Quantifiers, Complexity, and Degrees of Universals

A Large-Scale Analysis

Nima Motamed

October 8, 2020

CoSaQ Seminar

Introduction

Goal

We wish to find the reason underlying the appearance of certain semantic universals throughout natural language.

Previous work

One proposal posited in the literature: the learnability hypothesis.

Steinert-Threlkeld and Szymanik (2019)[1] have found evidence supporting the learnability hypothesis in the domain of quantifier expressions: neural networks find it easier to learn quantifiers satisfying certain semantic universals.

[1] S. Steinert-Threlkeld and J. Szymanik (2019). "Learnability and semantic universals". In: *Semantics and Pragmatics* 12 (4).

Previous work

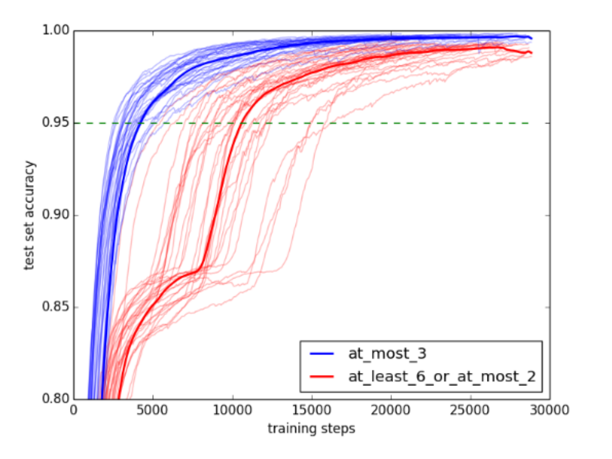

Figure 1: Learning curves from Steinert-Threlkeld and Szymanik (2019)

Previous work

Building upon these findings, van de Pol, Steinert-Threlkeld, and Szymanik (2019)[2] (henceforth called vdP&ST&S) have found evidence for the hypothesis that the presence of some semantic universals can also be explained by differences in complexity among quantifiers.

[2] I. van de Pol, S. Steinert-Threlkeld, and J. Szymanik (2019). "Complexity and learnability in the explanation of semantic universals". In: *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, pp. 3015–3021.

Previous work

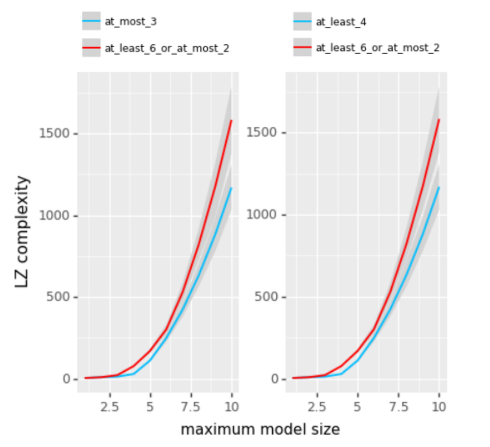

Figure 2: Complexity curves from vdP&ST&S

Previous work

Though their results are promising, both of their approaches lack generality due to their use of the minimal pair methodology.

We should upscale their approach!

In this talk, I will give an overview of my findings on the explanatory power of complexity for the presence of semantic universals, using an upscaled version of the approach taken by vdP&ST&S.[3]

[3] Along with some theoretical analysis towards the end!

Upscaling

This upscaling is done in two ways.

First, I do away with the minimal pair methodology, and instead measure the complexity of a large variety of logically possible quantifiers.

Second, I use the more general notion of the degree to which a universal is satisfied by a quantifier (as introduced by Carcassi, Steinert-Threlkeld, and Szymanik (2019)[4]), instead of the standard binary notion of satisfaction.

[4] F. Carcassi, S. Steinert-Threlkeld, and J. Szymanik (2019). "The emergence of monotone quantifiers via iterated learning". In: *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, pp. 190–196.

Let us first recap briefly on the relevant definitions.

Quantifiers & universals

Determiners

We consider quantifiers to be the semantic objects expressed by determiners, i.e. binary relations between sets of objects.

Generalised quantifiers

This is captured within the framework of generalised quantifier theory[5] as stating that determiners are type $\langle 1, 1 \rangle$ generalised quantifiers.

Definition

A type $\langle 1, 1 \rangle$ generalised quantifier Q is a set consisting of models $\mathbb{M} = \langle M, A, B \rangle$, where A, B and M are sets such that $A, B \subseteq M \neq \emptyset$.

From now on, we just refer to these as quantifiers.

[5] S. Peters and D. Westerståhl (2006). *Quantifiers in Language and Logic*. Oxford: Clarendon Press. ISBN: 9780 199291250.

Generalised quantifiers

Given a model $\mathbb{M} = \langle M, A, B \rangle$, we refer to M, A and B as the domain of discourse, restrictor, and scope of $\mathbb{M}$, respectively.

If $\mathbb{M} \in Q$, we say that within the domain of discourse M, the quantifier represented by Q applied to the restrictor A and scope B is satisfied, and we write $Q_M(A, B) = 1$. Similarly: $Q_M(A, B) = 0$ if $\mathbb{M} \notin Q$.

Extensionality

As an example, the representation of the natural language determiner ‘all’ as a formal quantifier is

$$all = \{\langle M, A, B\rangle \, ; \, A \subseteq B\}.$$

The definition of this quantifier, like that of any in natural language, does not contain any reference to M - it is extensional.

Definition

A quantifier Q is called extensional if for all sets $A, B \subseteq M \subseteq M'$, it holds that $Q_M(A, B) = Q_{M'}(A, B)$.

Semantic universals

There are three (categories of) semantic universals which we consider: monotonicity, quantitavitiy, and conservativity.

Monotonicity

Definition

A quantifier Q is called upward (right) monotone if for any sets $A \subseteq M$ and $B \subseteq B' \subseteq M$, it holds that $Q_M(A, B) \leqslant Q_M(A, B')$. Similarly, Q is called downward (right) monotone if for any $A \subseteq M$ and $B' \subseteq B \subseteq M$, it holds that $Q_M(A, B) \leqslant Q_M(A, B')$.

Most is upward monotone, while *an even number of* is not monotone at all.

Quantitativity

Definition

A quantifier Q is called quantitative if whenever we have $|A \cap B| = |A' \cap B'|$, $|A \setminus B| = |A' \setminus B'|$, $|B \setminus A| = |B' \setminus A'|$, and $|M \setminus (A \cup B)| = |M' \setminus (A' \cup B')|$, then $Q_M(A, B) = Q_{M'}(A', B')$.

Some is quantitative, while *the first three* is not.

CONSERVATIVITY

Definition

A quantifier Q is called (left) conservative if it always holds that $Q_M(A, B) = Q_M(A, A \cap B)$.

All is conservative, while the hypothetical determiner *EQ*, expressing that the restrictor and scope are equal in size, is not.

Conservativity is interesting: it is difficult to even express complex non-conservative determiners in natural language.

Degrees

Fine-grained distinctions

We have a binary notion of satisfaction for universals. This does not allow us to distinguish between quantifiers that do not satisfy a universal, even though intuitively, e.g. *at least three* satisfies monotonicity to a higher degree than *an even number of*.

Information-theoretical degrees

Carcassi, Steinert-Threlkeld, and Szymanik (2019) and Posdijk (2019)[6] have defined the notion of the degree to which a quantifier satisfies a universal. This definition is based on information theory, and (informally!) boils down to the normalised mutual information between the quantifier and the universal.

[6] W. Posdijk (2019). "The influence of the simplicity / informativeness trade-off on the sematic typology of quantifiers". Master's thesis. Universiteit van Amsterdam.

Formal definition

To facilitate theoretical analysis, we give a general and formal definition of the degree to which a quantifier is explained by some measure on models. The degrees of universals are specific instances of this definition.

Let us work through the preliminaries.

Models as strings

Assuming that we only consider finite models with objects from some countably infinite universe $U = \{o_i\,;\, i \in \omega\}$, there is a natural correspondence between models and quaternary strings (i.e. strings $\alpha \in 4^+$), since any $x \in M$ must be in exactly one of the sets $A \cap B$, $A \setminus B$, $B \setminus A$, and $M \setminus (A \cup B)$.

Quantifiers as random variables

Some notation: let $\mathcal{M}$ be the class of all models, and $\mathcal{M}_{\leqslant n}$ of those of size up to n.

For each n, we can place a uniform probability distribution over $\mathcal{M}_{\leqslant n}$. This allow us to view quantifiers Q as random variables $\mathbb{1}_{Q,n} : \mathcal{M}_{\leqslant n} \to 2$.

Degrees of explanation

We can now define the degrees!

Definition

Given some n, a quantifier Q, and some measure $X : \mathcal{M} \to \mathcal{X}$, the n-th degree of explanation of Q by X is defined as

$$\deg_n^X(Q) := 1 - \frac{H(\mathbb{1}_{Q,n} \mid X_n)}{H(\mathbb{1}_{Q,n})},$$

where H is the (conditional) Shannon entropy, and X_n is defined as the random variable obtained by restricting X to $\mathcal{M}_{\leqslant n}$.

Degrees of universals

Using this general definition, we can define degrees of universals by finding measures corresponding to the universals. For (upward right) monotonicity, quantitativity and (left) conservativity, respectively, we define the following measures on a model $\mathbb{M} = \langle M, A, B \rangle$:

- for each quantifier Q, a binary measure $\mathbb{1}_Q^{\nearrow}$ defined as $\mathbb{1}_Q^{\nearrow}(\mathbb{M}) = 1$ iff there is some $\mathbb{M}' = \langle M, A, B' \rangle$ with $B' \subseteq B$ and $Q_M(A, B') = 1$
- a measure $\#$ defined as $\#(\mathbb{M}) = \langle |A \cap B|, |A \setminus B|, |B \setminus A|, |M \setminus (A \cup B)| \rangle$
- a measure $\upharpoonright^{\leftarrow}$ defined as $\upharpoonright^{\leftarrow}(\mathbb{M}) = \langle M, A, A \cap B \rangle$

Degrees of universals

It is easily verified that the degree for some universal is equal to 1 for a quantifier if and only if that quantifier satisfies the universal. So this works.

Or does it? Note that degrees are parameterised by the maximum model size n. Do degrees stabilise as n increases? We will consider this question again towards the end.

Complexity

Kolmogorov complexity

The notion of complexity used by vdP&ST&S is that of (approximate) Kolmogorov complexity, from the field of algorithmic information theory.

This measures how well some sequence of symbols can be compressed by exploiting patterns and structures in the sequence. More complexity equals less structure.

Quantifiers as streams

To be able to apply this to quantifiers, we need another representation for them.

Using the correspondence between models and strings, we obtain a natural correspondence between quantifiers and infinite binary streams (i.e. streams $\beta \in 2^\omega$), since we can lexicographically order models.

Kolmogorov complexity

Since we can only really compute the approximate Kolmogorov complexity of finite strings, we need to work with finite parts of a quantifier's binary stream.

We define the n-th complexity value of a quantifier Q to be the average approximate Kolmogorov complexity[7] of the first $|\mathcal{M}_{\leqslant n}|$ bits in each of its binary streams (one for each possible lexicographical ordering of the set 4)

[7] We use the Lempel-Ziv algorithm by Lempel and Ziv (1976) to compute this.

Methods

Language of thought

As stated in the introduction, our goal is to measure degrees and complexity for a wide variety of logically possible quantifiers. We do this by way of a logical grammar producing quantifiers (c.f. the language of thought used by Piantadosi, Tenenbaum, and Goodman (2012)[8]).

Taking computational and practical concerns into consideration, we only produce extensional quantifiers.

[8] S. T. Piantadosi, J. B. Tenenbaum, and N. D. Goodman (2012). "Modeling the acquisition of quantifier semantics: a case study in function word learnability".

Language of thought

$$
\begin{aligned}
START &\rightarrow \lambda \text{ a b . } BOOL \\
BOOL &\rightarrow (SET = \emptyset) \mid (SET \neq \emptyset) \\
&\mid (SET \subseteq SET) \mid (SET \nsubseteq SET) \\
&\mid (SET \subset SET) \mid (SET \not\subset SET) \\
&\mid (\text{card}(SET) \text{ is even}) \mid (\text{card}(SET) \text{ is odd}) \\
&\mid (\text{card}(SET) = \text{card}(SET)) \\
&\mid (\text{card}(SET) \neq \text{card}(SET)) \\
&\mid (\text{card}(SET) \geqslant \text{card}(SET)) \\
&\mid (\text{card}(SET) > \text{card}(SET)) \\
&\mid (\text{card}(SET) = n) \mid (\text{card}(SET) \neq n) \\
&\mid (\text{card}(SET) \geqslant n) \mid (\text{card}(SET) \leqslant n) \\
&\mid (BOOL \text{ and } BOOL) \mid (BOOL \text{ or } BOOL) \\
SET &\rightarrow ORDER \mid (SET \setminus SET) \\
&\mid (SET \cap SET) \mid (SET \cup SET) \\
ORDER &\rightarrow \text{a} \mid \text{b} \mid (\text{first } n \text{ of } ORDER) \mid (\text{last } n \text{ of } ORDER)
\end{aligned}
$$

Figure 3: Non-terminals are colored, and n ranges over the positive integers.

Primitives

This grammar contains a lot of primitive notions, and does not explicitly contain negation. This is due to computational considerations: we want the grammar to produce a large amount of semantically distinct and interesting quantifiers, within reasonable time.

Quantifier generation

Using our grammar (with the variable n ranging from 1 to 10), we generate quantifiers by considering all productions at a maximum depth of 6. By comparing the quantifiers' binary streams, we can ensure that we only consider semantically unique quantifiers.

This process gives us 8044 quantifiers.

Statistic computation

For each quantifier, we then compute the 10-th degrees of monotonicity, quantitativity and conservativity, along with the 10-th complexity value.

Statistic distributions

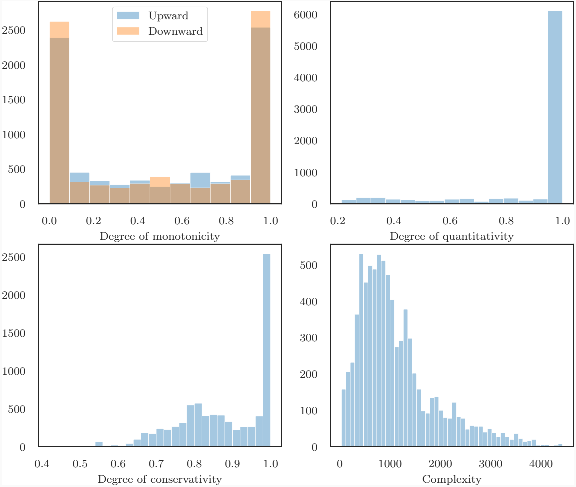

Figure 4: Distributions of the computed statistics.

Results

Correlations

To determine how well complexity explains a universal, we perform two types of correlation analysis with complexity as the independent and the degree of the universal as the dependent variable. We also perform these analyses for some statistics derived from the degrees.

We compute both R^2 and Kendall's τ[9], along with 95% confidence intervals for both.[10]

[9] This is the version described by Kendall (1945)

[10] These confidence intervals were obtained through non-parametric bootstrapping.

Linear regression

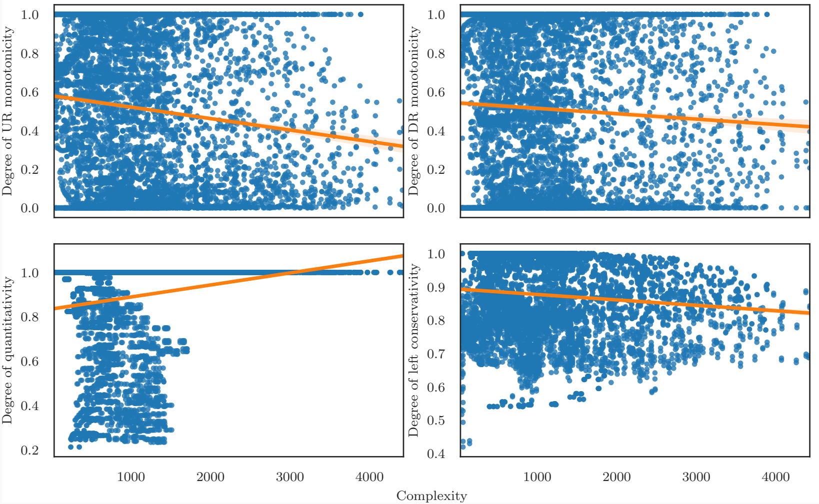

Figure 5: Linear regression fits for the basic statistics.

R^2 AND τ

Table 1: The values of R^2 and τ for a selection of the statistics. The 95% confidence intervals of each statistic are given within parentheses.

Dependent variable	R^2	τ
Upward right monotonicity	0.013 (0.008, 0.017)	-0.043 (-0.057, -0.029)
Downward right monotonicity	0.003 (0.000, 0.004)	-0.023 (-0.038, -0.008)
Maximum overall monotonicity	0.027 (0.018, 0.034)	-0.095 (-0.112, -0.078)
Maximum average monotonicity	0.051 (0.041, 0.059)	-0.142 (-0.156, -0.128)
Quantitativity	0.040 (0.035, 0.045)	0.175 (0.161, 0.189)
Left conservativity	0.013 (0.008, 0.016)	-0.069 (-0.082, -0.055)
Right conservativity	0.016 (0.011, 0.020)	-0.083 (-0.097, -0.070)
Average conservativity	0.040 (0.031, 0.049)	-0.147 (-0.162, -0.131)

R^2

Judging by the confidence intervals for R^2, it is highly plausible that all degrees of universals share a linear relationship with complexity, albeit a weak one.

Statistics based on monotonicity are best explained linearly by complexity, with $R^2 = 0.051$. This is in line with the findings of vdP&ST&S.

But conservativity, or at least some statistics based on it, seems to also correlate quite strongly with complexity (with $R^2 = 0.040$), which was not what they found.

τ

It gets worse when we consider τ. Quantitativity has significant and strong positive correlation, with $\tau = 0.175$.

But vdP&ST&S found that quantitative quantifiers do show a tendency towards being less complex, albeit not very robustly.

Correction

It can be verified from the regression and distribution plots that this positive correlation stems from the overwhelming amount of quantifiers that are fully quantitative. Same holds for other universals.

To determine whether complexity can make fine-grained distinctions between quantifiers when it matters - i.e. when the quantifiers neither satisfy, nor fully contradict universals - we perform our analysis again, this time having filtered out extreme quantifiers.

Corrected R^2 and τ

Table 2: Values of R^2 and τ, now taken for data without extreme values.

Dependent variable	R^2	τ
Upward right monotonicity	0.021 (0.013, 0.026)	-0.073 (-0.089, -0.057)
Downward right monotonicity	0.007 (0.002, 0.011)	-0.055 (-0.073, -0.037)
Maximum overall monotonicity	0.043 (0.026, 0.056)	-0.163 (-0.189, -0.137)
Maximum average monotonicity	0.050 (0.041, 0.059)	-0.132 (-0.147, -0.118)
Quantitativity	0.038 (0.020, 0.052)	-0.119 (-0.151, -0.089)
Left conservativity	0.001 (-0.001, 0.002)	-0.004 (-0.019, 0.010)
Right conservativity	0.002 (0.000, 0.003)	-0.015 (-0.029, -0.001)
Average conservativity	0.034 (0.025, 0.042)	-0.134 (-0.150, -0.118)

Quantitativity

Quantitativity no longer has any positive correlation, and instead displays relatively strong negative correlation now ($\tau = -0.119$).

But similar to what vdP&ST&S found, this is not a robust correlation: the confidence interval for both R^2 and τ is larger than that of any other statistic, even much larger for τ.

Monotonicity

In the case of monotonicity, removing extreme values has only strengthened all correlations. It appears that monotonicity is truly explained well by complexity.

Conservativity

Our findings for conservativity are quite different now, and are fully in line with vdP&ST&S. The correlations for both left and right conservativity are no longer statistically significant, with confidence intervals passing 0.

Note that results with similar implications were also found by Steinert-Threlkeld and Szymanik (2019) when looking at learnability.

Conclusion of experiments

So in conclusion, our findings indicate that the results of vdP&ST&S do in fact generalise: complexity does explain monotonicity, as well as quantitativity, though this latter result is not very robust. Finally, conservativity is not explained by complexity.

Degree robustness

Formal analysis

As promised, we now consider the question whether degrees are robust, in the sense that they actually approach some value as n increases. Answering this question for the three degrees of universals is difficult.

As a first step towards answering this question for degrees of universals, we have that degrees of explanation are generally not robust.

Formal analysis

Theorem

There exists a quantifier Q and a binary measure $X : \mathcal{M} \to 2$ for which the limit

$$\lim_{n\to\infty} \deg_n^X(Q)$$

does not exist.

Proof sketch.

Consider the quantifier Q with stream $101010\ldots$, and define X in such a way that it agrees with Q on all Q-true models of size at least 2, while having $X(\mathbb{M}) = 1$ for $\frac{1}{4}$-th of the Q-false models if n is even, and for $\frac{3}{4}$-th of them if n is odd. Then we get that $\lim_{n\to\infty} \deg_{2n}^X(Q) \neq \lim_{n\to\infty} \deg_{2n+1}^X(Q)$, with both of these even and odd limits existing. □

Formal analysis

The constructions in this proof are highly artificial, and may not have any bearing on degrees of actual universals. It is instead an invitation to give these degrees serious consideration.

Future work

Theoretical

- Expand upon the formal analysis: are there necessary and/or sufficient conditions on quantifiers and/or measures under which a degree of explanation converges? If so, do the measures of universals satisfy those for measures?
- Can we redefine the notion of the degree to which a universal is satisfied in such a way that we do not encounter any issues with convergence and model size? Possible option is to work with non-uniform distributions over models.

Methodological

- Use balanced samples of quantifiers w.r.t. the degrees of universals.
- Define a more natural language of thought with few primitives, and also consider production depth as another measure of complexity.
- Consider larger models (c.f. Steinert-Threlkeld and Szymanik (2019), who considered models of up to size 20).

Extensions

Just one important extension: reuse this approach with learnability instead of or in addition to complexity.

Questions?

R^2 AND τ

Table 3: Pre-correction

Dependent variable	R^2	τ
Upward right monotonicity	0.013 (0.008, 0.017)	-0.043 (-0.057, -0.029)
Downward right monotonicity	0.003 (0.000, 0.004)	-0.023 (-0.038, -0.008)
Upward left monotonicity	0.015 (0.010, 0.019)	-0.051 (-0.065, -0.036)
Downward left monotonicity	0.003 (0.000, 0.005)	-0.030 (-0.045, -0.015)
Maximum right monotonicity	0.019 (0.013, 0.025)	-0.094 (-0.110, -0.079)
Maximum left monotonicity	0.021 (0.015, 0.027)	-0.104 (-0.120, -0.089)
Average maximum monotonicity	0.039 (0.030, 0.047)	-0.143 (-0.158, -0.129)
Maximum overall monotonicity	0.027 (0.018, 0.034)	-0.095 (-0.112, -0.078)
Maximum average monotonicity	0.051 (0.041, 0.059)	-0.142 (-0.156, -0.128)
Quantitativity	0.040 (0.035, 0.045)	0.175 (0.161, 0.189)
Left conservativity	0.013 (0.008, 0.016)	-0.069 (-0.082, -0.055)
Right conservativity	0.016 (0.011, 0.020)	-0.083 (-0.097, -0.070)
Maximum conservativity	0.035 (0.024, 0.044)	-0.170 (-0.187, -0.152)
Average conservativity	0.040 (0.031, 0.049)	-0.147 (-0.162, -0.131)

Corrected R^2 and τ

Table 4: Post-correction.

Dependent variable	R^2	τ
Upward right monotonicity	0.021 (0.013, 0.026)	-0.073 (-0.089, -0.057)
Downward right monotonicity	0.007 (0.002, 0.011)	-0.055 (-0.073, -0.037)
Upward left monotonicity	0.023 (0.015, 0.029)	-0.076 (-0.092, -0.060)
Downward left monotonicity	0.007 (0.002, 0.010)	-0.053 (-0.072, -0.035)
Maximum right monotonicity	0.023 (0.014, 0.030)	-0.122 (-0.141, -0.104)
Maximum left monotonicity	0.023 (0.015, 0.032)	-0.124 (-0.143, -0.106)
Average maximum monotonicity	0.029 (0.020, 0.036)	-0.120 (-0.136, -0.104)
Maximum overall monotonicity	0.043 (0.026, 0.056)	-0.163 (-0.189, -0.137)
Maximum average monotonicity	0.050 (0.041, 0.059)	-0.132 (-0.147, -0.118)
Quantitativity	0.038 (0.020, 0.052)	-0.119 (-0.151, -0.089)
Left conservativity	0.001 (-0.001, 0.002)	-0.004 (-0.019, 0.010)
Right conservativity	0.002 (-0.000, 0.003)	-0.015 (-0.029, -0.001)
Maximum conservativity	0.007 (0.002, 0.011)	-0.085 (-0.104, -0.066)
Average conservativity	0.034 (0.025, 0.042)	-0.134 (-0.150, -0.118)

Carcassi, F., Steinert-Threlkeld, S., and Szymanik, J. (2019). “The emergence of monotone quantifiers via iterated learning”. In: *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, pp. 190–196.

Kendall, M. G. (1945). “The treatment of ties in ranking problems”. In: *Biometrika* 33 (3), pp. 239–251. DOI: 10.2307/2332303.

Lempel, A. and Ziv, J. (1976). “On the complexity of finite sequences”. In: *IEEE Transactions on Information Theory* 22 (1), pp. 75–81. DOI: 10.1109/TIT.1976.1055501.

References II

- Peters, S. and Westerståhl, D. (2006). *Quantifiers in Language and Logic*. Oxford: Clarendon Press. ISBN: 9780 199291250.
- Piantadosi, S. T., Tenenbaum, J. B., and Goodman, N. D. (2012). "Modeling the acquisition of quantifier semantics: a case study in function word learnability".
- van de Pol, I., Steinert-Threlkeld, S., and Szymanik, J. (2019). "Complexity and learnability in the explanation of semantic universals". In: *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, pp. 3015–3021.
- Posdijk, W. (2019). "The influence of the simplicity / informativeness trade-off on the sematic typology of quantifiers". Master's thesis. Universiteit van Amsterdam.

References iii

Steinert-Threlkeld, S. and Szymanik, J. (2019). “Learnability and semantic universals”. In: *Semantics and Pragmatics* 12 (4).