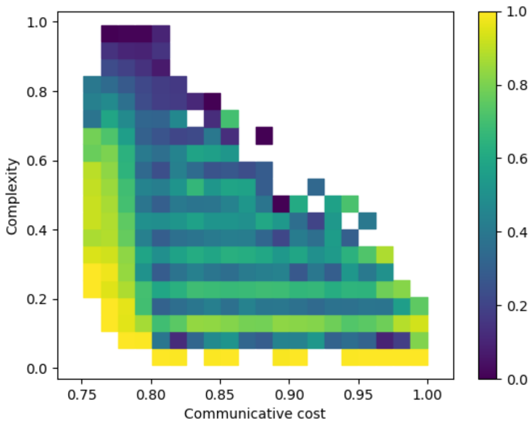
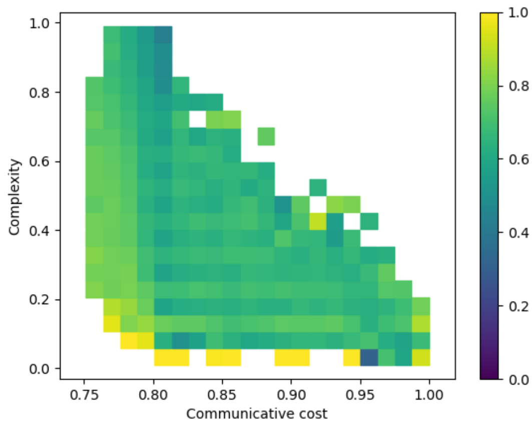


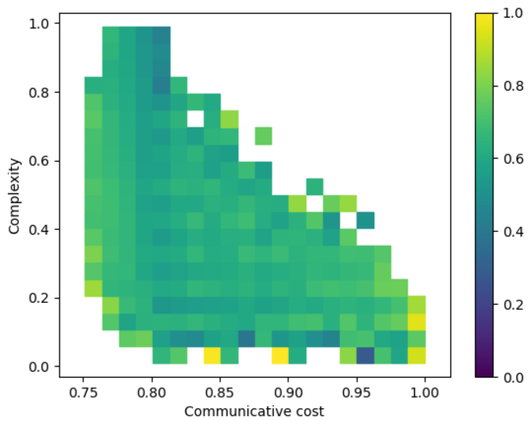
Figure 4.5: The communicative cost and complexity of the languages sampled in Experiment 2.



(a) Naturalness



(b) Monotonicity



(c) Conservativity

Figure 4.6: The average values of naturalness, monotonicity and conservativity in various ranges of complexity and communicative cost.

factor	coef.	p	ΔR^2
Intercept	.0828	0	
naturalness	-.0168	.0000	.040
monotonicity	-.0270	.0000	.013
conservativity	.0106	.0000	.005

Table 4.3: The results of a linear regression predicting optimality based on naturalness, monotonicity and conservativity in the sampled set of languages for Experiment 2. The ΔR^2 indicates how much R^2 increases upon adding that factor to a model with only the other two. Total R^2 of the model is .103

A proper way to verify the influence of these factors on optimality is through linear regression. See in Table 4.3 the effects of naturalness, monotonicity and conservativity on optimality. Naturalness and monotonicity have significant negative effects on the optimality, implying that they make languages more optimal. See in the rightmost column the increment in R^2 that adding each parameter to the model yields. This shows that, while monotonicity has a larger coefficient than naturalness, naturalness explains a larger portion of the variance.

4.7 Discussion

The results seem to imply that both quasi-naturalness and monotonicity predict better optimality than complete randomness. This implies that natural languages are indeed a relative optimization with respect to this trade-off. Indeed, the hypothesis testing done confirms this idea. In this section, I provide some points of discussion: strengths and weaknesses of the measures used and experiments shown in this chapter.

More sets of operators For this experiment, only a single set of operators was used. Using various sets of operators could provide a more complete picture. There are two possible problems with using only a single grammar. On the one hand, there is the desideratum of *expressivity*: the grammar needs to be able to express an adequate range of quantifiers to be an appropriate arena for evaluating the trade-off. On the other hand, there is the desideratum of *cognitive representativeness*.

Expressivity is not a problem; the setup used leaves space for adequate variance in communicative cost, complexity, monotonicity, conservativity and quasi-naturalness. It is cognitive representativeness and the consequent measure of complexity used that may form a problem—for example, one may wonder whether it is right to incur the same complexity for use of % as for use of $\cap$. For that reason, it would be good to see more verifications of the research results with different sets of operators, or different inherent complexities among the operators used.

Relatively close does not mean optimal Natural quantifiers lie *closer* to the Pareto front than completely random quantifiers. This does not imply that natural languages are *optimal* with respect to the trade-off, but it does imply that languages are *optimized*. There may be other factors at play, but natural language are certainly more optimized than completely random languages.

Simplicity is easier to achieve randomly than informativeness Many sampled languages, natural or random, are relatively uninformative and very simple. This can be attributed to the fact that languages were sampled uniformly across word count—supplying many languages of size 1 and 2, which are otherwise uncommon due to the combinatorial explosion of options as language size increases. This leads to a rather large area, rather close to the frontier, being populated by both quasi-natural and random languages. This might point to natural languages in general requiring a minimum level of informativeness.

Only quasi-natural languages While the examples of non-dominated quasi-natural languages seen in Table 4.2 look natural in principle, they are unlikely to match any languages actually attested in the world. However, this is not really problematic: the specific languages that end up the best in this mere approximation of real life interaction were never expected to be real natural languages. Indeed, since all our measures are mere approximations, it is sensible to stay focused on the general performance of general properties with respect to the trade-off, as was done in the experiments.

Conservativity is not related to optimality Although conservativity, one of the major semantic universals, does not make languages more optimal, this is not problematic for the hypothesis. If some aspects of natural languages improve the trade-off, that is already a strong enough result. Languages may be conservative for some other reason than optimizing the simplicity/informativeness trade-off. Steinert-Threlkeld and Szymanik (2018) have found a similar discrepancy between the results of conservativity and other universals in their research on the learnability of quantifiers. The current research further strengthens the idea that conservativity is of a very different nature than monotonicity.

4.8 Conclusion

It seems that out of the three properties of natural languages considered, two, namely naturalness and monotonicity, optimize with respect to the simplicity/informativeness trade-off. These results point towards a pressure for languages to optimize with respect to the simplicity/informativeness trade-off in the semantic domain of quantifiers.

Chapter 5

Future Directions

In this chapter, I discuss possible future directions that may build upon the present research. First, I discuss which computational restrictions applied in the present research may be alleviated. Second, I discuss how analysis of dynamics can provide more insight in the process at which languages optimize for the trade-off. Third, I discuss how the model may be extended to include pragmatics. Finally, I discuss how learnability may provide an alternative to the complexity measures used in the present research.

5.1 Computational Restriction

Perhaps the most obvious way to build upon the present research is alleviating some of the computational restrictions, such as the fact that the integers allowed in the grammar were limited, and that some semantic universals were built into the model and grammar.

All integers For computational reasons, the grammars used contained only even integers. Although I believe that the results attested in the experiments are representative for a scenario where all possible integers are considered, it would be good to verify that scenario using better computational means.

Extensionality and isomorphism invariance Another computational restriction built in the semantic universals extensionality and isomorphism invariance. This means that the results attested here measure the optimality of naturalness, monotonicity and conservativity *given* (EXT) and (ISOM). It remains to be seen whether this effect holds if these two factors are removed. In addition, these two factors, being semantic universals, are predicted to provide languages with relative optimality as well—although (EXT) is quite similar to conservativity, in that it considers a part of the model irrelevant, an may thus not provide much of an effect. This is something to be investigated. I suspect that gradual entropy-based measures can be developed for both (EXT) and (ISOM) with relative ease.

5.2 Dynamics

Kemp, Xu, and Regier (2018) mention the need for a more dynamic approach in the area of research: once one finds out that natural languages are optimized with respect to the simplicity/informativeness trade-off, a good next step is to figure out how that may have come to be. What are evolutionary paths from non-optimized languages to optimized languages? And do such paths ends up at natural languages per se? Such questions can be investigated by considering evolutionary algorithms, such as the one used to estimate the Pareto front, in more detail. For example, one may build in mutations that focus on naturalness, or one may compare convergence speed between the algorithm run with arbitrary sampling vs. natural sampling. Such approaches can provide more insight into the process that leads to the optimization witnessed.

5.3 Pragmatics

Although the results are already promising, one very important part of communication has not yet been included: pragmatics (Grice, Cole, and Morgan, 1975). Since pragmatics constitute a sizeable part of communication, and the measures for informativeness are based on the idea of efficient communication, incorporating pragmatics could greatly increase the accuracy of the measures. There are several forms of pragmatics that could be incorporated.

For one, scalar implicature (Carston, 1998) can be incorporated. Franke (2018) has modeled scalar implicature in quantifier languages using the Rational Speech Act framework (Goodman and Frank, 2016). I believe that this could be extended to the current model with relative ease. Scalar implicature can be expected to improve the informativeness of languages with overlap, since it effectively counteracts overlap. This might be more advantageous for natural languages than for random languages, which would further strengthen the results.

In addition, the maxim of quantity (the idea that one should not provide more information than needed) can be incorporated somehow—considering that oftentimes context dictates that people purposefully convey vague information, since more detail would not be relevant. A possible way to incorporate this would be to consider the informativeness of a language given various amounts of sim-max payoff. Alternatively, this can be incorporated by including a Question Under Discussion (Beaver et al., 2017) in the game, which must be answered with as little information as possible.

In short, pragmatics provide exciting extensions for the models used so far; it would be very interesting to see whether or not they strengthen the advantage natural languages have over random languages.

5.4 Learnability

Aside from the simlicity/informativeness trade-off as seen in this thesis, another possible factor in the shaping of the semantic typology of quantifiers is learnability: the idea that our languages are optimized to be easy to learn (note that this interpretation is rather recent and not to be confused with learnability in Universal Grammar (Pinker, 2009)). Steinert-Threlkeld and Szymanik (2018) have used this principle to explain the prevalence of several semantic universals among quantifiers.

Learnability is very related to simplicity (Gibson et al., 2019). Learnability implies simplicity, just as simplicity implies learnability. Thus, frameworks for measuring the learnability of a word or language may be used to measure the simplicity thereof. Steinert-Threlkeld and Szymanik (2018) use neural networks to investigate the learnability of quantifiers—this could lead to a measure of complexity therefor. Such an approach can likely be extended to whole languages as well. An obvious advantage of such an approach is that it does not depend on a Language of Thought and thus does not depend on the specific choice of operators. There are, of course, choices to be made in the form of the learning model used, but those choices are less domain-specific, allowing them to be motivated from general neuro- and cognitive science.

Thus, learnability provides a way to measure simplicity with fewer direct assumptions, providing exciting next steps in simplcity/informativeness research.

Chapter 6

Conclusion

In this thesis, I have developed measures of simplicity and informativeness for both individual quantifiers and quantifier languages, despite the fact that quantifiers are *function words*. This shows that the general approach applied to content words extends to function words as well. In addition, I have developed various ways to approach properties of natural quantifier languages, through either considering categories of words, or by measuring semantic universals gradually. By developing a measure of the optimality of languages, I have been able to relate said optimality to properties of natural languages through statistical analysis, providing a good view of the influence of the trade-off.

While there does not seem to be a trade-off at the level of individual quantifiers, a trade-off can certainly be identified at the level of quantifier languages. Statistical tests verify that monotonicity and naturalness improve the optimality of languages, implying that there are properties of natural languages that make them more optimized with respect to the trade-off than arbitrary, logically possible ones. This supports the hypothesis that the trade-off is an important factor in shaping the semantic typology of quantifiers. This is despite the fact that quantifiers are function words, strengthening the hypothesis that the trade-off shapes the semantic typology of *all* semantic domains.

Still, much can be done to strengthen this result further, by alleviating computational restrictions, analyzing dynamics, and incorporating pragmatics and learnability, not to mention extension to other domains of function words. These are all exciting directions sure to increase the knowledge about the nature and influence of the simplicity/informativeness trade-off.

Bibliography

Baker, Mark Cleland (2003). *Lexical categories: Verbs, nouns and adjectives.* Vol. 102. Cambridge University Press.

Barwise, Jon and Robin Cooper (1981). "Generalized quantifiers and natural language". In: *Philosophy, language, and artificial intelligence.* Springer, pp. 241–301.

Beaver, David Ian (1997). "Presupposition". In: *Handbook of logic and language.* Elsevier, pp. 939–1008.

Beaver, David Ian et al. (2017). "Questions under discussion: Where information structure meets projective content". In: *Annual Review of Linguistics* 3, pp. 265–284.

Carcassi, Fausto, Shane Steinert-Threlkeld, and Jakub Szymanik (2019). "The emergence of monotone quantifiers via iterated learning". In:

Carr, Jon W et al. (2018). "Conceptual structure is shaped by competing pressures for simplicity and informativeness". In:

Carston, Robyn (1998). "Informativeness, relevance and scalar implicature". In: *Pragmatics And Beyond New Series*, pp. 179–238.

Deb, Kalyanmoy (2014). "Multi-objective optimization". In: *Search methodologies.* Springer, pp. 403–449.

Deb, Kalyanmoy and Sachin Jain (2002). "Running performance metrics for evolutionary multi-objective optimization". In:

Evans, Nicholas (2010). "Semantic Typology". In: *The Oxford Handbook of Linguistic Typology.* Chap. 23.

Franke, Michael (2018). "Quantifier choice & approximate number". In: *Probabilistic language understanding: An introduction to the Rational Speech Act framework.*

Gibson, Edward et al. (2019). "How Efficiency Shapes Human Language, TICS 2019". In:

Goodman, Noah D and Michael C Frank (2016). “Pragmatic language interpretation as probabilistic inference”. In: *Trends in cognitive sciences* 20.11, pp. 818–829.

Grice, H Paul, Peter Cole, Jerry L Morgan, et al. (1975). “Logic and conversation”. In: *1975*, pp. 41–58.

Huttegger, Simon M and Kevin JS Zollman (2011). “Signaling games”. In: *Language, games, and evolution*. Springer, pp. 160–176.

Jäger, Gerhard (2008). “Applications of game theory in linguistics”. In: *Language and Linguistics compass* 2.3, pp. 406–421.

Keenan, Edward and Denis Paperno (2012). *Handbook of quantifiers in natural language*. Vol. 90. Springer.

Kemp, Charles and Terry Regier (2012). “Kinship categories across languages reflect general communicative principles”. In: *Science* 336.6084, pp. 1049–1054.

Kemp, Charles, Yang Xu, and Terry Regier (2018). “Semantic typology and efficient communication”. In: *Annual Review of Linguistics* 4, pp. 109–128.

Ouhalla, Jamal (2003). *Functional categories and parametric variation*. Routledge.

Piantadosi, Steven T, N Goodman, and J Tenenbaum (2013). “Modeling the acquisition of quantifier semantics: a case study in function word learnability”. In: *Under review*.

Piantadosi, Steven T and Robert A Jacobs (2016). “Four problems solved by the probabilistic language of thought”. In: *Current Directions in Psychological Science* 25.1, pp. 54–59.

Pinker, Steven (2009). *Language learnability and language development, with new commentary by the author*. Vol. 7. Harvard University Press.

Scott-Phillips, Thom (2014). *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*. Macmillan International Higher Education.

Shannon, Claude Elwood (1948). “A mathematical theory of communication”. In: *Bell system technical journal* 27.3, pp. 379–423.

Stalnaker, Robert (1978). “Assertion”. In: *Syntax and Semantics (New York Academic Press)* 9, pp. 315–332.

Steinert-Threlkeld, Shane and Jakub Szymanik (2018). “Learnability and semantic universals”. In: *Semantics and Pragmatics*.

Szymanik, Jakub et al. (2016). *Quantifiers and cognition: Logical and computational perspectives*. Vol. 96. Springer.

Tan, Kay Chen, Tong Heng Lee, and Eik Fun Khor (2002). “Evolutionary algorithms for multi-objective optimization: Performance assessments and comparisons”. In: *Artificial intelligence review* 17.4, pp. 251–290.

Tomasello, Michael (2009). *The cultural origins of human cognition*. Harvard university press.

Van Benthem, Johan et al. (1986). *Essays in logical semantics*. Springer.