Noname manuscript No.
(will be inserted by the editor)

Quantifiers, complexity, and degrees of semantic universals

A large-scale analysis

Nima Motamed

August 18, 2020

Abstract In spite of large diversity between natural languages, there exist some properties (referred to as universals) that are common to virtually all. Multiple universals on the semantic level have been formulated for quantifiers, and it has still not been fully determined why these universals appear. Past empirical research has pointed both towards learnability and complexity as explaining factors for their appearance: universal expressions are easier to learn and/or are less complex. To take another step towards their understanding, we generalise past work by giving a complete formal formulation of the notion of the degree to which a universal is satisfied by a quantifier. After proving a preliminary formal result on these degrees, we perform a large-scale empirical analysis of quantifiers generated from a logical grammar, determining correlations between the aforementioned degrees and the approximate Kolmogorov complexity of these quantifiers. Though overall correlations between these degrees and complexity were weak, we find that the monotonicity universal's degree correlates most strongly and robustly, the quantitativity universal's degree correlates reasonably, though not robustly, and finally that conservativity correlates weakly.

1 Introduction

While the diversity of natural languages is certainly enormous, there is reason to believe that there are features shared by most, if not all of them. Such features (called *universals*) have been attested at several levels of linguistic analysis (e.g., Hyman (2008) and Newmeyer (2008)), including at that of the analysis of natural language semantics (e.g., Barwise and Cooper (1981)). There is still no definitive explanation for the appearance of universals across languages.

Nima Motamed
E-mail: nima@motamed.nl

Seeking such an explanation, it is natural to turn our gaze inward and consider the relation between human cognition and universals. This relation has given rise to the oft-posited *learnability hypothesis* (e.g., Barwise and Cooper (1981), Keenan and Stavi (1986), and Peters and Westerståhl (2006)), stating that universals appear because expressions satisfying them are easier to learn. There is empirical evidence in the literature supporting this hypothesis, at least for semantic universals and *quantifier expressions*: Steinert-Threlkeld and Szymanik (2019, 2020) have found (amongst other results) that machine learning models (neural networks, specifically) find it easier to learn representations of quantifier expressions satisfying universals. Carcassi, Steinert-Threlkeld, and Szymanik (2019) introduce the information-theoretic notion of the *degree* to which a universal is satisfied by a quantifier, and show that quantifiers satisfying universals to a higher degree may naturally emerge in languages as a result of cultural evolution. Also of note is the work by Piantadosi, Tenenbaum, and Goodman (2012): they study the learnability of quantifiers using a learning model in which learners have access to a logical *grammar* generating possible semantic representations of quantifiers.

Learnability is not the only angle from which semantic universals of quantifiers can be studied. Van de Pol, Steinert-Threlkeld, and Szymanik (2019) have studied the relation between the satisfaction of semantic universals and a notion of the *complexity* of quantifiers stemming from (algorithmic) information theory. They found that one universal could be explained by complexity (in the sense that quantifiers satisfying universals are less complex), while others could not. Important to note is that they, like Steinert-Threlkeld and Szymanik (2019), determined this by manually selecting pairs of quantifiers that do and do not satisfy a universal, for which they then looked at the differences in complexity. This approach is limited in that it need not generalise to arbitrary quantifiers.

In this paper, we aim to overcome this limitation by generalising the work of van de Pol, Steinert-Threlkeld, and Szymanik (2019). This generalisation relies on two innovations. First, we introduce a logical grammar generating quantifier expressions as done by Piantadosi, Tenenbaum, and Goodman (2012), in order to do away with the manual selection of quantifiers. Second, we do not use binary notions of satisfaction of a universal, but instead work with degrees as introduced by Carcassi, Steinert-Threlkeld, and Szymanik (2019). This way, we can rigorously determine whether quantifiers' complexities explain the appearance of a universal by analysing the correlation between the complexity and the degree of that universal.

The paper is structured as follows. We introduce the framework of generalised quantifier theory along with the definitions of semantic universals for quantifiers in Section 2. We then move on to give a general definition of degrees of universals, and prove a negative result about them in Section 3. We then shortly define both the complexity of quantifiers and the logical grammar we use for our results in Sections 4 and 5, respectively. In Section 6 we give a short description of the process used to obtain our empirical data. Finally, we present and analyse the data in Section 7, before discussing and giving directions for future work in Section 8.

2 Quantifiers and universals

As is usual in this field of study, quantifiers are considered to be the semantic objects expressed by *determiners*, which express binary relations between sets of objects. These determiners can generally be classified as being either grammatically simple or complex determiners. In English, simple determiners include *all*, *some*, and *most*, while the complex determiners include expressions like *all but three* and *at least four*. The semantic universals we will discuss in this section all take the form of properties that all simple determiners appear to share. But before we can introduce these universals, we need to be able to rigorously discuss what constitutes a determiner.

2.1 Generalised quantifiers

To this end, we use the framework of generalised quantifier theory.[1] Within this framework, determiners in natural language correspond to type $\langle 1,1\rangle$ generalised quantifiers (which we will just refer to as quantifiers from now on, for the sake of brevity). These are sets of triples $\mathbb{M} = \langle M,A,B\rangle$ which consist of sets satisfying $A,B \subseteq M$. We refer to such triples as *models*. The sets M, A, and B are referred to respectively as the *domain of discourse*, *restrictor*, and *scope* of the model $\mathbb{M}$. If $\mathbb{M} \in \mathsf{Q}$ for some model $\mathbb{M}$ and quantifier Q, we interpret this as stating that within the domain of discourse M, the quantifier expression representing Q applied to the restrictor A and scope B is true. Note that we can also consider Q as a binary predicate sending models to truth values, defined by putting $\mathsf{Q}_M(A,B) = 1$ iff $\langle M,A,B\rangle \in \mathsf{Q}$. For brevity of notation, we will often write $\mathsf{Q}_M(A,B)$ as a statement meaning $\mathsf{Q}_M(A,B) = 1$.

Considering some natural language determiner *Det* representing quantifier Det, along with a noun N and verb phrase VP representing respectively the sets $[\![N]\!]$ and $[\![VP]\!]$ of individuals, the statement $\mathsf{Det}_M([\![N]\!],[\![VP]\!])$ is equivalent to the simple sentence *Det N VP* being true within the domain of discourse M. Some examples of quantifiers occurring in natural language include

$$\begin{aligned} \mathsf{all} &= \{\langle M,A,B\rangle \, ; A \subseteq B\}\,, \\ \mathsf{some} &= \{\langle M,A,B\rangle \, ; A \cap B \neq \emptyset\}\,, \\ \mathsf{most} &= \{\langle M,A,B\rangle \, ; |A \cap B| > |A \setminus B|\}\,, \text{ and} \\ \mathsf{all\_but\_three} &= \{\langle M,A,B\rangle \, ; |A \setminus B| = 3 \text{ and } |A| > 3\}\,. \end{aligned}$$

The astute reader may have noticed that all of the determiners we have mentioned so far correspond to quantifiers whose definitions only mention the restrictor and scope, not the exact domain of discourse. This property is called *extensionality*, and can formally be defined as follows.

Definition 1 (Extensionality) A quantifier Q is called *extensional* if for all sets $A,B \subseteq M \subseteq M'$, it holds that $\mathsf{Q}_M(A,B) = \mathsf{Q}_{M'}(A,B)$. ◇

All quantifiers (simple or complex) in natural language are extensional.[2] In part due to this, and in part due to computational considerations we discuss in Section 5, we limit our attention to extensional quantifiers for the experimental part of this paper.

2.2 Semantic universals

Having properly defined quantifiers, we can now introduce the three semantic universals this paper focuses on. These are properties that seemingly all simple determiners satisfy. It does not lie within the scope of this paper to thoroughly analyse them or give arguments for their universality - we refer the interested reader to (the literature referred to by) Peters and Westerståhl (2006) for more exposition.

[1] See Peters and Westerståhl (2006) for a thorough overview of the theory.

[2] This seemingly does not generally hold if you also consider type $\langle 1\rangle$ quantifiers. We do not treat these here, and instead refer the interested reader to Chapter 4.5 of Peters and Westerståhl (2006), containing arguments and proofs showing that these are essentially equivalent to quantifiers that are extensional.

The first of the universals we consider is *monotonicity*. This property captures that satisfaction of a quantifier is preserved when making the scope (or restrictor) more general or specific.

Definition 2 (Monotonicity) A quantifier Q is called *right monotone* if it is *upward* right monotone or *downward* right monotone. It is called upward [downward] right monotone if for any sets $A \subseteq M$ and $B \subseteq B'$ [$B' \subseteq B$] it holds that $\mathsf{Q}_M(A,B)$ implies $\mathsf{Q}_M(A,B')$. Analogously, we call Q *left monotone* if it is upward left monotone or downward left monotone. It is called upward [downward] left monotone if for any sets $B \subseteq M$ and $A \subseteq A'$ [$A' \subseteq A$] it holds that $\mathsf{Q}_M(A,B)$ implies $\mathsf{Q}_M(A',B)$. $\diamond$

We include the definition of left monotonicity for completeness: the semantic universal as formulated by Barwise and Cooper (1981) states that all simple determiners are *right* monotone.

The quantifier *most* is an upward right monotone quantifier: e.g., assuming that the sentence "most dogs bark" is true and that barking things also breathe, the sentence "most dogs breathe" must also be true. It is not downward right monotone, since the sentence does not guarantee the truth of "most dogs bark loudly." It is also not left monotone: neither "most animals bark" nor "most dalmatians bark" is guaranteed purely by the truth of the original sentence. It should not be difficult to see that *all* is both left and right monotone, while *an even number of* is not monotone at all.

The second universal is called *quantitativity*. This property captures that a quantifier is only concerned with quantities, as opposed to the particular presentation of or identity of members of the domain, restrictor and scope. This boils down to a quantifier's satisfaction being *isomorphism-invariant* in the sense that the quantifier gives the same truth value for isomorphic models, with isomorphisms as defined within generalised quantifier theory. Using an equivalent formulation of isomorphism-invariance, this gives rise to the following definition.

Definition 3 (Quantitativity) A quantifier Q is called *quantitative* if for any two models $\langle M,A,B\rangle$ and $\langle M',A',B'\rangle$ such that $|A\cap B| = |A'\cap B'|$, $|A\setminus B| = |A'\setminus B'|$, $|B\setminus A| = |B'\setminus A'|$, and $|M\setminus(A\cup B)| = |M'\setminus(A'\cup B')|$, it holds that $\mathsf{Q}_M(A,B) = \mathsf{Q}_{M'}(A',B')$. $\diamond$

Note that all simple determiners we have named so far are in fact quantitative, even if it might not seem so at first sight. For example, even though the denotation we gave for the determiner *all* did not contain any reference to cardinalities, we could have equivalently defined it as the set of models satisfying $|A\setminus B| = 0$. As expected however, complex determiners need not be quantitative: the determiner *the first three* relies also on the order of presentation of the restrictor.

The final universal we consider is called *conservativity*. Put simply, this property captures that restrictors are actually restricting the meaning of a sentence. It expresses that the validity of a sentence of the form *Det N VP* is not concerned with any instance of the *VP* that is not also an instance of the *N*. Generalising this idea, this gives rise to the following definitions.

Definition 4 (Conservativity) A quantifier Q is called *left conservative* [*right conservative*] if for any model $\langle M,A,B\rangle$ it holds that $\mathsf{Q}_M(A,B) = \mathsf{Q}_M(A,A\cap B)$ [$\mathsf{Q}_M(A,B) = \mathsf{Q}_M(A\cap B,B)$]. $\diamond$

Similar to how it was with monotonicity, the conservativity universal formulated by Barwise and Cooper (1981) states that all simple determiners are *left conservative*.

All simple determiners we introduced can be seen to be (left) conservative. None of them rely in any way on the set $B\setminus A$: e.g., to see that *some* is conservative, note that $A\cap B \neq \emptyset$ iff $A\cap(A\cap B)\neq\emptyset$. Looking at it in the context of language, we see that the sentence "some cars are hybrids" is equivalent to "some cars are cars that are hybrids." Considering a complex determiner E expressing that $|A| = |B|$, it follows that E is not conservative since $|A| = |A\cap B|$ does not guarantee that E holds.

3 Degrees of universals

Having introduced these semantic universals, we now move on to define one of the innovations we mentioned in Section 1: the definition of a degree to which a universal is satisfied. After giving this definition, we perform a preliminary formal analysis.

Our information-theoretic definition of a degree is based on the one formulated by Carcassi, Steinert-Threlkeld, and Szymanik (2019) and Posdijk (2019). This definition boils down to the normalised mutual information between (random variables representing) the truth values of the quantifier and some other source of information, dependent on the exact universal in question. Preparing ourselves for the aforementioned formal result, we give a very general definition of the degrees.

3.1 Degrees of explanation

Let us begin with some preliminaries. Our definition assumes that all models are finite (i.e. M is finite for each model). This guarantees that we are capable of computing the degrees. Furthermore, we consider only non-empty models that contain objects from some greater countably infinite universe $U = \{o_1, o_2, \dots\}$, such that a model of size n (i.e. a model with $|M| = n$) has domain of discourse $M = \{o_1, \dots, o_n\}$. This way we obtain a simple, topic-neutral and non-syntactic class of models. Because all models of the same size have the same domain of discourse, we are able to properly study relations between models as required for verifying universals.

These assumptions also give rise to a very natural way to represent models. For any model of size n, each of the objects $o_1, \ldots, o_n$ must be in exactly one of the 'zones' $A \cap B$, $A \setminus B$, $B \setminus A$, and $M \setminus (A \cup B)$. Specifying in which of these four zones objects lie uniquely characterises a model. It is not difficult to see that this means that we can represent models as non-empty quaternary strings by encoding zones by the digits 0 to 3. Any model of size n, uniquely corresponds to the string $\alpha \in 4^n$ defined by putting $\alpha_i := d$ if the object o_i is in the zone encoded by the digit d. Conversely, each string $\alpha \in 4^n$ uniquely corresponds to the model of size n defined by placing object o_i in the zone encoded by the digit d if $\alpha_i = d$. We denote the model corresponding to the string α by $\mathbb{M}_\alpha$.

By this correspondence, the standard increasing lexicographical ordering of quaternary strings gives us a standard ordering of all models as well. This ordering of models then gives rise to yet another natural representation, this time for quantifiers. Since a quantifier is fully determined by the models that validate it, we can now represent a quantifier Q by an infinite binary string β^{Q} defined by putting $\beta_i^{\mathsf{Q}} = 1$ iff the i-th model validates Q. This representation will not only come in handy in our proof in Section 3.2, but is also vital to the notion of complexity we introduce in Section 4.

We can now move on to the heart of the matter: the definition of the degrees. We wish to base our definition on random variables defined on the set of all models, but at the same time we wish to consider a uniform distribution on this set for the sake of generality. Since this is not possible for the set of all models, we incrementally construct degrees, built on a finite and increasing set of models.

For any $n \geqslant 1$, define the set $\mathcal{M}_n := \{\mathbb{M}_\alpha\,;\,\alpha \in 4^n\}$ of all models of size n. Using this, we define the set $\mathcal{M}_{\leqslant n} := \bigcup_{1 \leqslant k \leqslant n} \mathcal{M}_k$ of all models of size at most n. This will be the set over which our random variables are defined. Given a quantifier Q, define the random variable $\mathbb{1}_{\mathsf{Q},n} : \mathcal{M}_{\leqslant n} \to 2$ by setting $\mathbb{1}_{\mathsf{Q},n}(\mathbb{M})$ as the truth value that Q assigns to $\mathbb{M}$. This random variable is all we need to define the degrees we want.

Definition 5 (Degrees of explanation) Given some integer $n \geqslant 1$, a quantifier Q, and some function $X : \bigcup_{k \geqslant 1} \mathcal{M}_{\leqslant k} \to \mathcal{X}$, the *n-th degree of explanation of* Q *by* X is defined as

$$\begin{aligned}\mathsf{deg}_n^X(\mathsf{Q}) &:= \frac{I(\mathbb{1}_{\mathsf{Q},n}\,;X_n)}{H(\mathbb{1}_{\mathsf{Q},n})}\\ &= 1 - \frac{H(\mathbb{1}_{\mathsf{Q},n}\mid X_n)}{H(\mathbb{1}_{\mathsf{Q},n})},\end{aligned}$$

where X_n is the random variable over $\mathcal{M}_{\leqslant n}$ defined as the restriction $X\restriction_{\mathcal{M}_{\leqslant n}}$, H is the (conditional) Shannon entropy, and I is the mutual information. ◇

Intuitively, the degree of explanation of Q by X how much knowing about the information given by X reduces one's uncertainty about the truth of Q. If X is very informative, the degree will tend towards 1, while it will tend towards 0 if X is not informative at all.

By replacing X in this general definition with function measuring the main sources of information pertaining to a semantic universal, we finally obtain a degree measuring that universal's satisfaction.

Definition 6 (Degrees of universals) Define the following functions on $\bigcup_{k \geqslant 1} \mathcal{M}_{\leqslant k}$:

- The upward right, upward left, downward left, and downward right monotonicity functions $\mathbb{1}_{\mathsf{Q}}^{\nearrow}$, $\mathbb{1}_{\mathsf{Q}}^{\nwarrow}$, $\mathbb{1}_{\mathsf{Q}}^{\swarrow}$, and $\mathbb{1}_{\mathsf{Q}}^{\searrow}$, which given a model $\mathbb{M} = \langle M, A, B\rangle$ return 1 if there respectively is some $B' \subseteq B$, $A' \subseteq A$, $A' \supseteq A$, or $B' \supseteq B$ such that Q is satisfied on the model $\mathbb{M}$ with that sub- or superset instead, and 0 otherwise.
- The quantitativity function #, which given a model $\langle M, A,\ B\rangle$ returns the quadruple $\langle |A \cap B|, |A \setminus B|,\ |B \setminus A|, |M \setminus (A \cup B)|\rangle$.
- The left and right conservativity functions $\restriction^{\leftarrow}$ and $\restriction^{\rightarrow}$, which given a model $\langle M, A, B\rangle$ return the triples $\langle M, A, A \cap B\rangle$ and $\langle M, A \cap B, B\rangle$, respectively.

Given a quantifier Q, the *n-th degrees of monotonicity* $\mathsf{mon}_n^{\nearrow}(\mathsf{Q})$, $\mathsf{mon}_n^{\nwarrow}(\mathsf{Q})$, $\mathsf{mon}_n^{\swarrow}(\mathsf{Q})$, and $\mathsf{mon}_n^{\searrow}(\mathsf{Q})$, along with the *n-th degree of quantitativity* $\mathsf{quan}_n(\mathsf{Q})$ and the *n-th degrees of conservativity* $\mathsf{con}_n^{\leftarrow}(\mathsf{Q})$ and $\mathsf{con}_n^{\rightarrow}(\mathsf{Q})$ are the n-th degrees of explanation of Q by the corresponding functions defined above. ◇

It should not be difficult to verify that for each quantifier Q, these n-th degrees of universals are equal to 1 iff Q satisfies the corresponding universal's property when restricted to $\mathcal{M}_{\leqslant n}$.

3.2 Robustness of degrees

Our definition of the degrees has deviated slightly from that of Carcassi, Steinert-Threlkeld, and Szymanik (2019) and Posdijk (2019), not only in its generality but also in the additional parameterisation by an integer n. This parameterisation is in fact implicit in their results, as they also compute degrees over uniformly distributed finite sets of models. Explicitly including this parameter in our definition of the degrees allows us to study their properties.

For instance, it is natural to question whether degrees of explanation are 'robust', in the sense that as n increases, the n-th degree gradually approaches a fixed value value. We believe this is not at all obvious from the definition. As a first step towards answering this question, we show that degrees of explanation are generally *not* robust.

Proposition 1 *There exists a quantifier* Q *and some function* $X : \bigcup_{k\geqslant 1}\mathcal{M}_{\leqslant k} \to \mathcal{X}$ *for which the limit*

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

does not exist.

Proof A quick note: this proof involves large calculations, some of which are too large to consider here. We only sparingly show the details of these calculations, as they use quite elementary techniques, only being complicated by virtue of their size.

Consider the quantifier Q defined with corresponding binary string $\beta^{\mathsf{Q}} = 10101010\ldots$. Since $|\mathcal{M}_{\leqslant n}| = \frac{4}{3}(4^n - 1)$ is even for all $n \geqslant 1$, it follows that we always have that $\mathbb{1}_{\mathsf{Q},n} = 1$ with probability $\frac{1}{2}$. This already shows that $H(\mathbb{1}_{\mathsf{Q},n}) = 1$ for all $n \geqslant 1$, and so it suffices to find some X such that $H(\mathbb{1}_{\mathsf{Q},n} \mid X_n)$ diverges as $n \to \infty$.

We now define such a binary function $X : \bigcup_{k\geqslant 1}\mathcal{M}_{\leqslant k} \to 2$ by iteratively constructing its finite restrictions X_k - clearly this also uniquely defines the entire function X. Let $X_1 := \mathbb{1}_{\mathsf{Q},1}$ (i.e. the first finite restriction of $\mathbb{1}_{\mathsf{Q}}$), and then iteratively define X_n for $n \geqslant 2$ as follows. For all models $\mathbb{M} \in \mathcal{M}_{\leqslant n-1}$, we set $X_n(\mathbb{M}) := X_{n-1}(\mathbb{M})$. For the other $\mathbb{M}$ (i.e. those in $\mathcal{M}_n$), we first set $X_n(\mathbb{M}) := 1$ for all $\mathbb{M} \in \mathsf{Q}$ of size n. Depending on the parity of n, we additionally set $X_n(\mathbb{M}) := 1$ for a certain amount of models $\mathbb{M} \notin \mathsf{Q}$ based on the following scheme. If n is even, set $X_n(\mathbb{M}) := 1$ for $\frac{1}{4}$-th of the models $\mathbb{M} \notin \mathsf{Q}$ of size n. Otherwise, set $X_n(\mathbb{M}) := 1$ for $\frac{3}{4}$-th of these models. To see that this gives a well-defined family of random variables, note that $|\mathcal{M}_n \setminus \mathsf{Q}| = \frac{4^n}{2} = 2^{2n-1}$ is divisble by 4 for $n \geqslant 2$.

Expanding the definition of the conditional entropy, we know that

$$\begin{aligned} H(\mathbb{1}_{\mathsf{Q},n} \mid X_n) = {} & Pr[X_n = 1]H(\mathbb{1}_{\mathsf{Q},n} \mid X_n = 1) \\ & + Pr[X_n = 0]H(\mathbb{1}_{\mathsf{Q},n} \mid X_n = 0). \end{aligned}$$

Counting the amount of times $X_n = 1$ over its domain, we find that

$$Pr[X_n = 1] = \frac{2 + \sum\limits_{\substack{k=2 \\ k \text{ is even}}}^{n} \frac{5}{4}2^{2k-1} + \sum\limits_{\substack{k=2 \\ k \text{ is odd}}}^{n} \frac{7}{4}2^{2k-1}}{\frac{4}{3}(4^n - 1)},$$

which, after tedious algebraic manipulation, simplifies to

$$= \frac{7\cdot 4^{2\lfloor \frac{1}{2}(n-3)\rfloor + 3} + 5\cdot 16^{\lfloor \frac{n}{2}\rfloor} - 18}{10(4^n - 1)}.$$

When assuming that $X_n = 1$, it folows immediately from the construction of X_n that there are $\sum_{k=1}^{n} 2^{2k-1} = \frac{2}{3}(2^{2n} - 1)$ models of size n such that $\mathbb{1}_{\mathsf{Q},n} = 1$. We have that

$$Pr[\mathbb{1}_{\mathsf{Q},n} = 1 \mid X_n = 1] = \frac{\frac{2}{3}(2^{2n} - 1)}{2 + \sum\limits_{\substack{k=2 \\ k \text{ is even}}}^{n} \frac{5}{4}2^{2k-1} + \sum\limits_{\substack{k=2 \\ k \text{ is odd}}}^{n} \frac{7}{4}2^{2k-1}}$$

which, again after tedious algebraic manipulation, simplifies to

$$= \frac{5(4^n - 1)}{7\cdot 4^{2\lfloor \frac{1}{2}(n-3)\rfloor + 3} + 5\cdot 16^{\lfloor \frac{n}{2}\rfloor} - 18}.$$

By our construction of X_n, we have that $X_n = 0$ guarantees that $\mathbb{1}_{\mathsf{Q},n} = 0$, and so $H(\mathbb{1}_{\mathsf{Q},n} \mid X_n = 0) = 0$. It follows that

$$H(\mathbb{1}_{\mathsf{Q},n} \mid X_n) = Pr[X_n = 1]H(\mathbb{1}_{\mathsf{Q},n} \mid X_n = 1).$$

Working out this expression using the computes values of $Pr[\mathbb{1}_{\mathsf{Q},n} = 1 \mid X_n = 1]$ and $Pr[X_n = 1]$, and taking the limit as the even numbers n go to infinity, we find that

$$\begin{aligned} \lim_{n\to\infty} H(\mathbb{1}_{\mathsf{Q},2n} \mid X_{2n}) &= \frac{1}{2}\log_2\frac{27}{20} + \frac{7}{40}\log_2\frac{27}{7} \\ &\approx 0.557, \end{aligned}$$

while for the limit over the odd numbers n, we find that

$$\begin{aligned} \lim_{n\to\infty} H(\mathbb{1}_{\mathsf{Q},2n+1} \mid X_{2n+1}) &= \frac{1}{2}\log_2\frac{33}{20} + \frac{13}{40}\log_2\frac{33}{13} \\ &\approx 0.798. \end{aligned}$$

In other words, the value of $H(\mathbb{1}_{\mathsf{Q},n} \mid X_n)$ tends towards oscillating between the approximate values 0.557 for even n and 0.798 for odd n. Thus, the limit $\lim_{n\to\infty} H(\mathbb{1}_{\mathsf{Q},n} \mid X_n)$ does not exist, showing that the limit $\lim_{n\to\infty}\deg_n^X(\mathsf{Q})$ also does not exist. □

This result should not be interpreted as saying anything about the usefulness of degrees of explanation in the study of semantic universals - on the contrary, it is an invitation for future work studying the behavior of these degrees when specifically considering degrees of universals. The question whether degrees of universals remains open, and empirical observations have not yet found a quantifier and universal for which the degree diverges. If degrees of universals do not converge in general, however, it does not make much sense to continue using them as an objective measure of how much a quantifier satisfies a universal.

We could redefine degrees of explanation in such a way so that robustness is in fact guaranteed. By placing a probability distribution with countably infinite support over the set of all models, we could define degrees of explanation that take all models into account, removing the need for parameterisation by n. Aside from well-known ones like the geometric distribution, a potentially interesting candidate for this probability distribution is (an approximation of) the universal distribution from the field of algorithmic information theory[3], due to its relation with complexity. We leave this possibility to future work.

[3] We refer the interested reader to Kirchherr, Li, and Vitányi (1997) for a relatively informal introduction, and to Chapters 4 and 5 of Li and Vitányi (2008) for the full formal exposition.

4 Quantifier complexity

The measure of complexity used by van de Pol, Steinert-Threlkeld, and Szymanik (2019) is the *approximate Kolmogorov complexity*.[4] Intuitively, the Kolmogorov complexity measures how much a symbol of sequences can be compressed by recognising and simplifying patterns in the sequence. Formally, the Kolmogorov complexity $K(x)$ of a finite sequence or (usually binary) string x is defined as the size of the smallest (description of a) Turing machine that outputs x. Since it is a known result that K is an uncomputable function, approximations are needed. We define the approximate Kolmogorov complexity as

$$\hat{K}(x) := \log_2(|x|)\frac{LZ(x)+LZ(x^R)}{2},$$

where x^R is the reverse of the string x, and $LZ(x)$ is the Lempel-Ziv complexity (Lempel and Ziv 1976) of the string x.

This begs the question: how can we compute $\hat{K}$ for a quantifier? As stated in Section 3.1, we can view a quantifier Q as an infinite binary string β^{Q} by ordering all models according to the increasing lexicographical order, and then determining the truth value of the quantifier on all of these models. So we could compute $\hat{K}(\beta^{\mathsf{Q}})$. However, this would not be very robust. For this reason, we will compute the average value of $\hat{K}(x)$, with x ranging over all truth-value strings of Q there are for increasing lexicographical orderings of the set of all models, as done by van de Pol, Steinert-Threlkeld, and Szymanik (2019). To guarantee that this value is computable, we need to parameterise by the maximum model size we consider for the computation, as done in Definition 5.

Definition 7 (Quantifier complexity) For any quantifier Q and integer $n \geqslant 1$, the *n-th quantifier complexity of* Q is defined as

$$\mathsf{comp}_n(\mathsf{Q}) = \frac{1}{m}\sum_x \hat{K}(x_{1:|\mathcal{M}_{\leqslant n}|}),$$

where x ranges over the infinite truth-value strings of Q built on the increasing lexicographical orderings of the set of all models, m is the amount of such lexicographical orderings, and $x_{1:k} = x_1 x_2 \cdots x_k$. ◇

5 Quantifier grammar

As mentioned in Section 2.1, it is natural to restrict our attention to extensional quantifiers. This increases computational efficiency, since under this assumption models no longer correspond to quaternary strings, but instead to ternary ones. This greatly reduces not only the sizes of the sets of models and quantifiers we consider, but also the amount of different increasing lexicographical orderings on the set of models.[5]

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

Fig. 1 The logical grammar for generating quantifier expressions. Non-terminals are written in capital italics, while terminals are written in lowercase non-italics. The variable n ranges over the positive integers, and we use the symbol 'card(*SET*)' instead of $|\mathit{SET}|$ to prevent confusion with the guard symbols in the grammar. Expressions are generated by applying production rules from the symbol *START* until there are no non-terminals left.

We have given definitions for the degrees of universals, which were one of the innovations we mentioned in Section 1. Now we introduce the other innovation: the logical (and extensional) grammar in Figure 1, generating quantifier expressions which we use to obtain data on how much complexity explains semantic universals. Our grammar is based on those of Piantadosi, Tenenbaum, and Goodman (2012) and Posdijk (2019), but modified to be able to generate quantifiers that do and do not satisfy the universals we have given. In particular, this grammar generates all quantifiers used in the experiments of van de Pol, Steinert-Threlkeld, and Szymanik (2019) and Steinert-Threlkeld and Szymanik (2019). These include quantifiers referencing an underlying order of presentation of members of sets. We evaluate such quantifiers by assuming that objects in restrictors and scopes are ordered according to the enumeration of objects in the definition of the universe U, as described in Section 3.1.

Note that our grammar does not contain the Boolean not-operator, at least not explicitly. This is a design choice based

[4] For a strong formal introduction to the notion of Kolmogorov complexity as a whole, see Li and Vitányi (2008).

[5] This is a decrease from $4! = 24$ to $3! = 6$ orderings, and observing that $\hat{K}$ is invariant under mirrorings of strings, this actually is a decrease from 12 to 3 orderings.

on computational efficiency: by including negated versions of the primitive operations in the grammar as primitive operations themselves, we can greatly reduce the space of all possible expressions generated by the grammar while losing no expressivity.

6 Methods

Using the grammar in Figure 1 with the variable n ranging up to 10, we generate semantically unique quantifiers from the grammar by considering all productions at a maximal depth of 6. We guarantee semantic uniqueness over the set $\mathcal{M}_{\leqslant 10}$ of models of size at most 10 by comparing the binary strings β^{Q} for each quantifier, and then only considering those whose binary string has not appeared yet. This produces 8044 semantically unique quantifiers. For each quantifier found in this process, we then compute the 10-th degrees of monotonicity, quantitativity, and conservativity, along with the 10-th quantifier complexity. Figure 2 shows the distributions of each of these quantities.

The code necessary for running this process and our data analysis (along with the produced data and instructions) can be found at `https://github.com/nimota/quants-compl-degs`.

7 Results

We determine how well quantifier complexity explains the presence of universals by performing linear regression with the complexity as the independent, and the degrees of universals as dependent variables. The plots of this regression for the degrees of universals are shown in Figure 3. Additionally, we also apply regression to test how well complexity explains certain statistics derived from these degrees, such as the maximum of both degrees of (upward or downward) left and right monotonicity. The values of R^2 (along with 95% confidence intervals[6]) for all of these statistics are displayed in the second column of Table 1.

The confidence intervals for R^2 in Table 1 imply that it is highly plausible that the degrees of universals (and values computed from them) share a linear relationship with complexity, albeit a very weak one. The maximum average monotonicity (i.e. the maximum, taken over upward and downward, of the average monotonicity values, taken over left and right values) is best explained by complexity, with 5% of its variation being explained. Meanwhile, This is somewhat in line with the resuls of van de Pol, Steinert-Threlkeld, and Szymanik (2019), since they found complexity was best at differentiating between monotone and non-monotone quantifiers. The maximum average monotonicity

[6] These confidence intervals were determined non-parametrically through bootstrapping, resampling the original data 10,000 times.

Table 1 The values of R^2 for the regressions with named dependent variables and complexity, along with Kendall's τ for the same variables and complexity as well. The 95% confidence intervals of each statistic are given within parentheses. Maximums are generally taken over the upward and downward versions of a degree (where applicable), while averages are only taken over left and right versions.

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)

is then followed by quantitativity and the average conservativity, both having 4% of their variation explained.

Since none of the relations in Figure 3 seem that linear, we also compute Kendall's τ[7] in order to determine if there is a monotonic relation in general between the dependent variables and complexity. The results of these computations are in the last column of Table 1. These give largely similar results to what we found using R^2, with a few notable exceptions. Quantitativity correlates most strongly now, most likely due to the long uninterrupted line of points with degree 1 we see in Figure 3. Furthermore, it correlates *positively*, and looking at the confidence interval, this is significant. This would go completely against the findings by van de Pol, Steinert-Threlkeld, and Szymanik (2019), who found some (not very robust, by their own admission) evidence that quantitative quantifiers tended to be less complex.

This positive correlation largely seems to be an artifact originating in the large amount of quantifiers that are fully quantitative, as seen in both Figures 2 and 3: the other parts of the data seem to have negative correlation. This is not only true for quantitativity, but for the the other universals as well: many quantifiers' degrees lie on the two extreme ends of the spectrum, skewing our analysis. In order to determine whether complexity is capable of making fine-grained distinctions between quantifiers that neither satisfy nor fully contradict universals, we performed our analysis again on data where points on the extremes were filtered out. The results of this are displayed in Table 2.

The values for R^2 and τ now show different patterns. First off, quantitativity no longer displays any positive correlation, as expected. Its negative (general monotonic) correlation is of a noticeably lower magnitude than its former positive one, though still greater than that of the basic degrees of monotonicity. This fits nicely with the findings of van de Pol,

[7] Specifically, we use the version described by Kendall (1945).

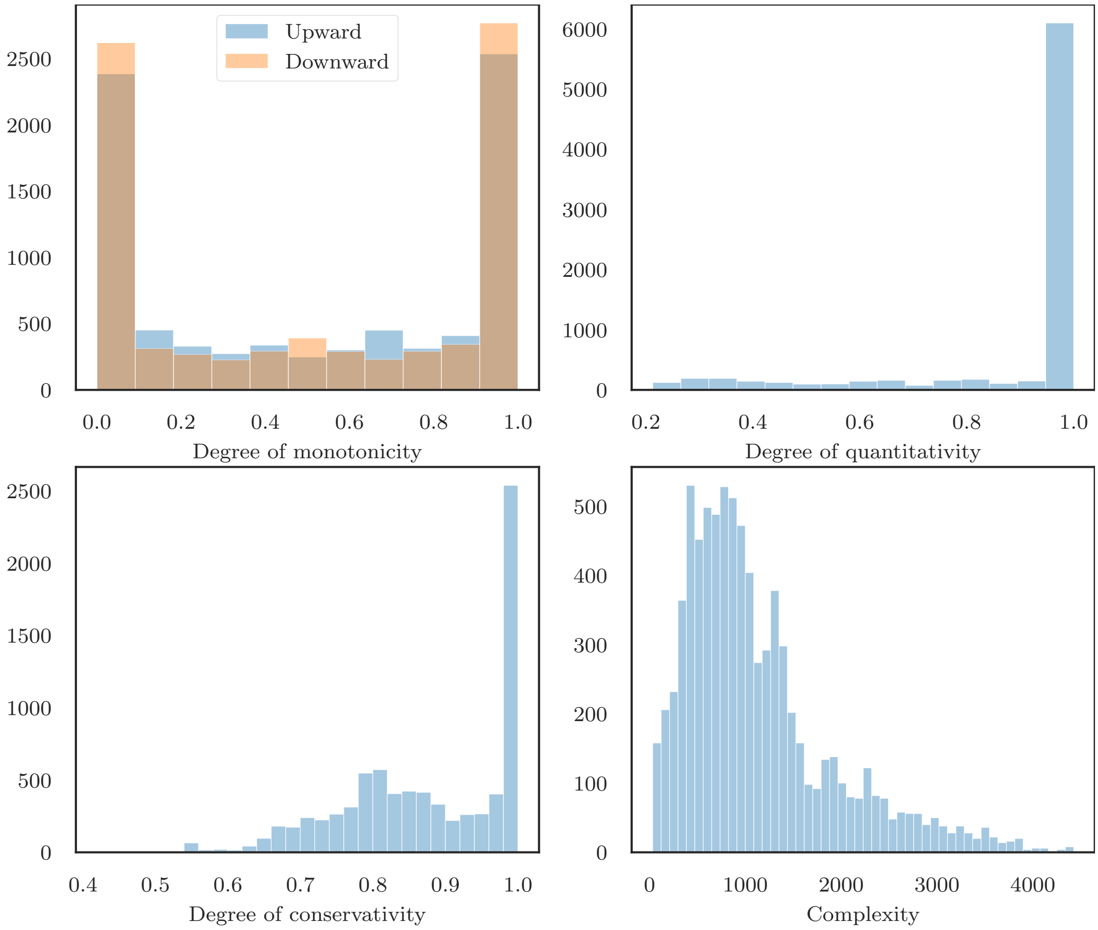

Fig. 2 Distributions of each of the computed measure over all generated quantifiers. Since the distributions of the right and left degrees of both monotonicity and conservativity are indistinguishable, we display these as single distributions (only differentiating between upward and downward monotonicity).

Table 2 Values of R^2 and τ, now taken for data without extreme (i.e. one or zero) 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)
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)

Steinert-Threlkeld, and Szymanik (2019): though we do see relatively strong linear correlation (and average monotonic correlation) for quantitativity, the result is not robust. Not only did the sign of the correlation switch after filtering, but the confidence intervals of both R^2 and τ are the largest of all variables.

The same can not be said of the monotonicity values: virtually all of their correlations have only strengthened. This is in line with the findings of van de Pol, Steinert-Threlkeld, and Szymanik (2019): though the explanatory power of complexity w.r.t. monotonicity we found is a lot weaker than theirs, it is a lot stronger than those w.r.t. other universals.

Also note that the basic degrees of conservativity no longer display significant correlation with complexity: it is consistent with both our found R^2 confidence intervals that there is no linear relation between the two, and the same holds for one of the τ intervals, showing the consistency of a lack of any monotonic relation in general. This supports both van de Pol, Steinert-Threlkeld, and Szymanik (2019) and Steinert-Threlkeld and Szymanik (2019), the former of whom found that conservativity is not explained by complexity, and the latter of whom argued that conservativity fundamentally differs in its source from the other two universals.

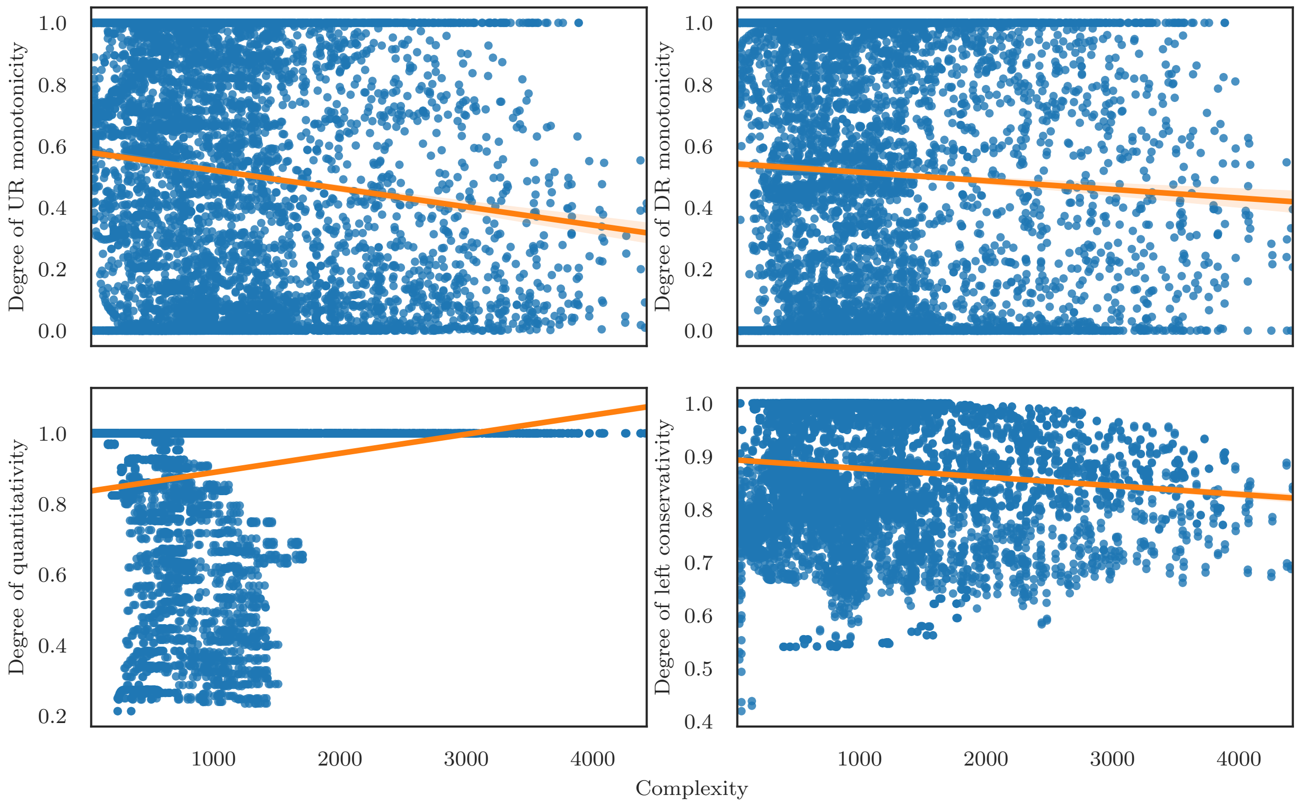

Fig. 3 The fits obtained by performing linear regression of the degrees of universals against quantifier complexity. We leave out the plots for the three other types of monotonicity and the other type of conservativity, since their plots are largely indistinguishable from those included already.

8 Discussion and conclusion

Using methods from both classical and algorithmic information theory, we have measured how well a quantifier's complexity explains the appearance of semantic universals by performing correlation analysis on a large amount of quantifiers generated by a grammar consisting of simple primitives. Our results are largely similar in interpretation to those of van de Pol, Steinert-Threlkeld, and Szymanik (2019), while being much larger in scale and more reliable. While the strength with which complexity explains the presence of universals is generally weak, the relative difference in explanation were analogous to those found in their work. We found that monotonic quantifiers are explained most strongly, that there is (not fully robust) evidence that quantitative ones are also reasonably explained, and that conservativity seems to be poorly explained.

While our results are strong and objective enough to stand on their own, we believe that they also show important avenues for future work. First, our work can be scaled up even further, by either generating quantifiers from the grammar up to an even greater depth, or by considering larger model sizes. Second, both the degrees of universals we defined as well as alternative approaches (such as probability distributions other than the uniform distribution) require thorough formal and empirical study in order to both increase our general understanding of the relation between information and quantifiers, as well as obtain more robust measures for future large-scale correlation studies. Finally, our large-scale approach is perfect for adaptation to the study of semantic universals and learnability, as done by Steinert-Threlkeld and Szymanik (2019). This would give much stronger evidence for or against the learnability hypothesis by determining correlations between ease of learning, degrees of universals, and complexity.

References

Barwise, J. and Cooper, R. (1981). "Generalized quantifiers and natural language". In: *Linguistics and Philosophy* 4, pp. 159–219. DOI: 10.1007/BF00350139.

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.

Hyman, L. M. (2008). "Universals in phonology". In: *The Linguistic Review* 25.1-2, pp. 83–137. DOI: 10.1515/TLIR.2008.003.

Keenan, E. L. and Stavi, J. (1986). “A semantic characterization of natural language determiners”. In: *Linguistics and Philosophy* 9, pp. 253–326. DOI: 10.1007/BF00630273.

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

Kirchherr, W., Li, M., and Vitányi, P. (1997). “The miraculous universal distribution”. In: *The Mathematical Intelligencer* 19, pp. 7–15. DOI: 10.1007/BF03024407.

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.

Li, M. and Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications*. 3rd ed. ISBN: 97814899 84456. DOI: 10.1007/978-0-387-49820-1.

Newmeyer, F. J. (2008). “Universals in syntax”. In: *The Linguistic Review* 25.1-2, pp. 35 –82. DOI: 10.1515/TLIR.2008.002.

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”. URL: http://colala.berkeley.edu/papers/piantadosi2012modeling.pdf.

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.

Steinert-Threlkeld, S. and Szymanik, J. (2019). “Learnability and semantic universals”. In: *Semantics and Pragmatics* 12 (4). DOI: 10.3765/sp.12.4.

— (2020). “Ease of learning explains semantic universals”. In: *Cognition* 195. DOI: 10.1016/j.cognition.2019.104076.