Reasoning about stored representations in semantics using the typology of lexicalized quantifiers[1]

Maike ZÜFLE — *University of Heidelberg*
Roni KATZIR — *Tel Aviv University*

Abstract. The typology of lexicalizations in natural languages is highly skewed: some meanings repeatedly receive their own expression as individual morphemes or words in language after language, while many other meanings rarely or never do. For example, while many languages have monomorphemic counterparts of English 'some' and 'all', no known language has a monomorphemic quantifier that means 'all or none' or a quantifier that asserts that its two arguments are of the same cardinality. It seems tempting to reason from this typological skew to properties of stored representations. However, it is not generally safe to assume that if something is typologically unattested then it simply cannot be represented or learned. The representational system for stored denotations is just one of several interacting factors that affect the typology, and other factors such as communicative pressure and learnability are likely to shape patterns of lexicalization. In this paper we propose to reason from the typology to stored representations by modeling the representational framework, communicative pressure, and learnability directly within an evolutionary model, building on work by Brochhagen et al. (2018). Our empirical focus is a lexicalization asymmetry noted by Horn (1972) in the domain of logical operators and framed within the Aristotelian Square of Opposition. We show that, on certain assumptions, Horn's lexicalization pattern depends on very particular representational costs in the lexicon: it arises if the storage costs for 'every' and 'some' are lower than those for 'not every' and 'not some' but not otherwise.

Keywords: typology, stored representations, lexicalization, logical operators, evolutionary modeling

1. Introduction

The typology of lexicalizations in natural languages is highly skewed: some meanings repeatedly receive their own expression as individual morphemes or words in language after language, while many other meanings rarely or never do. For example, while many languages have monomorphemic counterparts of English 'some' and 'every', no known language has a monomorphemic quantifier that means 'all or none' or a quantifier that asserts that the cardinality of its restrictor is greater than that of its nuclear scope (if 'gleeb' were such a quantifier in English, "Gleeb dogs smoke" would be true exactly when there are more dogs than smokers; nothing is *a priori* odd about this concept, as can be seen from the English verb 'outnumber', but no known quantifier in any language has this denotation). While this typological skew has been discussed across categories, some of its most striking manifestations have been observed for functional categories, and in particular for so-called logic words (elements such as 'and', 'or', 'every', and 'some'), which are the empirical focus of the present paper. In this domain, several typological asymmetries were observed in early work in semantics — notable examples include Horn (1972) and Barwise and Cooper (1981) — and have remained open challenges

[1]We would like to thank Moshe Bar-Lev, Adi Behar Medrano, Thomas Brochhagen, Maxime Cauté, Emmanuel Chemla, Milica Denić, Danny Fox, Michael Franke, Lilach Hadany, Aron Hirsch, Nur Lan, Ezer Rasin, and the audience at *Sinn und Bedeutung 26*.

and a matter of active research ever since.[2]

It seems tempting to reason from such typological asymmetries to properties of the representational system. On a particularly simplistic view, for example, one might take the attested meanings of logic words (e.g., 'some', 'every', 'no') to be possible stored denotations and everything else (e.g., 'every or no') as denotations that cannot be stored. However, this kind of mapping of lexicalizations to representations is not generally warranted. This is so because the representational system for stored semantic denotations is just one of several interacting factors that affect the typology.[3]

One important factor that complicates the reasoning from the typology to representations is language use. This factor concerns questions such as when speakers choose to use a word for a given (logical) term. For example, in a situation in which some but not all of the students arrived, what affects speakers' choice as to whether to use 'some', 'not all', or something else? The answer to this question can affect whether a given lexicalization is useful at all, and it is conceivable that denotations that are not useful will not be lexicalized even if they are perfectly representable. For example, it might perhaps be possible to store a binary sentential connective that returns *True* regardless of the truth values of its arguments, but even if that is the case it is not obvious that this connective will be useful to speakers, and it is plausible that this connective will be rare across languages.

The idea that useless denotations will be typologically rare points at a third important factor: learning. If the hypothetical binary connective that always returns *True* is rarely used, some learners might never be exposed to it, and the lexicalization will eventually be lost in the transmission of linguistic knowledge across generations. In this way learning interacts with usage to affect the typology. It also interacts with representations. Suppose, for example, that it is possible to represent a monomorphemic quantifier that means *some but not 217* through some complex combination of semantic building blocks. Justifying the acquisition of this complex and very specific quantifier from the perspective of the learner may well be nontrivial, and even if speakers use it, learners might fail to acquire it. Again, the result will be the loss of a lexicalization.[4]

The three factors mentioned above — representations, usage, and learning — are of course by no means the only factors that might affect the typology. Historical accident, sociolinguistic pressures, and various other considerations could also be relevant. However, representations, usage, and learning have been recognized as particularly important to the lexicalization asymmetries considered here, and in what follows we will focus exclusively on them.

[2]See von Fintel and Matthewson (2008), Szabolcsi (2010), Paperno (2011), and others for further discussion, and see Chemla et al. (2019) for important connections between generalizations for logic words and those for content words.

[3]In talking about representations we limit ourselves here to the representations involved in stored semantic knowledge. This is a gross oversimplification that ignores morpho-syntactic representations. We return to this matter briefly in section 5.

[4]See Paperno (2011), Magri (2015), and Steinert-Threlkeld and Szymanik (2019), among others, for proposals connecting learnability and the typology in the domain of quantificational determiners, and see Rasin and Aravind (2021) for an analysis of the actual linguistic input available to the child in this domain and its implications for possible learning mechanisms. For how iterations of learning across generations can affect language change and linguistic typology see Niyogi and Berwick (1997, 2009), Kirby (2001), and Kirby et al. (2004, 2007, 2008, 2015).

In order to reason about stored representations, then, it might be necessary to take into account all three factors and their interactions. The present note attempts to do so, focusing on a typological pattern noted by Horn (1972) and a particular view on stored representations growing out of Horn's original account of the pattern and developed further by Katzir and Singh (2013) and Uegaki (2021), a view that we will end up supporting. We present Horn's puzzle and the view on stored representations in section 2. In order to reason about stored representations we will use an evolutionary model by Brochhagen et al. (2018), who provide a framework for combining representations, communication, and learning in semantics. We present this model in section 3. In section 4 we use the evolutionary model to address three questions for the approach of Horn, Katzir and Singh, and Uegaki. In section 5 we discuss some of the many open questions for our account and consider some directions forward.

2. The puzzle

2.1. Horn's observation

Horn (1972) observes a striking typological asymmetry, which he frames in terms of the Aristotelian Square of Opposition, as schematized in Figure 1. Horn notes that while the *A* corner of the square (corresponding to 'every', 'and', and analogous operators) and the *I* corner (corresponding to 'some', 'or', and analogous operators) are often lexicalized and morphologically unmarked, the *E* corner (corresponding to 'no', 'neither/nor', and their like) is often not lexicalized (and is morphologically marked when it is), and the *O* corner is typically not lexicalized at all. The asymmetry, illustrated in Figure 1 and summarized below, appears to hold across languages and categories.

$$\begin{tikzcd} \forall,\neg\exists\neg & \mathbf{A} \arrow[r, dash] \arrow[d, dash] & ?\mathbf{E} \arrow[d, dash] & \forall\neg,\neg\exists \\ \neg\forall\neg,\exists & \mathbf{I} \arrow[r, dash] & *\mathbf{O} & \neg\forall,\exists\neg \end{tikzcd}$$

	A	**I**	**E**	**O**
connectives	and	or	**n**either ... **n**or	***n**and
2	both	either	**n**either	***n**both
quant. determiners	every	some	**n**o	***n**every
temp. adverbs	always	sometimes	**n**ever	***n**always
modals	must	may	***n**may (ca**n't**)	***n**must (nee**dn't**)

Figure 1: The Square of Opposition and Horn (1972)'s lexicalization generalization (diagram from Katzir and Singh 2013). While the *A* and *I* corners are routinely lexicalized as simplex, the *E* corner is morphologically marked, and the *O* corner is generally not lexicalized.

Horn's generalization. Across languages and categories of lexical operators, lexicalizations of the *A* and *I* corners are common and morphologically simple; lexicalizations of the *E* corner are less common and are often morphologically complex; and lexicalizations of the *O* corner are exceedingly rare.

2.2. An attempt to explain Horn's observation (H/KS/U)

One approach to the puzzle, already outlined by Horn (1972) and developed further by Katzir and Singh (2013) and Uegaki (2021) (H/KS/U), attributes the typological asymmetry to the interaction of three factors: communication (including scalar implicature), economy in lexical inventories (favoring simple inventories over complex ones), and specific cost asymmetries (which might be cashed out in various ways). More specifically, the I and O corners can both be strengthened through scalar implicature to mean $I \wedge O$ (e.g., 'some' and 'not all' can both be strengthened to mean $\exists \wedge \neg\forall$). This shared strengthening of I and O could make lexicalizing both corners redundant in some sense and dispreferred (or blocked categorically) by an economy condition on inventories. We will discuss this sense of redundancy shortly. Before that, however, note that even if an economy condition rules out lexicalizing both corners, there still remain two 3-corner inventories, $\{A,I,E\}$ and $\{A,E,O\}$, to choose from, with only the former being empirically attested. According to H/KS/U, this choice of $\{A,I,E\}$ over $\{A,E,O\}$ is due to I being less costly than O. Horn (1972) attributes the distinction in costs between I and O to a general preference for positivity, which favors I and A over O and E. Katzir and Singh (2013) and Uegaki (2021) cash out the preference for the positive corners over the negative ones in terms of decomposition, using specific assumptions about the innate primitives of stored semantic representations: A, I, and $\neg$ are taken to be primitives, while $E(=\neg I)$ and $O(=\neg A)$ are not.

Turning to the sense in which $\{A,I,E\}$ and $\{A,E,O\}$ compete with the 4-corner inventory $\{A,I,E,O\}$ and with each other, note that it will not do to simply let all inventories compete for simplicity: this would make all nonempty inventories lose to the empty one, and even if the empty inventory were somehow ruled out, the attested $\{A,I,E\}$ would lose to a smaller nonempty inventory such as $\{E\}$. The intuition mentioned above is that a particular notion of redundancy, defined with respect to meanings, plays a role. This notion of redundancy relies on the strengthening of operators (or sentences containing them) through scalar implicatures. The A and E corners do not get strengthened since they are maximally strong, but the I and the O corners both get strengthened to $I \wedge O$ ('some but not all'). Focusing exclusively on these strengthened meanings, the two 3-corner inventories $\{A,I,E\}$ and $\{A,E,O\}$ can be taken to be equally expressive.[5] Moreover, on certain assumptions the 4-corner inventory $\{A,I,E,O\}$ is comparable to each of the two 3-corner ones in terms of expressivity. This leads to the three inventories being valid competitors to one another and to $\{A,I,E\}$ being the winner based on economy. Moreover, it prevents a smaller inventory from preempting $\{A,I,E\}$: regardless of scalar implicatures, no smaller inventory can derive the same semantic coverage.

As Uegaki (2021) notes, a problem remains. While redundancy correctly predicts that a small unattested inventory such as $\{E\}$ does not preempt the attested (and bigger) $\{A,I,E\}$, nothing so far prevents $\{E\}$ from being a winner in its own right. Uegaki's remedy is to allow all inventories to compete with one another but make some inventories better than others in terms of their communicative success. Informally, an inventory is successful to the extent that it allows speakers to accurately communicate the relevant semantic distinctions in the domain under consideration. This more nuanced way of differentiating inventories is shared with recent work on the trade-off between complexity and informativity (see Kemp and Regier 2012, Zaslavsky et al. 2018, Steinert-Threlkeld 2019, 2021, Xu et al. 2020, and Denić et al. 2021), and as in that

[5] See Katzir and Singh (2013) and Enguehard and Spector (2021) for different ways to cash out this idea.

line of research Uegaki combines it with complexity through Pareto optimality: an inventory is taken to be optimal if it cannot be made more informative without increasing its complexity and if it cannot be made simpler without hurting its informativity. Focusing on the domain of binary sentential connectives and making certain further assumptions that we will not discuss here, Uegaki shows that Pareto optimization allows attested inventories such as $\{A,I\}$ and $\{A,I,E\}$ to win and prevents many unattested inventories such as $\{E\}$ and $\{A,I,E,O\}$ from winning.

2.3. Three questions for H/KS/U

The H/KS/U approach leaves open several issues that are important for a proper account of the pattern and its cognitive implications. In particular, it raises the following three questions.

Q1: Robustness. What might explain the appearance of an economy condition? H/KS/U propose relatively minor cost asymmetries and assume that those asymmetries lead to robust typological asymmetries. It does not seem plausible, however, that this mapping of costs to typology is mediated by a strict synchronic economy condition that rules out costly inventories. It would therefore be important to understand if cognitively plausible assumptions can give rise to this appearance of economy more indirectly.

Q2: Specific costs. Is the cost ranking of $A,I \ll E,O$ indeed warranted? H/KS/U do not systematically explore other possible cost patterns, and it is conceivable that other possibilities will lead to the observed typological pattern.

Q3: Usage. It has been argued recently by Enguehard and Spector (2021) that, separately from any matters of cost, there are usage asymmetries between meanings that might favor the attested $\{A,I,E\}$ over the unattested $\{A,E,O\}$. Given these usage asymmetries, do we still need to assume also cost asymmetries?

The three questions above concern storage costs and require an understanding of the role of these costs in shaping the typology. Following the discussion in the introduction we propose that storage costs affect the typology through learning. A given cost assignment makes some inventories cheap to store and other expensive. If learners are generally biased in favor of cheap inventories, this can result in pressure that differentiates between inventories in terms of how well they survive transmission over generations. In order to examine the effects of this cost-driven pressure and attempt to address the three questions for H/KS/U, we will model a learner with a simplicity bias explicitly as part of a larger model of cultural evolution. This larger model will also include usage factors, which counter in part the pressure for simple inventories by favoring inventories that support accurate communication (similarly to the effect of usage factors in Uegaki 2021's proposal). The concrete model that we will use, and with which we will attempt to address the three questions for H/KS/U, is the one developed by Brochhagen et al. (2018) and described in the next section.

3. The evolutionary model

The foundation for Brochhagen et al. (2018)'s model adopted here is the Replicator Mutator Dynamic (RMD; see Nowak et al. 2001, Page and Nowak 2002, Hofbauer and Sigmund 2003, Nowak 2006), a general framework for modeling adaptive systems such as biological and cultural evolution that has been used in a range of areas, including the study of language learning. Brochhagen et al. (2018) adapt the RMD to the communicative setting in semantics and pragmatics, and we base our work directly on their adaptation.

3.1. The RMD

Explicitly modeling the behavior of many individuals over generations is difficult. Instead, the RMD considers changes over generations to the *relative frequencies of types* in a population. In the present setting, a type is defined through its inventory of lexicalized logical operators, stated in terms of a meta-grammar for stored form-meaning pairings. For example, one type might have two stored pairings, the string 'xxx' paired with the denotation $\exists$, and 'yyy' paired with the denotation $\forall$. Another type might have three stored pairings, 'xxx' paired with the denotation $\forall$, 'yyy' with the denotation $\neg\exists$, and 'zzz' also with the denotation $\neg\exists$. The basic recurrence of the RMD relates type i's frequency in the next time step, x_i^{new}, to the frequencies of the different types in the current time step (x_j for each type j), the probability Q_{ji} of a learner hearing a speaker of type j and acquiring type i, and the communicative fitness f_j of type j in the current time step:

$$x_i^{new} = \frac{\sum_j Q_{ji} x_j f_j}{\Phi} \qquad \text{where } \Phi = \sum_k x_k f_k \text{ is the average fitness in the population.} \tag{1}$$

A type j that is currently common (high x_j), is currently communicatively successful (high f_j), and whose output is easily confused with that of type i (high Q_{ji}) will help raise the proportion x_i^{new} of type i in the next step.

The heart of the model concerns defining and implementing f_j and Q_{ji}. We briefly describe each in turn and refer the reader to Brochhagen et al. (2018) for further detail and discussion.

3.2. Communicative fitness

The communicative fitness of type i, f_i, is computed by ranging over all types j and evaluating how well type i and j communicate with one another — the expected utility between types i and j, $EU(i,j)$ for each j — weighted by x_j, the current frequency of type j:

$$f_i = \sum_j x_j EU(i,j) \tag{2}$$

If a type j is currently frequent (high x_j) and can communicate well with type i (high $EU(i,j)$), it improves the fitness of type i. The expected utility $EU(i,j)$ factors in both how well a speaker of type i can get their message across to a hearer of type j, $EU_{speaker}(i,j)$, and how well a speaker of type j can get their message across to a hearer of type i, $EU_{speaker}(j,i)$. We follow Brochhagen et al. (2018) in assuming that there is no inherent preference for any given type to be a speaker or a hearer: $EU(i,j) = \frac{1}{2}[EU_{speaker}(i,j) + EU_{speaker}(j,i)]$

As to $EU_{speaker}(i,j)$, this is defined by considering each state s of the world that the speaker (of type i) may wish to convey (with $P(s)$ the prior probability of states s), the message m they will choose to do so (given the speaker's distribution $S_i(\cdot|s)$ of messages given state s), the state s' that the hearer (of type j) will infer given the message m (given the hearer's distribution $H_j(\cdot|m)$ of states given message m), and how close the intended state s and the inferred state s' are ($\delta(s,s')$, here taken to be 1 exactly when $s = s'$ and 0 otherwise, but see Steinert-Threlkeld 2019 and Uegaki 2021 for more nuanced possibilities):

$$EU_{speaker}(i,j) = \sum_{s} P(s) \sum_{m} S_i(m|s;L_i) \sum_{s'} H_j(s'|m;L_j)\delta(s,s') \tag{3}$$

Speaker and hearer in the current implementation are modeled within the Rational Speech Act (RSA; Frank and Goodman 2012; Goodman and Stuhlmüller 2013; Bergen et al. 2016), a choice that offers a convenient approximation to scalar implicatures, which are a key aspect of the H/KS/U approach, as discussed above.[6] Within the RSA, speakers and hearers can be either fully literal or pragmatic at various levels of reasoning. **Literal hearers** assign probabilities to states given a message based only on the lexicon and the prior probabilities of states. Specifically, as summarized in (4), the probability $H_{lit}(s|m;L)$ that a literal hearer with lexicon L assigns to state s given message m is proportional to the product of two factors: (a) $P(s)$, the prior probability of state s; and (b) $L_{[s,m]}$, which is 1 if state s makes message m true and is 0 otherwise. This ensures that literal hearers spread their probability of states given a message m among those states that make m true and in a way that is proportional to the prior probability of those states that make m true. **Literal speakers** assign probabilities to messages given a state based only on the lexicon, as in (5): all messages that are false in a state receive the same low probability, and all messages that are true in that state receive the same higher probability.[7] The speaker's preference for using better messages is mediated by a so-called *rationality parameter* λ. When λ is high the speaker will tend to use the best messages and will assign a very low probability to messages that are not the best. With a low λ, the speaker will spread their probability mass over messages more evenly. In our simulations below we used a high rationality parameter: $\lambda = 20$. This choice is listed in Table 1, along with the values of the other parameters that we use.

$$H_{lit}(s|m;L) \propto P(s)\, L_{[s,m]} \tag{4}$$
$$S_{lit}(m|s;L) \propto exp(\lambda\, L_{[s,m]}) \tag{5}$$

Given the definitions of literal hearers and speakers we can now define pragmatic discourse participants. **Pragmatic hearers** assign probabilities to states given a message through Bayesian inference about a literal speaker. Specifically, as stated in (6), a pragmatic hearer's probability of state s given message m is proportional to the product of two factors: (a) $P(s)$, the prior probability of state s; and (b) $S_{lit}(m|s;L)$, the probability that a literal speaker assigns to message m given state s. **Pragmatic speakers** assign probabilities to messages given a state based on the utility of a literal hearer, as in (7).[8]

[6]We use the RSA for modeling convenience only. Recent work has provided evidence that even if something like the RSA is relevant for global disambiguation, accounting for scalar implicatures requires a grammatical exhaustivity operator (see Champollion et al. 2019, Franke and Bergen 2020, Fox and Katzir 2021, and Asherov et al. 2021). We hope that the simplification of using the RSA and not incorporating such an exhaustivity operator into the model does not affect our conclusions below.

[7]In principle, the speaker's probabilities can also incorporate costs reflecting their effort in making the utterance. These costs can depend at least in part on the morpho-syntactic representations made available by the grammar. Our current implementation does not incorporate speaker costs.

[8]Equations (6) and (7) generalize straightforwardly to the definition of higher-level pragmatic hearers and speakers, where a hearer of level $n+1$ bases their probability distribution on a speaker of level n, and a speaker of level $n+1$ bases their probability distribution on a hearer of level n. Such higher-level discourse participants are used in much of the RSA literature, but in line with Brochhagen et al. (2018) we will limit the pragmatic participants to those of the first level, with distributions that are based on literal participants.

$$H_{prag}(s|m;L) \propto P(s)\, S_{lit}(m|s;L) \quad (6)$$
$$S_{prag}(m|s;L) \propto \; exp(\lambda\, H_{lit}(s|m;L)) \quad (7)$$

For the purposes of our model — including both communicative fitness (equation (3) above) and learning (equation (8) below — we will assume that all discourse participants are pragmatic.

3.3. Learning

Q_{ji}, the probability of acquiring type i from a teacher of type j, is computed by considering the distribution over corpora by a pragmatic speaker of type j, written as t_j, and the probability that a learner exposed to each corpus will acquire type i. Each corpus consists of a sequence of k pairs $\langle s_l, m_l\rangle$ of a state s_l and message m_l, where the pairs are taken to be independent. The probability of a given corpus, then, is as follows, where S_j is the speaker's probability distribution over messages given states:

$$P(d = \{\langle s_1, m_1\rangle \ldots, \langle s_k, m_k\rangle\}|t_j) = \prod_{i=1}^{k} S_j(m_i|s_i)P(s_i) \quad (8)$$

In much of our discussion we will make the simplifying assumption that the prior over states is uniform. We will revisit this assumption in section 4.3 below, where we discuss the proposal of Enguehard and Spector (2021), which relies crucially on biased priors.

$Q_{j,i}$ is based on the distribution over corpora given speaker types as defined in (8) and on an inference to which we now turn in which the learner reasons from a given corpus to the types that might have generated it. For each corpus d, generated with probability $P(d|t_j)$ by a speaker of type j as described above, the learner attempts to infer the generating type. The probability $P(t_i|d)$ of d being generated by type i is obtained using Bayes' Rule, factoring in both the prior probability $P(t_i)$ of type i (as described below) and the likelihood $P(d|t_i)$ of d being generated by type i (as summarized in equation (8) above). Following Brochhagen et al. (2018), a learning parameter l is used to determine the learner's behavior: when $l = 1$, the learner samples from the posterior probability, and as l increases the learner becomes more inclined to choose the type that maximizes the posterior.[9]

$$Q_{j,i} = \sum_{d \in D} P(d|t_j) \quad \underbrace{P(t_i|d)}_{\propto [P(t_i)P(d|t_i)]^l} \quad (9)$$

Turning to the prior probability over types, Brochhagen et al. (2018) assume that the learner prefers types with short lexica (see Piantadosi et al. 2012, 2016 and Katzir et al. 2020 for learners in this domain that follow a similar principle). The specific prior $P(t_i)$ that Brochhagen et al. (2018) choose and that we inherit is as follows. The prior probability of the lexicon of type i is defined through the product of the prior probability of the individual words in t_i's lexicon, as stated in (10). The prior probability of an individual word w, in turn, is defined as in (11) by the difference between its storage cost and the maximal storage cost of a word in the lexicon: the higher the storage cost of w, the lower its prior.

[9]See Griffiths and Kalish (2007) for analysis of the two choices, and see Brochhagen et al. 2018 for a study of interpolations between the two.

$$P(t_i) \propto \prod_{w \in Lex(t_i)} P(w) \tag{10}$$

$$P(w) \propto \max_{w' \in Lex(t_i)} (cost(w')) - cost(w) + 1 \tag{11}$$

3.4. Parameters
In all the simulations reported below the parameters were as follows:

parameter	value
λ	20
k (sequence length)	5
$\|D\|$ (samples per type)	50
learning parameter	15
generations	at least 50 [10]
number of independent simulation runs	50

Table 1: General setting

4. Using the RMD to probe the three questions

Recall the three questions for H/KS/U, repeated here:

Q1: Robustness. What might explain the appearance of an economy condition? H/KS/U propose relatively minor cost asymmetries and assume that those asymmetries lead to robust typological asymmetries. It does not seem plausible, however, that this mapping of costs to typology is mediated by a strict synchronic economy condition that rules out costly inventories. It would therefore be important to understand if cognitively plausible assumptions can give rise to this appearance of economy more indirectly.

Q2: Specific costs. Is the cost ranking of $A, I \ll E, O$ indeed warranted? H/KS/U do not systematically explore other possible cost patterns, and it is conceivable that other possibilities will lead to the observed typological pattern.

Q3: Usage. It has been argued recently by Enguehard and Spector (2021) that, separately from any matters of cost, there are usage asymmetries between meanings that might favor the attested $\{A, I, E\}$ over the unattested $\{A, E, O\}$. Given these usage asymmetries, do we still need to assume also cost asymmetries?

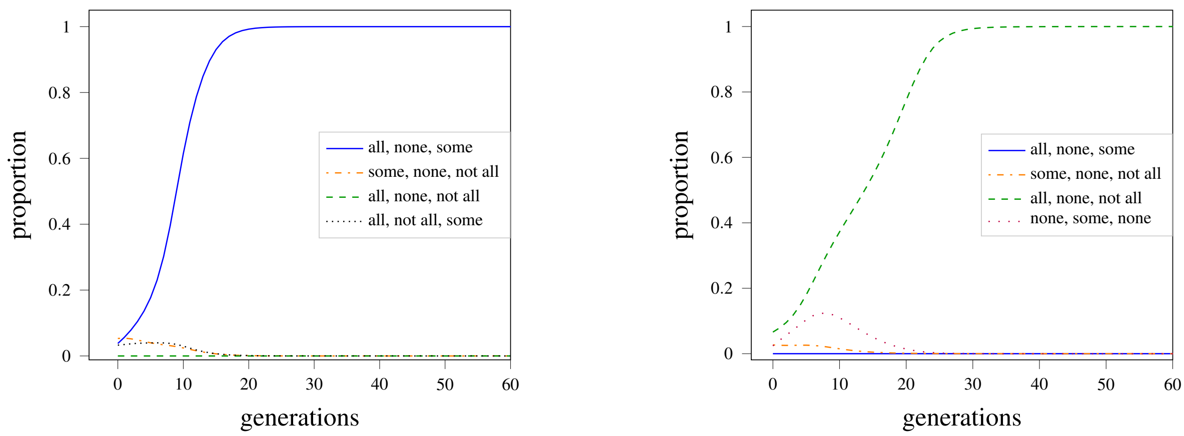
With the help of the RMD we can now attempt to address these questions.[11]

4.1. Q1: Why should minor cost asymmetries lead to robust typological asymmetries?
Probing the first question, regarding robustness, is the most straightforward: As has been observed in the literature, even very small biases at the level of the individual can be amplified through multiple iterations of transmission to robust population-level asymmetries (see Kirby et al. 2007).

[10]The simulation ends if the winning inventory remains the same and has the same proportion for 10 consecutive generations.
[11]The code to all the simulations below is at `https://github.com/taucompling/rmd_square`.

In more detail, consider again the learning component in the RMD, from (9): $Q_{j,i} = \sum_{d \in D} P(d|t_j) \underbrace{P(t_i|d)}_{\propto [P(t_i)P(d|t_i)]^l}$. Recall that the prior probability of a type, $P(t_i)$, is defined via lexicon complexity, as in (10) and (11). This prior favors less costly lexica over costlier ones. The strength of this preference depends on the actual costs of the lexica under consideration, but even when it is relatively weak, its effect on $Q_{j,i}$ can be expected to lead over sufficiently many generations to a robust typological preference for simplicity.

The amplification of the potentially small bias at the level of an individual to a robust one at the population level is illustrated in Figure 2, which shows two cases in which a minor cost asymmetry leads to a single type taking over the population. When the costs are $A(8)$, $I(8)$, $O(10)$, $E(10)$ (so $A, I \ll O, E$, as in the H/KS/U approach), the language type winner over 60 generations is the one with *some, all, none*, as shown in Figure 2a. With small costs changes ($A(10)$, $I(10)$, $O(8)$, $E(8)$, so $E, O \ll A, I$) the winner type changes to *all, not all, none*, as shown in Figure 2b.



(a) With the costs $A(8)$, $I(8)$, $O(10)$, $E(10)$, the language type winner is that containing a lexicon with *some, all, none*.

(b) With small costs changes ($A(10)$, $I(10)$, $O(8)$, $E(8)$) the winner type changes to *all, not all, none*.

Figure 2: Amplification of small cost asymmetries to robust typological asymmetries. The graphs show the change of the proportions of several language types over multiple generations using two different cost patterns.

4.2. Q2: What costs lead to the attested typology?

The second question, concerning what costs matter, requires looking at a range of cost assignments and checking the predicted typology according to each assignment against the attested pattern of lexicalization. Table 2 summarizes the results of such a test. The inventory of the winner type is shown in the second column depending on the relative costs (first column). The average proportion of this winner type after at least 50 generations is stated in column 3, and the the percentage of the rounds where this type ends up winning is listed in the last column. As table 2 shows, two distinct cost patterns yield the attested typology in the case of 3-corner inventories: (1) $A, I, E \ll O$; and (2) $A, I \ll E, O$.

Significantly, however, the tie between the two cost patterns in the case of 3-corner invento-

relative costs	inventory of winner type	proportion of winner type	runs won
A, I, E, O	$A, I(/O), E$	50.0%	51.0 %
$A, I, O \ll E$	$A, I(/O), E$	49.15 %	50.6 %
$E, I, O \ll A$	$A, I(/O), E$	50.01 %	49.52 %
$A, E \ll I, O$	A, O, E	70.0%	60.0 %
$A, I, E \ll O$	**A, I, E**	100.0%	99.7 %
$A, I \ll E, O$	**A, I, E**	100.0%	100.0 %

Table 2: Quantificational determiners with three corners. For the cost rankings in the first three rows of the table the population at the end of the simulations is divided almost evenly between the the inventories $\{A, I, E\}$ and $\{A, I, O\}$. This is abbreviated in the table as $A, I(/O), E$. In the cost rankings schematized in the last two rows of the table, the attested 3-corner inventory takes over the entire population in almost every run.

ries does not carry over to the case of 2-corner inventories, where as far as we can tell the only attested inventory is $\{A, I\}$. While in the 3-corner case both the H/KS/U cost pattern of $A, I, E \ll O$ and the cost pattern $A, I \ll E, O$ gave rise to the attested inventory $\{A, I, E\}$, only the former yields the attested 2-corner inventory $\{A, I\}$ whereas the latter yields the unattested 2-corner inventory $\{I, E\}$. Table 3 compares the two cost patterns in the 3-corner and the 2-corner case.

inventory size	relative costs	inventory of winner type	prop. of winner type	runs won
3	$A, I, E \ll O$	**A, I, E**	100.0 %	99.7 %
3	$A, I \ll E, O$	**A, I, E**	100.0 %	100.0 %
2	$A, I, E \ll O$	I, E	100.0 %	84.2 %
2	$A, I \ll E, O$	**A, I**	100.0 %	100.0 %

Table 3: Quantificational determiners with three corners (first two rows) and with two corners (last two rows). While two different cost rankings yield the attested 3-corner inventory, only one ranking (corresponding to the the H/KS/U costs) leads to the attested 2-corner inventory.

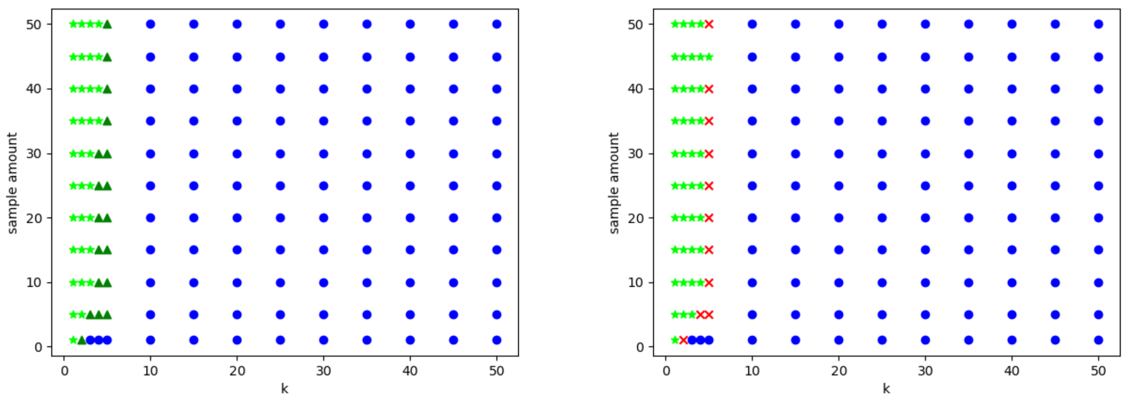
Differently from 3- and 2-corner inventories, it is unclear whether there are any attested 1-corner inventories. Based on the general markedness asymmetries stated in Horn's generalization and on analogy with the typological evidence for sentential connectives as analyzed by Uegaki (2021), it seems reasonable to expect that if there are any true 1-corner inventories they would be either $\{A\}$ or $\{I\}$ and never $\{E\}$ or $\{O\}$. On this — admittedly speculative — assumption, 1-corner inventories, too, help choose between the H/KS/U cost pattern $A, I \ll E, O$ and the competing cost pattern $A, I, E \ll O$. As table 4 shows, the former cost pattern predicts only $\{A\}$ and $\{I\}$ among the 1-corner inventories, while the latter cost function predicts also $\{E\}$.

Tables 3 and 4 illustrated the winning inventories of sizes 2 and 1 in an artificial way: for these results we ran the RMD under an explicit restriction of the types to those of the desired inventory size. While at present we do not have a fully organic way to obtain winning types of different sizes, we observe that a solution that is at least somewhat less artificial than explicit

relative costs	inventory of winner type	proportion of winner type	runs won
A, I, E, O	$A(/I/O/E)$	25.0 %	100.0 %
$A, I, O \ll E$	$A(/I/O)$	33.3%	100.0 %
$E, I, O \ll A$	$I(/E/O)$	33.3 %	100.0 %
$A, E \ll I, O$	$A(/E)$	50.0 %	100.0 %
$A, I, E \ll O$	$A(/I/E)$	33.3 %	100.0 %
$A, I \ll E, O$	**A(/I)**	50.0 %	100.0 %

Table 4: Quantificational determiners with one corner. With all cost rankings the population is split almost evenly between two or more types. This is written in the table in abbreviated form using slashes. E.g., $A(/I/O/E)$ in the first row corresponds to an even split between all four 1-corner inventories. On the assumptions discussed in the text, only the H/KS/U costs lead to acceptable predictions.

size restrictions is offered by varying the k parameter that determines the sample size for the learner. Figures 3a and 3b show the result of varying k for the two cost patterns under consideration. The emerging picture is similar to the one from the artificial restriction of inventory sizes: under the H/KS/U cost pattern of $A, I \ll E, O$ we obtain the attested inventories of the different sizes, while under the alternative cost pattern of $A, I, E \ll O$ we do not.



(a) Cost assignment: $A(8)$, $I(8)$, $O(10)$, $E(10)$ (b) Cost assignment: $A(8)$, $I(8)$, $O(10)$, $E(8)$

Figure 3: Language type winners depending on sequence length k and sample amount. Lime stars represent populations that are almost evenly split between the two 1-corner inventories $\{A\}$ and $\{I\}$. Green triangles represent the 2-corner inventory $\{A, I\}$. Blue dots represent the 3-corner inventory $\{A, I, E\}$. Red cross marks represent other inventories.

The results for 1- and 2-corner inventories address a key aspect of Horn (1972)'s generalization: that lexicalizations of the A and I corners are common while lexicalizations of the E corner are less common. On the H/KS/U cost pattern of $A, I \ll E, O$, our model derives a strong version of this distributional asymmetry between the corners, one in which the lexicalization of E implies the lexicalization of both A and I, while either of the latter can be lexicalized on its own. However, there is a related aspect of Horn's generalization that our account fails to address: that lexicalizations of the A and I corners appear to be morphologically unmarked while lexicalizations of the E corner appear to be morphologically marked (and might in fact be morpho-syntactically complex; see Sauerland 2000). This morphological asymmetry is handled by stipulation by Katzir and Singh (2013) and is often set aside elsewhere in the literature. We leave this matter open here.

4.3. Q3: Can usage asymmetries account for the typological pattern?

Turning to the third question, regarding usage instead of costs, Enguehard and Spector (2021) suggest that I gets lexicalized instead of O simply because it happens to be more useful. Specifically, they make the case that I is more informative than O and that speakers generally prefer to use more informative statements when possible. On certain further assumptions, this leads to I being used more often than O. This, in turn, means that if languages need to choose between lexicalizing I and O, they will be better off lexicalizing the former. If successful, the usage-based perspective could offer an alternative explanation that does away with cost asymmetries for semantic representations.

Enguehard and Spector (2021) note that in their version, the usage-based proposal only helps in making the narrow choice between the lexical inventories $\{A,E,I\}$ and $\{A,E,O\}$ (where the missing corner in each case is expressed analytically). In particular, there is no account of why there is no lexicalization of all four corners ($\{A,E,I,O\}$) or of the status of 2-corner or 1-corner inventories. The current framework, however, allows us to explore Enguehard and Spector (2021)'s direction further. For example, we find that for 3-corner inventories, usage asymmetries indeed lead to the correct typological results (table 5): if $P(\forall) < P(\neg\exists)$, as Enguehard and Spector (2021) assume, the correct $\{A,I,E\}$ inventory wins, and in any event, the 4-corner $\{A,E,I,O\}$ is often ruled out (see Figure 4b below).

usage prob.	inventory of winner type	proportion of winner type	runs won
$P(\neg\exists) = P(\forall)$	$A, I(/O), E$	50.02 %	50.00 %
$P(\neg\exists) < P(\forall)$	A, O, E	100.0 %	87.47 %
$P(\forall) < P(\neg\exists)$	**A, I, E**	100.0 %	87.44 %

Table 5: Inventories (3 corners) by usage probabilities when storage costs are uniform. In the first row the population is roughly evenly split between the inventories $\{A,I,E\}$ and $\{A,I,O\}$. The attested 3-corner inventory wins when the usage probabilities are as in Enguehard and Spector (2021).

Our second finding is that costs trump usage: if both usage and cost asymmetries are assumed, usage asymmetries are irrelevant (table 6).

Our third and most significant finding with respect to Enguehard and Spector (2021) is that usage does not work for 2-corner inventories: differently from the 3-corner case, in the 2-corner case the attested inventory $\{A,E\}$ does not arise on any of the relevant usage rankings when the storage costs are uniform. This is shown using a restriction to two corners in table 7, where Enguehard and Spector (2021)'s usage probabilities lead to the unattested inventory $\{I,E\}$ and where the attested 2-corner inventory $\{A,I\}$ does not emerge, regardless of usage. Figure 4a illustrates the emergence of the unattested inventory $\{I,E\}$ across different values of k and

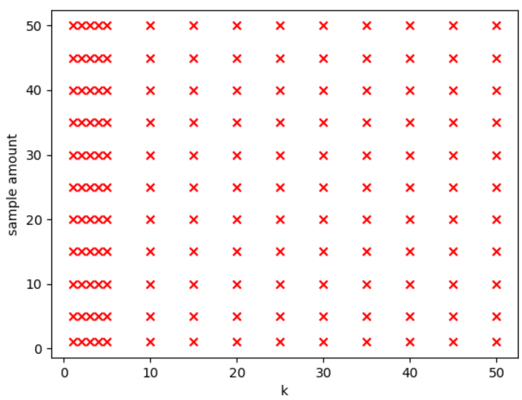
relative costs	usage	inventory of WT	proportion of winner type	runs won
$A,I \ll E,O$	$P(\neg\exists) < P(\forall)$	**A, I, E**	98.73 %	92.0 %
$A,I \ll E,O$	$P(\forall) < P(\neg\exists)$	**A, I, E**	99.54 %	96.0 %
$E,O \ll A,I$	$P(\neg\exists) < P(\forall)$	A,O,E	98.53 %	98.0 %
$E,O \ll A,I$	$P(\forall) < P(\neg\exists)$	A,O,E	100.0 %	98.0 %

Table 6: Inventories (3 corners) by usage probabilities when storage costs vary. Usage probabilities do not overcome biases induced by storage costs.

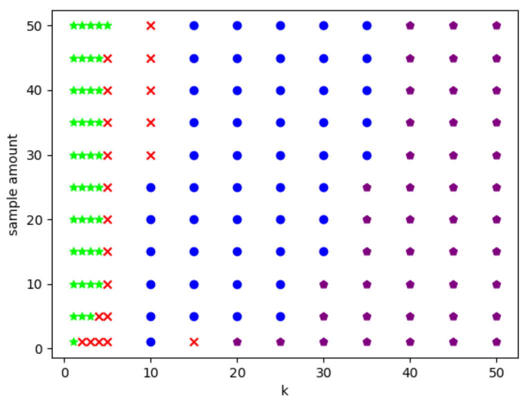
sample size when the inventory size is fixed at 2. In Figure 4b, where in addition to 2-corner inventories also 1-, 3-, and 4-corner inventories are allowed — while usage probabilities follow Enguehard and Spector 2021 and storage costs are uniform — we see again that incorrect inventories are predicted (including unattested 3-corner inventories for various parameter values) and the attested inventory of $\{A,I\}$ does not arise. We can conclude that regardless of usage, we still need the H/KS/U cost asymmetries $A,I \ll E,O$.

usage prob.	inventory of winner type	proportion of winner type	runs won
$P(\neg\exists) = P(\forall)$	A,E	100.0 %	40.44 %
$P(\neg\exists) < P(\forall)$	A,O	100.0 %	83.29 %
$P(\forall) < P(\neg\exists)$	I,E	100.0 %	83.73 %
$P(\forall) \ll P(\neg\exists)$	I,E	100.0 %	71.49 %

Table 7: Inventories of two corners by usage probabilities when storage costs are uniform. The attested 2-corner inventory does not arise.



(a) Results for 2-corner-inventories.



(b) Results for 1-, 2-, 3- and 4-corner inventories.

Figure 4: Language type winners depending on sequence length k and sample amount with equal costs for A,I,E,O but with usage probability $P(\forall) < P(\neg\exists)$. Lime stars represent the 1-corner inventory $\{A\}$. Green triangles represent the 2-corner inventory $\{A,I\}$. Blue dots represent the 3-corner inventory $\{A,I,E\}$. Purple pentagon represents the full 4-corner inventory. Red cross marks represent other inventories.

5. Remaining issues

The discussion above illustrates how the proposed framework, even in the present highly simplified setting, makes it possible to improve upon existing accounts of a typological pattern of lexicalization and probe the pattern's cognitive implications (in this case, implications regarding the relative storage costs of the quantificational determiners on the Square of Opposition).

Obviously, this preliminary investigation leaves open many important issues. Some limitations, such as our model's inability to derive 1- and 2-corner inventories organically, have already been noted above. Below is a far from exhaustive list of further open questions that we hope to come back to in future work.

Morpho-syntax I: the negative corners. In section 4.2 we noted that our model does not explain why actual lexicalizations of the *E* corner appear to be morphologically marked (and might be morpho-syntactically complex, as argued by Sauerland 2000). A related question is whether the two negative corners, *E* and *O*, can at all be lexicalized as morphologically simplex, at least in principle. Starting from Horn (1972), much of the discussion of the empirical puzzle assumes that simplex lexicalizations of the negative corners are indeed possible.[13] It is unclear, however, whether this is indeed the case; and if it is not, we are not modeling the right thing. Given that actual lexicalizations in natural languages do not seem to be helpful in this respect, one would need other sources of evidence. A source of evidence that seems particularly promising is data from artificial-grammar learning experiments. To date, however, the results available from such experiments appear inconclusive (see Hunter and Lidz 2013 and Spenader and de Villiers 2019 for data and discussion).

Morpho-syntax II: the broader context. We have assumed throughout that quantifiers appear in simple sentences that do not involve other operators. As far as our discussion above is concerned, utterances might be "Some dogs bark" or "Every dog barks" but not "Some dog barks and it is not the case that every dog barks". But a more realistic grammatical setting with sentences that can include multiple operators makes it possible to convey messages that might be impossible based on a small lexicon in the present setting. This means that what looks like a hopelessly impoverished inventory in terms of the present model could end up being communicatively viable within a more realistic setting. At the same time, the relevant messages will now start varying in length, which might affect the speaker's will to use them, another nuance that is lost on the current model. The broader morpho-syntactic context can affect the typology in other ways as well. For example, some denotations might be possible in principle but ruled out as actual lexicalizations due to interactions with the morpho-syntax. This might be the case for meanings such as the quantificational counterpart of 'outnumber' (see Fox 2002).

Adequacy of the evolutionary model. We have assumed that the RMD is an adequate model for language change and that its predictions with respect to the emergent typology are meaningful. This assumption, however, requires support. Like all models, the RMD incorporates various simplifications. In the present case, the assumption that the types in the population are homogeneously spread, so every type can converse with every type, is one such sim-

[13] An exception is Katzir and Singh (2013), who stipulate that only the *A* and *I* corners can be lexicalized as morphologically simplex, while the *E* and *O* corners are morpho-syntactically complex and include a negation node.

plification: actual human populations are of course concentrated quite non-homogeneously. Another simplification is that learners acquire their grammar from a single teacher, again contrary to the actual human setting. It is quite possible that some of those simplifications affect the predictions of the model in important ways (see, e.g., Niyogi and Berwick 2009 for how the choice of single- vs. multiple-teacher acquisition can dramatically affect the predicted dynamic of language change).

Communicative setting. Another simplification of the current model concerns the knowledge and aims of the conversational participants. In particular, we have assumed that the speaker knows the exact state of the world and wishes to convey it perfectly to the hearer. But not all situations are like that, and deviations from this idealization might affect the typology. A connective like 'or', for example, might be useful exactly because of its utility in situations where the speaker does not know which of the two disjuncts is true or does not care if the hearer finds out. Incorporating a more sophisticated model of the participants' epistemic states and conversational goals seems important.

6. Summary

We presented a preliminary attempt to reason about stored representations in semantics using the typology of lexicalized logic words. We noted that reasoning directly from the typology to the representational system is complicated by the interaction of representations with usage and learning. In order to address this complication we modeled the three factors and their interaction through a specific evolutionary model, namely Brochhagen et al. (2018)'s adaptation of the Replicator-Mutator Dynamic. Our empirical focus was Horn (1972)'s puzzle concerning the Square of Opposition. Using the evolutionary model we showed that, on certain assumptions, a particular proposal in the literature (H/KS/U) concerning the relative costs of stored representations, according to which 'some' and 'all' are less costly to store than 'no' and 'not all', is supported. We used the model to address three open questions for the H/KS/U view. First, the model explained how a small cost asymmetry in the lexicon translates into a robust typological asymmetry in lexicalization. Second, the model confirmed that only the specific cost asymmetry that has been proposed yields the attested pattern, while other cost asymmetries make incorrect predictions. Finally, the model allowed us to investigate the relative roles of cost asymmetries and usage asymmetries in deriving the typology, with evidence that the latter cannot replace the former. We noted that our work leaves several major issues open, which makes our conclusions highly tentative.

References

Asherov, D., D. Fox, and R. Katzir (2021, October). Reference games and the nature of exhaustification. Ms., MIT and TAU, available at https://ling.auf.net/lingbuzz/006257.

Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy 4*, 159–219.

Bergen, L., R. Levy, and N. Goodman (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics 9*(20).

Brochhagen, T., M. Franke, and R. van Rooij (2018). Coevolution of lexical meaning and pragmatic use. *Cognitive Science 42*(8), 2757–2789.

Champollion, L., A. Alsop, and I. Grosu (2019). Free choice disjunction as a rational speech act. In *Proceedings of SALT 29*, pp. 238–257.

Chemla, E., B. Buccola, and I. Dautriche (2019). Connecting content and logical words. *Journal of Semantics*.

Denić, M., S. Steinert-Threlkeld, and J. Szymanik (2021). Complexity/informativeness trade-off in the domain of indefinite pronouns. In *Semantics and linguistic theory*, Volume 30, pp. 166–184.

Enguehard, É. and B. Spector (2021). Explaining gaps in the logical lexicon of natural languages: A decision-theoretic perspective on the square of Aristotle. *Semantics and Pragmatics 14*, 5.

Fox, D. (2002). Antecedent contained deletion and the copy theory of movement. *Linguistic Inquiry 33*(1), 63–96.

Fox, D. and R. Katzir (2021). Notes on iterated rationality models of scalar implicatures. *Journal of Semantics 38*(4), 571–600.

Frank, M. C. and N. D. Goodman (2012). Predicting pragmatic reasoning in language games. *Science 336*(6084), 998.

Franke, M. and L. Bergen (2020). Theory-driven statistical modeling for semantics and pragmatics: A case study on grammatically generated implicature readings. *Language 96*(2), e77–e96.

Goodman, N. D. and A. Stuhlmüller (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science 5*(1), 173–184.

Griffiths, T. L. and M. L. Kalish (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science 31*(3), 441–480.

Hofbauer, J. and K. Sigmund (2003). Evolutionary game dynamics. *Bulletin of the American Mathematical Society 40*(4), 479–519.

Horn, L. (1972). *On the Semantic Properties of the Logical Operators in English*. Ph. D. thesis, UCLA.

Hunter, T. and J. Lidz (2013). Conservativity and learnability of determiners. *Journal of Semantics 30*(3), 315–334.

Katzir, R., N. Lan, and N. Peled (2020). A note on the representation and learning of quantificational determiners. In M. Franke, N. Kompa, M. Liu, J. L. Mueller, and J. Schwab (Eds.), *Proceedings of Sinn und Bedeutung 24*, Volume 1, Osnabrück/Berlin, pp. 392–410. Osnabrück University and Humboldt University of Berlin.

Katzir, R. and R. Singh (2013). Constraints on the lexicalization of logical operators. *Linguistics and Philosophy 36*(1), 1–29.

Kemp, C. and T. Regier (2012). Kinship categories across languages reflect general communicative principles. *Science 336*(6084), 1049–1054.

Kirby, S. (2001). Spontaneous evolution of linguistic structure — an iterated learning model of the emergence of regularity and irregularity. *Evolutionary Computation, IEEE Transactions on 5*(2), 102–110.

Kirby, S., H. Cornish, and K. Smith (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences 105*(31), 10681.

Kirby, S., M. Dowman, and T. L. Griffiths (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences 104*(12), 5241–5245.

Kirby, S., K. Smith, and H. Brighton (2004). From UG to universals. *Studies in Language 28*(3), 587–607.

Kirby, S., M. Tamariz, H. Cornish, and K. Smith (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition 141*, 87–102.

Magri, G. (2015). Universal restrictions on natural language determiners from a PAC-learnability perspective. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimini, T. Matlock, C. J. Jennings, and P. P. Maglio (Eds.), *Proceedings of the 37th annual meeting of the Cognitive Science Society*, pp. 1494–1499.

Niyogi, P. and R. C. Berwick (1997). Evolutionary consequences of language learning. *Linguistics and Philosophy 20*(6), 697–719.

Niyogi, P. and R. C. Berwick (2009). The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences 106*, 10124–10129.

Nowak, M., N. Komarova, and P. Niyogi (2001). Evolution of Universal Grammar. *Science 291*, 114–118.

Nowak, M. A. (2006). *Evolutionary dynamics*. Harvard University Press.

Page, K. M. and M. A. Nowak (2002). Unifying evolutionary dynamics. *Journal of Theoretical Biology 219*(1), 93–98.

Paperno, D. (2011). Learnable classes of natural language quantifiers: Two perspectives. Ms., UCLA.

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

Piantadosi, S. T., J. B. Tenenbaum, and N. D. Goodman (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological review 123*(4), 392–424.

Rasin, E. and A. Aravind (2021). The nature of the semantic stimulus: the acquisition of every as a case study. *Natural Language Semantics 29*(2), 339–375.

Sauerland, U. (2000). No 'no': On the crosslinguistic absence of a determiner 'no'. In *Proceedings of the Tsukuba Workshop on Determiners and Quantification*, pp. 415–444.

Spenader, J. and J. de Villiers (2019). Are conservative quantifiers easier to learn? evidence from novel quantifier experiments. In J. J. Schlöder, D. McHugh, and F. Roelofsen (Eds.), *Proceedings of the 22nd Amsterdam Colloquium*, pp. 504–512.

Steinert-Threlkeld, S. (2019). Quantifiers in natural language optimize the simplicity/informativeness trade-off. In J. J. Schlöder, D. McHugh, and F. Roelofsen (Eds.), *Proceedings of the 22nd Amsterdam colloquium*, pp. 513–522.

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

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

Szabolcsi, A. (2010). *Quantification*. Cambridge University Press.

Uegaki, W. (2021). The informativeness/complexity trade-off in the domain of Boolean connectives. To appear, Linguistic Inquiry.

von Fintel, K. and L. Matthewson (2008). Universals in semantics. *The Linguistic Review 25*(1–2), 139–201.

Xu, Y., E. Liu, and T. Regier (2020). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind 4*, 57–70.

Zaslavsky, N., C. Kemp, T. Regier, and N. Tishby (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences 115*(31), 7937.