Tel Aviv University
The Lester and Sally Entin Faculty of Humanities
Department of Linguistics

A Frequency Based Model for Lexicalization of Logical Operators

MA thesis submitted by:
Adam Rimon

Thesis advisor:
Dr. Roni Katzir

November 2018

Contents

1 Introduction **1**

2 Katzir and Singh's condition **2**

2.1 Issues with the economy condition 4

2.2 Issues with the maximal coverage motivation 5

3 Usefulness and frequency based theory **5**

4 Proposal **6**

4.1 Lexicalization of frequently used constructs 7

4.2 Relations conveyed by structures 8

4.3 Tendency towards simpler structures 10

4.4 Usefulness pattern 12

4.5 Deriving the attested lexicalization pattern 13

5 Computational model and simulation **14**

5.1 The process 15

5.2 Simulation 16

6 Discussion **16**

6.1 The basic operators 17

6.2 The usefulness pattern 17

Abstract

Horn (1972) observed the peculiar absence of the *not-all* quantifier from the lexicon of all studied languages and suggested that it is so because a lexicalization for it would be redundant, as *some* has *not-all* as an implicature. He also argued that even though the redundancy could also be solved by lexicalization of *not-all* and the absence of *some*, languages "choose" lexicons with less negations.

Katzir and Singh (2013) revisited Horn's puzzle, created a generalization that could capture lower-order and higher-order logical operators more easily and proposed a more explicit system that predicts the absence of the *not-all* quantifier. Yet, their system lacks some cognitive motivations, and makes some wrong predictions.

I suggest a different, frequency-based approach to the puzzle, based on the interaction between usage frequency effects, tendency toward simplicity, and the primitives used for encoding the operators. I propose a computational model that seems cognitively motivated and accounts for the lexicalization pattern observed by Horn.

Acknowledgments

Before presenting my thesis, I would like to express my gratitude to some of the people who helped me through my journey as a linguistics student.

First and foremost, I would like to thank my advisor, Dr. Roni Katzir. His patience and respect for any idea brought before him inspired me and played a profound part in shaping me as a linguist and researcher. Roni helped and guided me while writing my thesis with his vast knowledge and his endless ability to yield helpful references, and I am very grateful for that.

I would also like to thank Professor Tal Siloni. She shaped my experience as a linguistics student since the beginning, and provided counsel whenever needed. I am deeply thankful for her support and advice.

I am very thankful to Professor Mira Ariel, Dr. Nirit Kadmon, and Professor Fred Landman for being available for any question, providing insights and broadening my perspective when tackling different aspects of this thesis.

I extend my gratitude to other faculty members of the Lingusitics Department whose courses exposed me to many different angles of the linguistics study: Professor Outi Bat-El, Dr. Evan Gary Cohen, Dr. Aya Meltzer-Asscher and Professor Julia Horvath.

I would also like to express my gratitude to the faculty and staff members of the Adi Lautman Interdisciplinary Program for Outstanding Students, for the opportunity they have given me to broaden my horizons and learn so many different things that shaped my thought.

I am greatly indebted to my colleague and friend Daniel Asherov. His friendship, support and advice deeply helped me in this journey. I thank him for his insights and his encouragement, and for helping me organize my thoughts when my thesis has seemingly arrived at a dead end.

Finally, I would like to thank my life partner Yael Firer, for her support, interest, and insights. Her sharp mind, enthusiasm and love helped me through this journey both academically and personally.

1 Introduction

Horn (1972) observed a peculiar absence of a set of lexicalized logical elements from all researched languages. It is framed in terms of the traditional "Square of Opposition" which contains the quantifiers *every/all* and *some* along with their negations *no* and *not-every* at the corners named A, I, E and O respectively (presented in a square as in Figure 1 below). In those terms, the observation states that the O corner (*not-every*) is never lexicalized - it is never represented by one single word in any natural language. The E is also rarely lexicalized, so logical inventories in languages consist of the other corners of the square: either $\{A, I, E\}$ or $\{A, I\}$.[1]

$$\begin{tikzcd}
\textit{all} & A \arrow[r, dash] \arrow[d, dash] & E \arrow[d, dash] & \textit{no} \\
\textit{some} & I \arrow[r, dash] & O & \textit{not all}
\end{tikzcd}$$

Figure 1: Square of Opposition

In order to explain the absence of a lexicalization of O (*not all*), Horn has proposed conditions on the ability to lexicalize logical operators, followed by Katzir and Singh (2013) who phrased the conditions in a more explicit and generalized manner. Horn's idea was based on scalar implicatures; when (1a) is uttered (1b) is also understood (by implicature).

(1) a. Some cats like fish
    b. Not all cats like fish

The implicature makes it possible for a speaker wanting to express a *some but not all* (referred to as Y) relation to do so using the single word *some* (E). We could say informally that much of the time expressing Y instead of O seems to be good enough, and therefore an inventory of $\{A, I, E\}$ is enough for expressing each corner of the square using a single word (A and E for expressing themselves, and I for expressing both I and O) and there is no need for a dedicated lexicalization for the negation of *all* (O).[2]

The same argument could predict the validity of the unattested inventory of

[1] In some rare cases languages have even smaller inventories of $\{A\}$ or $\{I\}$, but I'll set them aside in this thesis.

[2] The redundancy of O is not fully explained by Horn (1972), and there is no account for how and why we can ignore the meaning of O that is not covered by Y (namely, *not all and maybe no*). Katzir and Singh (2013) later explain how the meaning is ignored, but do not deal with the "why".

$\{A, E, O\}$, as the *not all* (A) expression has *some* (E) as an implicature (when (1b) is uttered (1a) is understood), but no language seems to have an inventory like the second one (at least, none of the researched languages has it). To account for this difference between I and O, Horn has suggested that languages prefer to lexicalize *some* rather than *not all* because negativity in general is "marked", and *not all* is negative and thus more "marked" than *some*.

2 Katzir and Singh's condition

Following Horn, Katzir and Singh (2013) offer a system (K&S henceforth) consisting of an explicit lexicalization mechanism explaining how lexicalization is done, and a single condition which captures some of the intuitions of Horn. That system successfully predicts the validity of the attested inventories of $\{A, I, E\}$ and $\{A, I\}$ while invalidating other inventories such as $\{A, I, E, O\}$.

The Lexicalization Mechanism (Constraint on Lexicalization in Katzir and Singh 2013) defines the means for a language to lexicalize operators. First, it defines the two basic operators, the simplex ones, which are the ones corresponding to the A and I corners. Then, it defines a negation-based lexicalization mechanism that can lexicalize a negated operator. The lexicalization created in this manner is "marked". The definition in K&S is generalized for all categories of operators (connectives like *and* and *or*, quantifiers like *every/all* and *some*, and others), but here in (2) the mechanism is defined only for quantifiers for the sake of simplicity.

(2) Lexicalization Mechanism (based on Constraint on Lexicalization, 39 in Katzir and Singh 2013)

a. Basic case: The only simplex operators are the ones corresponding to the A corner (*every/all*) and the I corner (*some*) of the *Square of Opposition*.

b. Marked case: For an operator μ defined as in (2a), it may be possible to lexicalize $\neg\mu$, and the result is marked.

The Gricean Condition (see (3) below) is the explicit mechanism responsible for filtering out inventories containing "redundancies", as Horn put it. As "redundancies" are not defined explicitly by Horn, Katzir and Singh attempt to do it by defining the Coverage (see (4) below) of an inventory as the set of all the operators the inventory either contains or are implicatures of operators (see (5) below) in the inventory. Then, an inventory sharing the same Coverage with a proper subset of it is said to contain "redundancies" and should be blocked. This means that if we had an inventory of all the logical operators in the square, and the operator *not-all* in it is the implicature of *some* that is also in the inventory (thus covered by it), we would say the inventory contains "redundancies" and it would be filtered out.

(3) Gricean Condition (based on 41 in K&S)
Let X and Y be two inventories of logical operators such that $[X] = [Y]$. If $Y \subset X$, X cannot be lexicalized.

(4) Coverage (based on 4 in K&S)
Let X be an inventory of logical elements. For any logical element z, we say that X covers z if (a) $z \in X$, or (b) there is some $y \in X$ such that $y \rightsquigarrow_X z$ (that is, X covers z if z is either a member of X or the scalar implicature of some member of X). We write $[X]$ for the set of all elements that are covered by X.

(5) Operator-level Scalar Implicature (based on 40 in K&S)
Let Y be a set of operators. For any two operators y and z we say that z is an operator-level scalar implicature (OSI) of y given Y, written $y \rightsquigarrow_Y z$, iff the following conditions hold:

a. $\neg z \in Y$

b. $\neg z$ is *innocently excludable* (see (6) below) given y and $Alt(y, Y)$, where $Alt(y, Y)$ is the set of elements in Y that are at most as marked as y

(6) Innocent Exclusion (adapted from Fox 2007)
An element x is *innocently excludable* given an element a and a set A if x is in every maximal subset of A that can be negated consistently with a, formally $x \in IE(a, A)$ where:

a. $IE(a, A) := \bigcap\{B \subseteq A : B \text{ is a maximal set in } A \text{ s.t. } \neg B \cup \{a\} \text{ is consistent}\}$

b. $\neg B := \{\neg b : b \in B\}$

An important difference between Horn's and K&S' approaches is that instead of filtering out "marked inventories" by comparing the amount of negatives in the inventories like in Horn (1972), in K&S the sensitivity to markedness is incorporated into the scalar implicature calculation (see (5)).

Informally speaking, Horn argued that the inventory $\{A, O, E\}$ has more negations and thus is more marked than $\{A, I, E\}$, and although they are both able to express all the corners of the Square, the former is blocked because it is more marked. In K&S the sensitivity to markedness comes into play already when calculating implicatures for Coverage. O contains negation and thus is more complex and marked, and both E and I are equally considered as alternatives for calculating the implicature of O. They contradict each other so neither of them is *innocently excludable*, so O does not implicate any of them. I is a simplex and thus only A is a possible alternative for it, leading to the implicature of O. Here, the inventory $\{A, O, E\}$ simply does not cover the whole Square as opposed to $\{A, I, E\}$. This way, $\{A, I, E, O\}$ is blocked because $\{A, I, E\}$ is a subset with the same coverage.[3] Generally, the unattested inventories should not be valid options in the first place.

Apparently, K&S seems to get the right results in a well motivated way. The Gricean Condition serves as an economy condition, eliminating "redundant" inventories and limiting the number of the operators in the inventory only to the required ones. The coverage notion seems to imply an importance for inventories to "cover" the entire Square of Opposition, thus it serves as a maximization condition.

Yet, a few issues arise from Katzir and Singh (2013). In the next sections I'll go through some of them and propose a different frequency-based approach, avoiding those issues.

2.1 Issues with the economy condition

The first issue has to do with the Gricean Condition serving as an economy condition. It does prove as a useful condition, seeming to lead to the right predictions, shrinking the inventory of logical operators and preventing inventories such as $\{A, I, E, O\}$.

The condition entirely rules out the possibility for an inventory consisting of the 4 operators. If indeed such a condition governs the language acquisition process, the predictions state that whenever such an inventory is presented to an infant it is always discarded for a subset of that inventory. In other words, it predicts that a word for the O corner can never be acquired.

However, in an experiment reported in Hunter and Lidz (2012), also cited in Katzir and Singh (2013), it is shown that infants can definitely acquire such a word. Hunter and Lidz (2012) studied infants' acquisition of conservative and non-conservatives logical operators, and during the experiment a novel conservative operator named 'gleeb', corresponding to *not-all*, was successfully acquired by the subjects.

The economy condition, then, needs to be derived from a finer mechanism that can allow lexicalization of the whole square, but somehow induces a penalty that would make any natural language avoid such a lexicalization. In the fre-

[3] The only alternative in $Alt(I, \{A, I, E\})$ that is at most marked as I is A, and $\neg A$ is *innocently excludable* given I, thus $[\{A, I, E, O\}] = [\{A, I, E\}]$.

quency based model proposed below I suggest a lexicalization mechanism that does allow an inventory with a lexicalized O, but makes it a very rare event.

2.2 Issues with the maximal coverage motivation

The second issue is fundamental to the language acquisition process assumed in K&S. The system works under the assumption that there is some "coverage motivation" to the acquisition process that aspires to expand the inventory as much as possible. It is not spelled out explicitly by K&S but, as detailed below, it seems critical to the theory.

Without such motivation, one could imagine a system in which an infant would just infer the most plausible operator represented by each lexicalized item, and as long as the inventory assembled meets the Gricean Condition (namely, there is no proper subset that has the same coverage of that inventory) it is a valid one. This would lead us to expect a smaller, degenerated inventory of only weak operators (I and O corners) to be attested, as most of the times when A (*all*) is used, I (*some*) is true, and most of the times when E (*no*) is used, O (*not-all*) is also true. Moreover, a single-operator inventory of $\{O\}$ for example, or even an empty inventory also satisfy the Gricean Condition.

On the other hand, assuming the ability of an infant to compare the coverage of inventories is a very non-trivial assumption. It seems hard to imagine how an infant's language acquisition system could evaluate the coverage of an inventory and compare it to other inventories without having this rule hard wired into the system, but having such a rule hard wired seems stipulative.[4]

3 Usefulness and frequency based theory

While K&S offers a single mechanism that captures most of the phenomena, it requires individuals to be able to perform some complex tasks such as calculating inventories' coverage, and does not provide proper motivations for such tasks. Although it is unclear when the Gricean Condition (3) in K&S is checked, it seems natural for it to be part of the acquisition process, but as I showed before, its requirements are very non-trivial for the acquisition process to have.

Other processes that may be responsible for the phenomena are the "produc-

[4] When the coverage motivation is phrased as in (4) it is also hard to justify it, because, as said before, O cannot really be expressed in a single word with the inventory of $\{A, I, E\}$ but rather the strengthened meaning of it ($I \wedge O$). What motivates Coverage to be defined like this remains a mystery.

tion" process (i.e. word choice and sentence generation) and the grammaticization process (including lexicalization of new words). The "production" process seems to have a tendency toward simpler structure (in the spirit of Grice's Maxim of Manner), and the grammaticization process seems to occur on frequently used phrases ('*grammars code best what speakers do most*' — Du Bois 1985, p. 363).

In this thesis I propose a lexicalization model based mainly on those two processes. I define a simple generative speaker model that generates phrases for meanings a speaker may want to convey. This model then enables us to discuss how the phrases and meanings may be chosen by analyzing the expected typologies resulted by different assumptions added throughout the thesis, and comparing them to the attested pattern.

If we assume that the model lexicalizes frequently used phrases, we first need some initially lexicalized building blocks to create those phrases. The initially lexicalized building blocks are stipulated to be the *basic operators* (7) below.

As the model is based on the tendency of speakers to choose simpler structures for the meanings they want to convey, we also need to add those meanings and their possible representing structures to the model. The meanings a speaker might want to convey are captured by logical relations and their *usefulness pattern*. The usefulness pattern represents the usefulness of each relation, i.e. the frequency in which a speaker would choose to convey each meaning. As we see below, the usefulness pattern is critical for the model to get the attested lexicalization pattern. Instead of directly stipulating the attested lexicalization typology as the usefulness pattern, I will propose a pattern that will be more "natural", but may seem unintuitive at first.

4 Proposal

I propose an alternative model to K&S that derives the attested inventory from the interaction between frequency effects, tendency toward simplicity, and the primitives used for encoding the operators. Those effects will be realized through the "production" and lexicalization processes.[5]

In order to derive inventories through usage, we first need some initial state. In this thesis, I stipulate an initial state of lexicalized A and I as a working premise. This stipulated initial state of the inventory allows a speaker with such a lexicon and a simple grammar allowing negation to express any relation

[5] I assume both processes are carried out by the speaker, but another possibility would be to have the lexicalization carried out by a learner acquiring the inventory. This won't have any effect on the behavior of the processes or the overall inventory derivation.

at any corner of the Square of Opposition. It is stated formally in (7), and discussed below.

(7) BASIC OPERATORS: A and I are the basic operators[6], and are already lexicalized by other means.[7]

4.1 Lexicalization of frequently used constructs

Let's see how lexicalized inventories are derived. For simplicity, I assume a syntax where any operator can be negated by simply inserting *not* beforehand (even though in English you clearly can't just insert *not* before *some*, for example) and without an inverse scope reading for the negation. The derivation then begins, and the speaker is able to use the lexicalized operators in a sentence with or without negation, and thus express any relation they want.

For example, let's say a speaker wanted to say at this stage that *there are no blue cats*. The sentence describes an E relation, but as only *all* and *some* are lexicalized, the speaker could choose one of two constructs:

(8) a. All cats are not blue
    b. Not some cats are blue

Let's add an assumption that if some construct is useful for the speaker, thus used frequently, it gets lexicalized. For example, if in addition to the sentences in (8) the speaker wanted to describe many other E relations, they would have to use more constructs similar to those in (8). If this happened *frequently enough* and the phrase *not-some* is used, we would want it to undergo lexicalization.

The *all...not* construct might also be used frequently, but similarly to K&S the proposed mechanism does not allow it to get lexicalized. Hoeksema (1999) suggests that one main process involved in lexicalization of operators is contraction of adjacent elements, and following that idea I assume that lexicalization can occur only for phrases containing negation immediately followed by an operator. Effectively, this means that frequently negated operators get lexicalized, and it is stated formally in (9).

(9) LEXICALIZATION MECHANISM: Negation of the basic operators can be lex-

[6] This is strengthened by K&S' generalization over logical operators of different orders. K&S uses *sup* function as the base for the I operators and *inf* function as the base for the A operators to support that idea.

[7] The proposed system assumes A and I to have already have lexicalizations. Cases where one of them is not lexicalized are out of scope for this thesis.

icalized. A negated operator undergoes lexicalization if it occurs *frequently enough*.[8]

This mechanism relies on the definition of *frequently enough*. While there may be many different definitions of *frequently enough* that could work just as well, for the sake of being explicit and concrete we will use the definition in (10).

(10) FREQUENTLY ENOUGH (a concrete definition for lexicalization):
An operator occurs frequently enough if it occurs more than a certain number of times (*lexicalization threshold*) in a window of a certain number of used operators (*lexicalization window size*).

The assumptions of basic operators (7) and lexicalization mechanism (9) effectively limit the set of potentially lexicalized operators to the four corners of the Square of Opposition (Figure 1).[9] The two basic operators are already-lexicalized by other means, so in order to derive the attested lexicalization pattern, we need a model in which *not-some* is used more frequently than *not-all* in such a way that the former is used *frequently enough* (above the *lexicalization threshold*) while the latter is not. This would mean that *not-some* gets lexicalized and *not-some* does not, attesting for the lexicalization pattern.

4.2 Relations conveyed by structures

The next thing we need, after we have a mechanism for lexicalizing frequently used operators, is to figure out what affects the occurrence frequency. The first thing the operator occurrence frequencies are affected by is what kind of relations are conveyed by a speaker. Although the lexicalized set of operators is limited to the four corners of the Square of Opposition, we can imagine different relations that a speaker may want to convey, such as any of the corners of the Square of Opposition mentioned above, or even a more "complex" relation such as *some but not all*, referred to as the Y corner of the Extended Square of Opposition (Horn 2011; Horn 1990). The speaker can then use structures that may contain operators or negations to convey the wanted relations directly or by implicature.

The occurrence frequency of a structure containing an operator (or a negated operator) is derived from the usefulness of relations (i.e how frequently a speaker wants to convey them), and the constructs available for saying it. One could stipulate the attested lexicalization typology as the usefulness pattern directly,

[8]This mechanism may be a part of a general phrase lexicalization mechanism, but for our purposes I assume one that deals deals only with operators.

[9]The Constraint on lexicalization, (17) in Katzir and Singh (2013), also limits the lexicalized operators to the same set.

but that would not gain any explanatory power. It would be better if we could assume a simple, natural usefulness pattern for the relations and derive the attested lexicalizations from the available structures and their expected occurrence frequency. In this section I stipulate a usefulness pattern which is quite different from the typology, and below I try to argue for its naturalness.

I assume that the main operators a speaker may want to say are the five relations $\{A, I, E, O, Y\}$[10]. Initially, let's assume uniform usefulness for all five relations.[11]

We can now analyze how each relation affects the occurrence frequency of each operator by counting the expected number occurrences of each construct in various potential scenarios. For example, in the case exemplified in (8), as we did not assume any reason to prefer one construct over the other, 50% of the times (8a) is used and 50% of the times (8b) is used. This means that if a speaker wanted to convey the E relation 100 times, *all* would be used 50 times and *not-some* (and thus also *some*) would be used 50 times. We can say that 100 uses of the E relation "contribute" 50 occurrences each to *all*, to *some*, and to *not-some*.

Let's look at what would happen if a speaker wanted to convey a Y relation between the set of cats and the set of black things – *some, but not all cats are black*. The speaker could either say that explicitly, or use implicatures to convey that meaning (I assume that implicatures are independent of the currently lexicalized operator inventory). Here is the list of constructs that would be available for the speaker to choose from (implicatures in parentheses):

(11)
a. Some cats are black (but not all are)
b. Not all cats are black (but some are)
c. Some cats are not black (but some are)
d. Not all cats are not black (but not all are)
e. Some cats are black but not all are
f. Not all cats are black but some are
g. Some cats are not black but some are
h. Not all cats are not black but not all are

Similarly to the case before, in (11) the speaker can choose one of the 8 constructs. In those 8 constructs each of the operators *some*, *all* and *not-all* appears exactly 6 times, so generally if a speaker wanted to convey the Y relation

[10] The proposed mechanism is dependent on the usefulness of each element of the set and not the set itself, so we could add more complex relations to this list without making much of a change to the theory as long as the added ones are considered much less useful.

[11] Other relations are assumed for now to be close to non-useful, and thus to have almost no influence on the occurrence frequencies.

100 times, each operator would be expected to appear $\frac{6}{8}$ of them, meaning 75 times. We can say that 100 uses of the Y relation "contribute" 75 occurrences to *some*, to *all*, and to *not-all*.

We could do similar analyses to the other relations. Let's take an overall look at the expected numbers, when each of the relations is used 100 times:[12]

(12)
- A "contributes" 50 occurrences to *all*, *some* and *not-some*.
- E "contributes" 50 occurrences to *all*, *some* and *not-some*.
- I "contributes" 50 occurrences to *all*, *some* and *not-all*.
- O "contributes" 50 occurrences to *all*, *some* and *not-all*.
- Y "contributes" 75 occurrences to *all*, *some* and *not-all*.

While it is plausible that speakers would prefer some relations over others, we don't have any reason to assume so yet (will be revisited in section 4.4), so first let's assume all relations are equally useful. If all five relations have the same usefulness then in 500 propositions that a speaker may want to say, each of the assumed relations is expected to be used 100 times. In such a case, if we count the expected occurrences of each operator we get that *not-some* occurs 100 times (i.e. 20% of the time), while *not-all* occurs 175 times (i.e. 35% of the time). This means that by default, *not-all* would occur more frequently than *not-some* and thus is expected to be lexicalized before *not-some*, still not accounting for the attested lexicalization pattern.

4.3 Tendency towards simpler structures

Let's add a tendency towards simpler structures (stated informally in (14) below) to our system. Intuitively this means that when a speaker wants to convey an I relation like in (13), they will prefer to use (13a) over (13b).

(13) a. Some cats are black
b. Not all cats are not black

(14) Complexity tendency (to be defined explicitly in (17)): For conveying a certain meaning, the simpler a structure is, the more frequently it will be used.

For explicitly defining the complexity tendency, we first need to be able to compare two structures. Let's use a definition based on Katzir (2007), stated in

[12] Note that when the negation of *some* is used (or *all*), both *some* and *not-some* (or *all* and *not-all*) occurrences are counted

(15). This definition basically states that if there is a way to convert a structure to another by only replacing and deleting parts of it then it is more complex than the other:

(15) STRUCTURAL COMPLEXITY, based on (19) in Katzir (2007):
Let ϕ, ψ be parse trees. ψ is at most as complex as ϕ (and possibly simpler) if we can transform ϕ to ψ using a finite series of deletions, contractions, and replacements of constituents in ϕ with constituents of the same category.[13]

Now let's see how this notion of complexity affects the usage frequency. We need to be able to measure the penalty induced by a difference in complexity. *Structural complexity* is based on a series operations transforming one structure to another, so let's define the *structural distance* (16) between sentences as the minimal length of such series, not counting replacements[14]:

(16) STRUCTURAL DISTANCE:
Let ϕ, ψ be parse trees, such that ψ is simpler than ϕ. The distance between them is the length of the minimal series of deletions, contractions, and replacements of constituents in ϕ with constituents of the same category that would result in ψ, not counting replacements.

The definition allows us to measure the distance between (13a) and (13b), and say that (13a) is simpler than (13b) by 2. With this measured distance we can now define that a sentence which is simpler by 1 than another sentence occurs more times by a certain coefficient, say 2. This means that as (13a) is simpler than (13b) by 2, it occurs 2^2 times more. Now we can define the *complexity tendency* explicitly in (17).

(17) COMPLEXITY TENDENCY: For any ϕ and ψ structures available for a speaker to convey a certain meaning, such that ψ is simpler than ϕ by n, ψ is 2^n times more likely to be chosen than ϕ.

A consequence of the *complexity tendency* is that now, if a speaker wanted to convey a Y relation, there would be a tendency to use *some* over *all*, as the relevant structures involving *some* in (11) are generally simpler than the ones that do not involve it.

[13] While K&S assumes an internal structure for an operator and derives an operator's complexity, I don't assume such a thing and the complexity defined here does not care about an internal structure of an operator.

[14] If we counted replacements, two similar sentences could be regarded as simpler than each other by the same amount. *every cat is black* and *some cat is white* exemplify this unwanted scenario.

Let's analyze again the way each relation influences the operator occurrences, and assume that a speaker wanted to convey the relation I 100 times. This time, as we saw before, a sentence containing *some* (like the one in (13a)) is used 4 times the number of times a sentence containing *not-all* with another negation (like (13b) is used. So out of the 100 sentences, 80 would contain *some* while 20 sentences would contain *not-all*.

Overall, those are the expected numbers we get when each of the relations is used 100 times:

(18)
- A yields 80 occurrences of *all*, and 20 of *some* and of *not-some*.
- E yields 50 occurrences of *all*, *some* and *not-some*.
- I yields 80 occurrences of *some*, and 20 of *all* and of *not-all*.
- O yields 50 occurrences of *some*, *all* and *not-all*.
- Y yields ~ 75 occurrences of *some* and 40 of *all* and *not-all*.

Now, both A and E still contribute to the frequency of *not-some*, and O and I still contribute to the frequency of *not-all*, but now Y contributes much less than before to *not-all*.

Still, if we assume a uniform usefulness for the relations, the expected occurrences number of *not-some* is 70 while the expected occurrence number of *not-all* is 110. This means that *not-all* still occurs more frequently than *not-some*, thus *not-all* is expected to be lexicalized before *not-some* and the attested lexicalization pattern is still not accounted for.

4.4 Usefulness pattern

As mentioned before, in order to account for the attested lexicalization pattern, the frequency based model needs to show a clear difference between the frequency of *not-some* and the frequency of *not-all*. This would allow them to be lexicalized differently if the *lexicalization threshold* was below the frequency of *not-some* but above the frequency of *not-all*. The relations A and E contribute to the frequency of *not-some* while the relations I and O contribute to the frequency of *not-all* in the same manner. Thus, it is crucial that the usefulness of A and E would be greater than the usefulness of I and O.

In many cases, when using *some* or *not-all* in discourse, Y is actually what is meant (by implicatures) rather than I or O (see discussion in section 6.2), so it might be reasonable to assume that Y is more useful than I and O. Adding this assumption about Y into the picture actually increases the frequency of *not-all*, but the tendency toward simpler constructs reduces the frequency of *not-all* (as *some* tends to be used instead) back again.[15]

If we do assume that in a certain system the operators A, Y, and E are more useful for speakers (i.e. there are more contexts in which a speaker would want to say them) than the other relations (i.e. I, O), say 4 times more, the proposed mechanism predicts the attested inventories.[16]

Let's analyze what happens if A, Y, and E are used 400 times each, but I and O are used just 100 times each:

(19)
- A yields 320 occurrences of *all*, and 80 of *some* and of *not-some*.
- E yields 200 occurrences of *all*, *some* and *not-some*.
- I yields 80 occurrences of *some*, and 20 of *all* and of *not-all*.
- O yields 50 occurrences of *some*, *all* and *not-all*.
- Y yields $\sim$ 300 occurrences of *some* and 160 of *all* and *not-all*.

In this case, *not-some* is expected to occur 280 times while *not-all* is expected to occur just 230 times. This predicts the attested lexicalization pattern by the derivation process described below.

4.5 Deriving the attested lexicalization pattern

Let's assume the A, Y, and E relations are 4 times more useful than I, and O, and that the *complexity coefficient* is 2 as assumed above, and see how the derivation lexicalizes *not-some* before *not-all*:

First, A and I are lexicalized by default. This is the base state of lexicalization. In that base state, when a speaker wants to say A, they will usually use *all*. They could also use *not-some* with the negation of the predicate, but those are much more complex and thus less probable, specifically 2^2 times less probable as we saw above.

When a speaker wants to say Y, they will use *some* for it most of the times, as it can mean Y by implicature. They could also use *not-all* for it, or any other construct in (11), but again this is more complex and thus less probable (see

[15] Actually, if we wanted to assume a much higher ratio between A,E,Y and I,O (at least 5 to 1), we could avoid having the model rely on complexity and still get the frequency of *not-some* to be higher than of *not-all*.

[16] The usefulness pattern we assume here predicts the lexicalization of E but the lack of lexicalization of O, like the attested inventory. It was claimed before that *needn't* could be a lexicalization of O (Auwera 2014; Horn 1989, p. 260). If that is the case, a different usefulness pattern in those contexts (one in which O is more useful) could give us a simple explanation. The existence of words like *unnecessary* refer to a similar meaning and may support this idea.

(14)). In those constructs *some* would be used almost twice as much as *not-all*. For E, a speaker will have to use one of the structures *not-some* and *all* with a negated predicate, which are equally complex. A speaker might also want to say I or O specifically. I behaves similarly to A in respect to the frequency pattern, and O behaves similarly to E.

If we count all the expected occurrences we will see that while *not-some* is used 20% of the time, *not-all* is used only about 16% of the time. Thus, the base state would usually lead to a more frequent use of *not-some* than *not-all*. As the lexicalization mechanism is based on the number of occurrences in a window (10), if we set the *lexicalization threshold* to be about 25% of the *window size*, there will be a chance for the system to move to a state where *all*, *some* and *not-some* are lexicalized.

If the derivation continues in that state of the inventory, there won't be any change in the likelihood of the use of *not-all*, thus it still wouldn't be probable for it to get lexicalized.[17]

This process extends over generations of speakers, and when creating the computational model it is important to make sure the state of the process can be transfered to a next generation. That's why a learner will be part of the model, even though acquisition or speaker-learner dynamics do not play part in accounting for the phenomena.

5 Computational model and simulation

The computational model presented is a very simple iterative model, in which on each iteration there are three components at play: a speaker, a learner and a world described by the speaker.

In such models, the learner is usually a part of speaker-learner dynamics that play an important role in the model. This is not the case in the current model, where the phenomena could be accounted for by a single speaker speaking for a long time. The learner exists in the model only to make sure that the model does not rely on any internal state of the speaker that cannot be transfered through learning.

The world is actually just a collection of individuals and named groups of these individuals, representing predicates. On each iteration (a.k.a generation), the speaker describes relations between two groups for many different groups

[17]The proposed mechanism does not have a process for removing lexicalizations, thus an event of lexicalizing *not-all* will lead to *not-all* being lexicalized for all generations. Removing lexicalizations actually might be possible, and it might affect the dynamics assumed here.

in many sentences. The generated sentences are of the form shown in Figure 2 where Q is the chosen quantifier, R is the quantification range and P is the predicate. For example, the relation of *every cat is black* is formed as "*every cat black*", where $Q = every, R = cat, P = black$. Q and P may be negated if prefixed with *not*. In some complex cases, a sentence consisting of a couple of sub-sentences of that form may be generated.

$$\left\{ \begin{array}{c} not \\ \emptyset \end{array} \right\} \quad Q \quad R \quad \left\{ \begin{array}{c} not \\ \emptyset \end{array} \right\} \quad P$$

Figure 2: The form of generated sentences

The sentences tend to be simpler (17), ranked as follows (in line with (15)):

$$Q\ \ R\ \ P \quad << \quad \begin{array}{c} not\ \ Q\ \ R\ \ P \\ Q\ \ R\ \ not\ \ P \end{array} \quad << \quad not\ \ Q\ \ R\ \ not\ \ P$$

For the sake of simplicity, lexicalization will be incorporated into the speaker, where when a phrase of the form *notQ* occurs *frequently enough* (10), it gets lexicalized.

The learner then induces the meaning of the quantifiers involved in the descriptions, but does nothing more.

5.1 The process

Initially, a world is generated randomly and a speaker is initialized with a lexicalized inventory of operators consisting of the two basic operators (i.e. *all* and *some*). As said before, this reflects the idea that this is the base state for the lexicalization process, on which the proposed mechanism works.

During each generation, the speaker generates numerous true sentences. For each sentence it makes a weighted choice of an operator out of the set $\{A, Y, E, I, O\}$. For example, suppose $\{A, Y, E\}$ are 4 times more probable then $\{I, O\}$ so the probability for each of the three useful operators is 0.3125 and the probability of each of the not-useful operators is 0.03125.

After that, the speaker finds two groups having the chosen relation between them, and then, it finds the possible structures to use for expressing the chosen operator (depending on what is lexicalized).

From the generated structures it chooses one, giving better probabilities to simpler structures. In our example, the complexity of the structure is in direct relation to the sentence length. The probability is calculated such that a sentence of length n has more weight than that of a sentence of length $n+1$ in the ratio defined by a *complexity coefficient*.

Finally, a sentence is generated. Each time a sentence is generated, a lexicalization process may happen. The speaker looks back at the recently used operators (up to certain number of operators, determined by the *window size*), and if a negated operator was used above a certain number of times (*lexicalization threshold*), it undergoes lexicalization.

The learner actually does not play any real part in the model except for showing us that the process can continue across generations, and that the state of the model can be easily learned. It takes the generated sentence and infers a valid grammar, i.e. a grammar for which the sentences are true, in respect to the generated world. The search space is very small and usually there is only one possible grammar, so the learner does not employ any interesting learning algorithm.

5.2 Simulation

A simulation of the computational model presented above was written in Python[18]. I ran the simulation a few times for 200 generations, each generating 1000 sentences, with the following parameters:

- The usefulness is such that A, Y, E weigh 4 times more than I, O.
- The *complexity coefficient* is 2.
- For the lexicalization mechanism (9), the *lexicalization window size* is 120 occurrences and the *lexicalization threshold* is 30 occurrences.

This simulation indeed resulted in the attested lexicalization pattern. In some runs the resulting inventory contained only lexicalizations for the initial state of A (*all*) and I (*some*), and in some it contained also a lexicalization of E (*not-some*), as expected.

6 Discussion

The model I propose here takes a different approach to the typological puzzle raised by Horn (1972) than the ones attempted before. By separating the use of

[18]The code can be found at: `https://bitbucket.org/taucompling/oplexi/`

operators into a level of meaning (the relations a speaker wants to convey) and a level of structure (the operators used in a structure conveying a relation) the puzzle became a question about the basic operators starting the whole process, and about the set of available meanings and their usefulness.

While the lexical mechanism (9) and complexity tendency (14) seem well motivated by universally attested principles, the basic operators and the usefulness pattern assumed here are merely stipulations and are in need of more justification.

6.1 The basic operators

The first stipulation the proposed model relies on is that the A and I operators, namely *all* and *some*, are "basic" and are lexicalized by a different process. This is critical, because the lexicalization mechanism I assume takes already-lexicalized items as building blocks, and if the system did not have the two building blocks, the other two could never be lexicalized.

The idea of these operators being cognitively basic was suggested before in the literature. For example, in the recent years, Buccola et al. (2016) argued *all* and *some* are conceptually simpler, and Katzir and Singh (2013) made generalizations over logical operators of different orders and suggested the simple and cognitively motivated *sup* and *inf* functions as the base for *some* and *all* (and other I and A operators).

6.2 The usefulness pattern

The second stipulation is the specific usefulness pattern leading to the expected lexicalizations, which relies on I and O being much less useful than Y and assumes that the basic relations a speaker might want to convey are the five corners of the Extended Square of Opposition, A, E, Y, I, O.

There is much reason to assume that theoretically, in the spirit of Grice's Maxim of Quantity, whenever Y is applicable, speakers will prefer to convey Y over I or O. Also, the use of *some* and *not-all* usually has the meaning of Y. This led different logicians, philosophers and semanticists to model the logic of language with Y as an independent status, sometimes even on the expense of I and O (see overview in Horn 2011).[19]

[19] The frequent use of *some* for conveying Y rather than I even led Ariel (2015) to argue that *some* lexically has the meaning of Y and that the I reading it sometimes has is an explicature. Even though the discussion of the lexical meaning of *some* or *not-all* is rather irrelevant to the model I propose, this gives further evidence for the usefulness of Y over I and O.

The five basic relations can actually be derived from A and I being the basic operators. If we look at the different combinations we can have with either of the two basic operators, each may or may not be participating in the meaning to be conveyed and with or without negation, we get exactly those five relations:

		$\boldsymbol{A}$	$\neg\boldsymbol{A}$
		A	O
$\boldsymbol{I}$	I	A[20]	Y
$\neg\boldsymbol{I}$	E	?[21]	E

Yet, the usefulness pattern still needs to be empirically attested, either by a computational model or by a typological study of the meanings conveyed by sentences.

References

Ariel, Mira (2015). "Doubling up: Two upper bounds for scalars". In: *Linguistics* 53.3, pp. 561–610.

Auwera, Johan Van der (2014). "More ado about *nall". In: *Papers dedicated to Jacques Moeschler*. Ed. by J. Blochowiak. Université de Genève, 1–10.

Buccola, Brian, Manuel Križ, and Emmanuel Chemla (2016). "Conceptual alternatives". In:

Du Bois, John W. (1985). "Competing Motivations". In: *Iconicity in Syntax*. Ed. by J. Haiman. John Benjamins Publishing Co., 343—365.

Fox, Danny (2007). "Free Choice and the Theory of Scalar Implicatures". In: *Presupposition and Implicature in Compositional Semantics*. Palgrave Macmillan UK, pp. 71–120. DOI: 10.1057/9780230210752_4. URL: https://doi.org/10.1057%2F9780230210752_4.

Hoeksema, Jack (1999). "Blocking effects and polarity sensitivity". In: *JFAK: Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*.

Horn, Laurence (1972). "On the semantic properties of logical operators in English". PhD thesis. University of California, Los Angeles.

— (1989). "A natural history of negation". In: *IL: University of Chicago Press, Chicago*.

— (2011). "Histoire d'* O: Lexical pragmatics and the geometry of opposition". In: *Bern: Peter Lang*, pp. 383–416.

[20] There is some discussion in literature about the existential import of the universal quantifier, i.e. A. In this table I assume that the basic universal operator does have an existential import.

[21] If we assume the existential import, A contradicts $\neg I$.

Horn, Laurence R. (1990). “Hamburgers and Truth: Why Gricean Explanation is Gricean”. In: *Annual Meeting of the Berkeley Linguistics Society* 16.1, pp. 454–471. DOI: 10.3765/bls.v16i0.1693. URL: https://doi.org/10.3765%2Fbls.v16i0.1693.

Hunter, T. and J. Lidz (2012). “Conservativity and Learnability of Determiners”. In: *Journal of Semantics* 30.3, 315–334. DOI: 10.1093/jos/ffs014.

Katzir, Roni (2007). “Structurally-defined alternatives”. In: *Linguistics and Philosophy* 30.6, pp. 669–690. DOI: 10.1007/s10988-008-9029-y. URL: https://doi.org/10.1007%2Fs10988-008-9029-y.

Katzir, Roni and Raj Singh (2013). “Constraints on the lexicalization of logical operators”. In: *Linguistics and Philosophy* 36.1, pp. 1–29. DOI: 10.1007/s10988-013-9130-8. URL: https://doi.org/10.1007%2Fs10988-013-9130-8.

תקציר

ב־Horn (1972) מוצג היעדרו של הכמת *not-all* מכלל השפות הנחקרות, ומוצע כי התופעה הזו נובעת מכך שלקסיקליזציה של הכמת הזה מיותרת מכיוון שהכמת *not-all* משתמע כאימפליקטורה מהכמת *some*. בנוסף נטען כי למרות שהמיותרות הזו היתה יכולה להפתר גם על ידי לקסיקליזציה של הכמת *not-all* והיעדר הכמת *some*, שפות "בוחרות" בלקסיקון שמכיל פחות שלילות.

ב־Katzir (2013) ה"תעלומה" של Horn נבדקת שוב, מוצגת הכללה שתופסת אופרטורים לוגיים הן מסדר נמוך והן מסדר גבוה, ומוצעת מערכת מפורשת יותר שחוזה את היעדר הכמת *not-all*. למרות זאת, חסרות במערכת שהוצעה מוטיבציות קוגניטיביות וכן המערכת מייצרת כמה תחזיות שגויות.

בעבודה זו, אני מציע גישה אחרת, מבוססת שכיחויות, המתבססת על האינטרקציה בין האפקטים של שכיחות שימוש, הנטיה לפשטות, ואבני הבניין המשמשות לקידוד האופרטורים הללו. אני מציג מודל חישובי הנראה בעל מוטיבציה קוגניטיבית ומסביר את תבנית הלקסיקליזציה שהוצגה על ידי Horn.

אוניברסיטת תל אביב
הפקולטה למדעי הרוח ע"ש לסטר וסאלי אנטין
החוג לבלשנות

מודל מבוסס שכיחויות עבור לקסיקליזציה של אופרטורים לוגיים

חיבור זה הוגש כעבודת גמר לקראת התואר
"מוסמך אוניברסיטה" באוניברסיטת ת"א

על ידי:
אדם רימון

בהנחיית:
ד"ר רוני קציר

נובמבר 2018