Semantic triviality leads to ungrammaticality through iterated learning[1]

Ciyang QING — *University of Edinburgh*
Wataru UEGAKI — *University of Edinburgh*

Abstract. A major line of research in semantics concerns meaning-driven explanations of combinatorial restrictions of various operators. Such explanations rely on a link from semantic triviality to ungrammaticality, standardly explicated in terms of LOGICALITY. According to the logicality approach, the grammar contains a natural deductive system that contributes to speakers' grammaticality judgments, besides the purely syntactic combinatorial system. However, this involves a non-trivial architectural assumption about the interaction between syntax and semantics, and there is no consensus on the exact specification of the natural deductive system. In this paper, we provide an alternative account of the link between semantic triviality and ungrammaticality based on ITERATED LEARNING, an independently motivated model of language evolution. Within this model, it can be shown that for certain trivial sentences, a population of speakers possessing a grammar that in principle generates them is overtaken by learners who induce a grammar that rules them out after several generations. Crucially, our account does not need to postulate an additional natural deductive system.

Keywords: triviality, logicality, ungrammaticality, iterated learning.

1. Introduction

A major line of research in the semantics of natural language concerns MEANING-DRIVEN EXPLANATIONS of combinatorial restrictions of various operators in the natural language. According to a meaning-driven explanation, a combinatorial restriction arises if a composition of operators is predicted to give rise to semantic triviality. This line of explanation can be given for the definiteness effect of the English existential-*there* construction (Milsark, 1977; Barwise and Cooper, 1981), restrictions on connected exceptives (von Fintel, 1993), weak islands (Abrusán, 2014; Schwarz and Simonenko, 2018), licensing of polarity-sensitive items (Chierchia, 2006, 2013) and selectional restrictions of clause-embedding predicates (Theiler et al., 2019) among others.

Crucially, such explanations assume that there is a link from semantic triviality to ungrammaticality. According to a prominent hypothesis, this assumption is explicated in terms of the LOGICALITY of grammar. That is, the grammar contains a natural deductive system that contributes to speakers' grammaticality judgments, besides the purely syntactic combinatorial system (e.g., Gajewski, 2002; Fox and Hackl, 2006; Chierchia, 2013). This deductive system, or the 'natural logic', automatically computes logical properties of structures built by syntax and filters out those that are logically trivial. Thus, according to this hypothesis, syntax together with the deductive system determines the set of grammatical sentences. Following Del Pinal (2019), we call this hypothesis the logicality hypothesis.

[1]We would like to thank Chris Barker, Kajsa Djärv, Patrick Elliott, Michael Glanzberg, Simon Kirby, Dan Lassiter, Anne Mucha, Deniz Özyıldız, Floris Roelofsen, and Maribel Romero, as well as audiences at The New York Philosophy of Language Workshop, Sinn und Bedeutung 29, and Amsterdam Colloquium 2024. All remaining errors are our own. This work is supported by the UKRI Future Leaders Fellowship (MR/V023438/1).

In this paper, we will provide an alternative account of the link between semantic triviality and ungrammaticality. Our alternative replaces the natural deductive system with ITERATED LEARNING, an independently motivated model of language evolution (Kirby and Hurford, 2002). Within this model, it can be shown that for certain trivial sentences, a population of speakers possessing a grammar that in principle generates them is overtaken by learners who induce a grammar that rules them out after several generations. Crucially, our account does not need to postulate an additional natural deductive system.

We would like to stress that our aim in this paper is not to provide an argument against the logicality hypothesis. Rather, our aim is to put our hypothesis on the table as an alternative and compare its predictions and assumptions with those of logicality-based theories. It is also important to note that logicality and our model can co-exist. In particular, it is possible that certain cases of meaning-driven ungrammaticality are accounted for by logicality while other cases are accounted for by our alternative. We leave the discussion of such phenomenon-specific considerations to another paper (Qing and Uegaki, prep).

This paper is structured as follows. In §2, we present examples of meaning-driven ungrammaticality. In particular, we will present how the definiteness effect of the English existential-*there* construction is accounted for in terms of semantic triviality, following Barwise and Cooper (1981). §3 reviews the logicality hypothesis and discusses how it deals with the issue of distinguishing between grammatical and ungrammatical triviality, following Gajewski (2002) and Del Pinal (2019). We present our alternative in terms of iterated learning in §4. After presenting the architectural assumptions and the overview of the model, we will provide a toy model that accounts for the evolution of the definiteness effect in the existential-*there* construction. In §5 , we present how our account teases apart grammatical triviality from ungrammatical triviality. §6 compares assumptions and predictions of the logicality hypothesis and our alternative. In §7, we discuss further nuances regarding the interpretation of our alternative as well as its potential phenomenon-specificity. We conclude in §8.

2. Meaning-driven ungrammaticality

In this section, we will survey phenomena that have been argued to involve meaning-driven ungrammaticality in the literature. Following the existing literature on logicality (e.g. Gajewski, 2002; Del Pinal, 2019), we will use Barwise and Cooper's (1981) account of the definiteness effect in the English existential-*there* construction (ETC) as our leading example. As such, the toy model we will develop in §4 assumes Barwise and Cooper's (1981) derivation of semantic triviality for the ETC and explains how such semantic triviality leads to ungrammaticality. However, the focus on this particular phenomenon is only for practical purposes and it is not crucial for our general claim that Barwise and Cooper's (1981) is ultimately correct. What is crucial is that meaning-driven ungrammaticality exists (whether or not the definiteness effect is its instance) and that it can be stated in terms of distributions of closed-class items. The relevance of the latter condition in terms of closed-class items will be discussed in detail in §5.

As exemplified below, the English existential-*there* construction is compatible with some determiners, but not others.

(1) a. There is { a / no / *every } smiling cat.

b. There are { a few / *most } smiling cats.

As observed by Milsark (1977), the distinction between the two classes of determiners correlates with the semantic distinction between STRONG and WEAK DETERMINERS. We take Barwise and Cooper's (1981) definition of positive and weak determiners, as follows:

(2) a. Determiner *det* is POSITIVE STRONG iff for every model $M = \langle [\![\ ]\!], D_e \rangle$ and every $A \subseteq D_e$, if $[\![det]\!](A)$ is defined, then $[\![det]\!](A)(A) = 1$
b. *det* is NEGATIVE STRONG iff for every model $M = \langle [\![\ ]\!], D_e \rangle$ and every $A \subseteq D_e$, if $[\![det]\!](A)$ is defined, then $[\![det]\!](A)(A) = 0$
c. *det* is WEAK if *det* is not strong.

For instance, *every* and *most* are positive strong while *a*, *no*, *a few* are weak.

Given these definitions of strong and weak determiners, it can be shown that strong determiners in an ETC give rise to semantically trivial truth conditions. First, an ETC with a determiner *det* is analyzed as involving generalized quantification as follows:

(3) $[\![\textit{there be det NP}]\!] \Leftrightarrow [\![det]\!]([\![NP]\!])(D)$ (D: the domain of individuals)

We assume that $[\![NP]\!] \subseteq D_e$. Given conservativity, $[\![det]\!]([\![NP]\!])(D) \Leftrightarrow [\![det]\!]([\![NP]\!])([\![NP]\!] \cap D)$ $\Leftrightarrow [\![det]\!]([\![NP]\!])([\![NP]\!])$. If *det* is positive strong, by (2a), this is trivially true. If *det* is negative strong, by (2b), this is trivially false. Thus, if *det* is strong, $[\![\textit{there be det NP}]\!]$ is tautological or contradictory, as long as $[\![NP]\!] \subseteq D$. On the other hand, a weak determiner like *no* does not give rise to semantic triviality. With conservativity, $[\![no]\!]([\![NP]\!])(D) \Leftrightarrow [\![no]\!]([\![NP]\!])([\![NP]\!] \cap D)$ $\Leftrightarrow [\![no]\!]([\![NP]\!])([\![NP]\!])$. This truth condition is contingent, as it depends on whether $[\![NP]\!]$ has a member.

Assuming that semantically trivial sentences are filtered out from the set of grammatical sentences in some way, this explains the ungrammaticality of strong determiners such as *every* and *most* in the ETC, while the account rules in weak determiners in the same environment. Other phenomena hypothesized to involve meaning-driven ungrammaticality include the following.

Connected exceptives As exemplified in (4), connected exceptive phrases of the form *Det but NP* only allow universal determiners such as *all* and *no*. von Fintel (1993) shows that this pattern can be explained by semantics, as non-universal determiners give rise to triviality in this environment.

(4) a. **All** students but Sue passed the exam.
b. ***Few** students but Sue passed the exam.

Licensing of polarity items Negative polarity items such as *any* is licensed under the scope of negation among other downward-entailing environments, as in (5a), while it is anti-licensed in an episodic positive environment, as in (5b).

(5) a. Sue does**n't** have any eggs.
b. *Sue has any eggs.

Chierchia (2006, 2013) provides a meaning-driven explanation of this pattern based on an obligatory exhaustification operator which, roughly, negates all subdomain alternatives for the polarity item (cf. Krifka, 1994; Lahiri, 1998).

Negative islands in comparatives *Than*-clauses in English comparatives permit quantification by certain determiners, but not by others, as can be seen in (6).

(6) a. Mary is taller than **every** other student is.
    b. *Mary is taller than **no** other student is.

Gajewski (2008) offers the generalization that determiners which are (right) downward entailing are ruled out in this environment while the others are ruled in. He provides a meaning-driven explanation of the distribution.

Weak islands in manner questions Manner questions that involve a *wh*-movement from within the complement of a factive predicate yields ungrammaticality while those that involve a movement from within the complement of a non-factive predicate are grammatical. Abrusán (2014) shows that this pattern can be explained in terms of triviality of the presupposition predicated for the factive case.

(7) a. How does John **hope** that Peter fixed the car?
    b. *How does John **regret** that Peter fixed the car?

See also Oshima (2007) and Schwarz and Simonenko (2018) for extensions of the generalization and modified views on the nature of the semantic triviality involved in the ungrammatical cases.

Selectional restrictions of clause-embedding predicates Certain clause-embedding predicates exhibit selectional restrictions with respect to the types of clausal complements they combine with (Grimshaw, 1979). For example, whereas *know* can combine with both declarative and interrogative complements, *think* can only combine with declarative complements. This pattern is exemplified below:

(8) a. Xander **knows** { that Silver / which horse } won the race.
    b. Xander **thinks** { that Silver / *which horse } won the race.

Zuber (1982) offers the generalization that this pattern is partly correlated with the NEG-RAISING PROPERTY of predicates—neg-raising predicates are incompatible with interrogative complements. Theiler et al. (2019) and Mayr (2019) provide a meaning-driven explanation for the generalization, in terms of semantic triviality predicted for the combination of neg-raising predicates and interrogative complements. See also Uegaki (2023) and Uegaki and Sudo (2019) for meaning-driven accounts of the selectional restrictions of other classes of predicates.

3. The logicality hypothesis

3.1. Overall architecture

The accounts of meaning-driven ungrammaticality surveyed in the previous section all assume that there is a link between semantic triviality and ungrammaticality. This assumption is a nontrivial one, given separation between syntax and semantics, as standardly assumed. In this section, we will review the logicality hypothesis, which is designed to explicate the link in terms of an additional logical deductive component in grammar.

The architectural assumptions of the logicality hypothesis are as follows. Grammar contains syntax as well as NATURAL LOGIC, i.e., a logical deductive system that automatically computes whether structures built by syntax are logically trivial or contingent. Natural logic filters out logically trivial structures from the set of grammatical sentences. By positing this automatic filtering mechanism, the hypothesis accounts for the fact that certain semantic triviality leads to ungrammaticality.

3.2. Grammatical vs. ungrammatical triviality

An issue with any account of the triviality-ungrammaticality link concerns the distinction between grammatical and ungrammatical triviality. Not all semantically trivial sentences are ungrammatical. The following sentences, for example, can be taken to be semantically trivial yet grammatical.

(9) a. This atom is bivalent and is not bivalent. (contradictory)
b. This atom is bivalent and is trivalent. (contradictory)
c. It is raining or it is not raining. (tautological)

Under the logicality hypothesis, natural logic should *rule out* meaning-driven ungrammaticality of the kind mentioned in §2 but *rule in* the grammatical triviality of the kind in (9). How can natural logic be specified to make this distinction? An analytical intuition pursued in the literature to address this issue is this: ungrammatical triviality involves triviality that arises purely from the **logical** words in the structure, whereas grammatical triviality as in (9), arises due to the meaning of **content** words. Below, we will review two prominent solutions based on this analytical intuition—one based on LOGICAL SKELETONS (Gajewski, 2002, 2009) and the other based on the RESCALE operation (Del Pinal, 2019).

A remark is in order at this point regarding the distinction between logical and content vocabularies. For the purpose of this paper, we will not delve into the nature of this distinction.[2] Rather, we—along with proponents of logicality such as Gajewski and Del Pinal—will assume that such a distinction exists and that it corresponds to the functional/lexical distinction and/or the closed/open-class distinction traditionally assumed in syntax. However, in §5 below, we will speculate on the possibility of departing from the dichotomous view and adopting a more gradient distinction between the two classes.

Logical skeletons Gajewski (2002, 2009) proposes that natural logic operates on the logical skeleton of a sentence, which is a syntactic representation where each token of a content word is replaced with a new variable. Natural logic determines a sentence to be logically trivial iff its logical skeleton is necessarily true or necessarily false under all variable assignments. To see how this analysis makes the correct distinction between grammatical and ungrammatical triviality, consider the following pair of examples.

(10) a. *There is every cat.

[2]See van Benthem (1989); van Benthem (2002) for a possible definition of logical vocabulary in terms of permutation invariance, and Bonnay (2008) for its refinement.

b. Hydrogen is bivalent and is not bivalent.

In (10a), we have an ungrammatical example with the violation of the definiteness constraint in the ETC. In (10b), we have grammatical triviality. The logical skeletons of (10) can be represented as follows, with x/X_i representing distinct variables replacing the content words:

(11) a. there is every X_1.
b. x_1 is X_2 and is not X_3.

Here, (11a) is necessarily true, regardless of the value we assign to the variable X_1, following the derivation of triviality presented in §2. On the other hand, (11b) is contingent, crucially because X_2 and X_3 are distinct variables. Note that each token of a content word in the original sentence—even if they are the same lexeme—is replaced with a new variable. Thus, Gajewski (2002, 2009) allows the account to make an appropriate distinction between grammatical and ungrammatical triviality by restricting natural logic to operate on logical skeletons.

Rescaling Del Pinal (2019) problematizes the non-standard nature of natural logic under the skeleton view. In particular, according to the skeleton view, natural logic does not validate the classical inference rules such as Law of Non-Contradiction and Modus Ponens. This motivates Del Pinal to adopt an alternative view within the logicality hypothesis in terms of a meaning modulation mechanism dubbed RESCALING. According to this view, natural logic is equipped with the RESCALE operator, which narrows down the meaning of an open class term. The operator is defined as follows:

(12) For any open class term P, argument of suitable type x and context c,
$\{x|\textsc{Rescale}_c(P)(x)\} \subseteq \{x|P(x)\}$. (Del Pinal, 2019: 11)

That is, the meaning modulation based on RESCALE is constrained to narrow down meanings, where the precise refinement depends on the context parameter c. Crucially, Del Pinal assumes the dichotomy between closed-class and open-class vocabularies (in a way parallel to Gajewski's reliance on the logical vs. content vocabulary distinction) and stipulates RESCALE to be only applicable to the open-class terms.

With the RESCALE operator in place, one can distinguish grammatical triviality and ungrammatical triviality—grammatical cases are those that allow a non-trivial interpretation when RESCALE is inserted while ungrammatical cases are those that remain trivial even after insertion of RESCALE. For example, the pair in (10) can be distinguished based on the following logical forms:

(13) a. There is every $\textsc{Rescale}_c$(cat).
b. Hydrogen is bivalent and is not $\textsc{Rescale}_c$(bivalent).

Even with rescaling, (13a) remains to be trivial, given Barwise and Cooper's (1981) derivation we sketched in §2. On the other hand, (13b) is non-trivial, as the meaning of *bivalent* can be modulated to something narrower, making the statement consistent and contingent. Thus, Del Pinal (2019) captures the distinction between grammatical and ungrammatical triviality by postulating the additional rescale operator within natural logic, without making the logic itself non-standard as in the skeleton view.

3.3. Upshot

In this section, we provided an overview of the logicality hypothesis as a theory designed to explain the link between semantic triviality and ungrammaticality. The hypothesis assumes that grammar contains a natural deductive system, natural logic, which filters out logically trivial structures. We have discussed two specific views within the logicality hypothesis on how to tease apart grammatical triviality and ungrammatical triviality: the skeleton view due to Gajewski (2002, 2009) and the rescaling view due to Del Pinal (2019).

Regardless of the specific stance one adopts within the logicality hypothesis, the hypothesis entails a substantial architectural assumption about the interaction between syntax and semantics—namely, that the grammar contains a logic with its own deductive system, which acts as a filter for grammatical structures. Furthermore, as we have seen, there is currently no consensus on the exact specification of this natural deductive system. See e.g., Abrusán (2019); Del Pinal (2019); Pistoia-Reda and Sauerland (2021) for further discussion of the latter issue.

As we have emphasized in the introduction, our aim is not to provide an argument against the logicality hypothesis per se. Nonetheless, we believe that the significant architectural assumptions it entails justify exploring an alternative that does not rely on such assumptions. In the following sections, we present our alternative, which seeks to replace natural logic with an independently motivated model of language evolution based on iterated learning.

4. An alternative: Meaning-driven ungrammaticality through iterated learning

In this section, we propose an alternative analysis of how semantic triviality can lead to ungrammaticality, without assuming that natural logic is a part of the grammar. Instead, we assume that the process of ITERATED LEARNING filters out (certain but not all) semantically trivial sentences in the diachronic process.[3] Specificially, our iterated learning model makes the following architectural assumptions (14).

(14)
a. **Separation of syntax and semantics**: The syntax module only concerns possible combinations of forms without reference to their meanings.[4]
b. **Constrained hypothesis space of grammars:** The language faculty defines the space of possible grammars that learners can in principle entertain. Furthermore, learners may have initial preferences for certain grammars than others.
c. **Iterated learning**: Learners acquire/induce a grammar based on the input produced by the speakers of the previous generation together with their initial preferences. Then they produce sentences according to their grammar, which in turn will serve as input for the learners in the next generation.
d. **Pragmatic speakers**: Each generation of speakers are pragmatic speakers, who produce utterances that maximize the chance of a listener choosing the intended meaning.

[3] Iterated learning is developed in the theory of language evolution and has been used to model a variety of phenomena, such as emergence of compositionality (Kirby and Hurford, 2002; Kirby et al., 2008, 2015), emergence of morphological regularity (Smith and Wonnacott, 2010), and word-order harmony (Culbertson and Kirby, 2016).
[4] Note that this assumption does not preclude speakers from noticing and making use of certain aspects of meanings of an expression that correlate with its formal (morpho-syntactic) features, e.g., gender or number agreement.

e. **Bayesian syntax learning**: Learners induce (through a Bayesian induction process) a syntax that fits the linguistic input the previous generation produces.

As a proof of concept, we introduce below a toy iterated-learning model of the definiteness effect. In this case, learners' hypothesis space contains various possible grammars with different combinatorial restrictions about what types of DPs may appear in an existential-*there* construction. For simplicity, we focus on the following two hypotheses that will be the most relevant for our purposes (15).

(15) a. G_+: existential *there* is compatible with *a*, *no*, and *every*
b. G_{-every}: existential *there* is compatible with *a* and *no* but not with *every*.

As an overview, our interated-learning analysis of the definiteness effect is schematically illustrated in Fig. 1. Speakers of generation 0 possess G_+, i.e., they consider *every* in an existential *there* construction to be well-formed.[5] However, under our assumption that they are pragmatic speakers, they almost never produce *every* in existential *there* constructions, as it would be tautogical and thus informationally useless. Consequently, some learners in generation 1 never encounter *every* in existential *there* as part of their input data. Assuming that these learners are Bayesian when they are learning syntax, they are likely to induce G_{-every} instead of G_+ since the likelihood of observing an input without any occurrence of *every* in existential *there* is significantly greater in G_{-every} than in G_+. When such learners become the speakers who produce utterances that serve as the input for the learners in the next generation, they will not produce *every* in existential *there* constructions because it is simply ill-formed according to their grammar. The effect of the grammar induction can amplify as the leaning process is iterated through generations.

Below, we spell out the technical details of the iterated learning model.

4.1. Pragmatic speakers in a coordination game

We start with a population whose speakers all have the grammar G_+, i.e., they consider all 3 determiners *no*, *a*, and *every* syntactically compatible with ETCs. They need to use ETCs to communicate to each other facts about the world. Let u_D be the utterance *there is D NP*. For concreteness, we consider 4 possible worlds and use w_i ($i = 0,1,2,3$) to represent the world where the cardinality of the set of the relevant entities denoted by the NP is *i*. Given the semantics of the determiners and ETCs, we have the following truth conditions (16).

(16) $[\![u_{\text{no}}]\!] = \{w_0\}, \quad [\![u_{\text{a}}]\!] = \{w_1, w_2, w_3\}, \quad [\![u_{\text{every}}]\!] = \{w_0, w_1, w_2, w_3\}$

In order to communicate with each other effectively about the actual world, pragmatic speakers will tend to choose a true utterance that is most informative. Concretely, we adopt a probabilistic pragmatic speaker model from the Rational Speech Act (RSA) framework (Goodman and Frank, 2016), according to which a literal listener randomly chooses a world where the utterance is true (17) and a pragmatic speaker favors utterances that maximize the chance of a literal listener choosing the actual world (18).

[5] This is a simplifying assumption and not a crucial commitment of our analysis. We will return to this issue in §7.

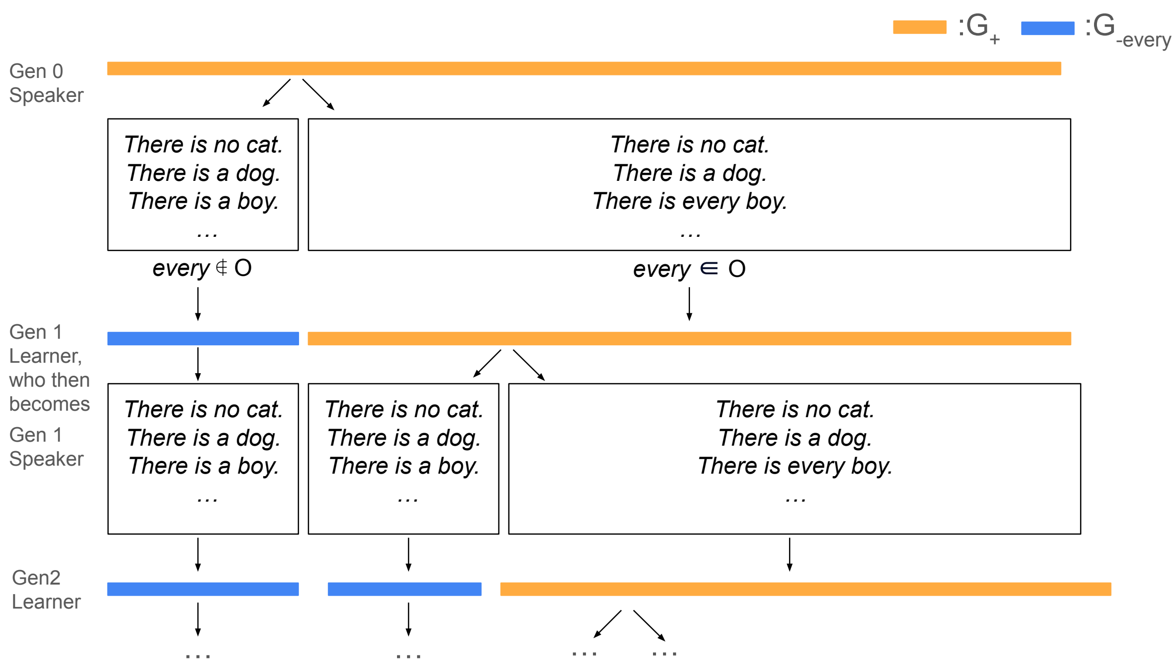

Figure 1: A schematic illustration of the iterated learning model for the definiteness effect

(17) $L_{lit}(w \mid u) = K \cdot \Pr(w)$ if $w \in [\![u]\!]$, and 0 otherwise
(where $\Pr(w)$ is the prior probability of w, and K is a normalizing constant ensuring that literal listener's probabilities over worlds add up to 1)

(18) $S(u \mid w) = K' \cdot [L_{lit}(w \mid u)]^{\alpha}$
(where α is a free parameter controlling how much the pragmatic speaker favors utterances that are more informative, and K' is a normalizing constant ensuring that speaker's probabilities over utterances add up to 1)

For concreteness, we assume a uniform prior over worlds and $\alpha = 10$. Now, as an example, suppose w_3 is the actual world. The probability of the literal listener choosing w_3 is 0 after hearing u_{no} (since u_{no} is not true in w_3), $\frac{1}{3}$ after hearing u_{a} (since u_{a} is true in w_1, w_2 and w_3), and $\frac{1}{4}$ after hearing u_{every} (since u_{every} is true in all the worlds). Then, the probability of the pragmatic speaker choosing u_{a} in w_3 is

$$S(u_{\text{a}} \mid w_3) = \frac{(\frac{1}{3})^{10}}{0^{10} + (\frac{1}{3})^{10} + (\frac{1}{4})^{10}} \approx 0.947.$$

That is, even though u_{a} and u_{every} are both true in w_3, a pragmatic speaker would nevertheless prefer u_{a} because it is more informative. In general, since tautological expressions are, by definition, true in every possible world, they are the least likely to be used by the pragmatic speaker among the true expressions.[6]

[6] Technically, the standard formulation of a literal listener (17) is undefined for a contradictory utterance. One quick fix is to assume that the literal listener will default to the prior in this case, i.e., when they hear a contradictory utterance they simply ignore it. This way, contradictory and tautological utterances are treated in parallel.

Similarly, suppose w_0 is the actual world. The probability of the literal listener choosing w_0 is 1 after hearing u_{no} (since u_{no} is only true in w_0), 0 after hearing u_{a} (since u_{a} is not true in w_0), and $\frac{1}{4}$ after hearing u_{every} (since u_{every} is true in all the worlds). Then, the probability of the pragmatic speaker choosing u_{no} in w_0 is

$$S(u_{\text{no}} \mid w_0) = \frac{1^{10}}{1^{10} + 0^{10} + (\frac{1}{4})^{10}} \approx 0.999999.$$

The probability of a pragmatic speaker using an utterance in a given world is summarized below (19). The main takeaway is that a pragmatic speaker is very unlikely to use *every* in an ETC because the resulting meaning is tautological and hence uninformative.

(19)

$S(u \mid w)$	w_0	w_1	w_2	w_3
u_{no}	0.999999	0	0	0
u_{a}	0	0.947	0.947	0.947
u_{every}	0.000001	0.053	0.053	0.053

It is worth reiterating that this point does not hinge on the specific RSA framework we adopted here. Any satisfactory pragmatic theory should make qualitatively the same prediction.

4.2. Bayesian learners

So far our pragmatic speaker model provides the production probabilities in one instance of a coordination game. We further assume that learners in the next generation observe the outputs of pragmatic speakers playing the coordination game N times for each world. For concreteness we assume $N = 10$, which means that a learner observes 40 utterances of ETCs. The probability that *every* is never used among these 40 utterances is

$$P(u_{\text{every}} \notin O) = [S(u_{\text{no}} \mid w_0)]^{10} \cdot [S(u_{\text{a}} \mid w_1)]^{10} \cdot [S(u_{\text{a}} \mid w_2)]^{10} \cdot [S(u_{\text{a}} \mid w_3)]^{10} \approx 0.195.$$

That is, about 80% of the learners in the next generation observes *every* in an ETC at least once in their input. Such learners will acquire the same grammar G_+ as the previous generation, since this is the only grammar compatible with the input. In contrast, the remaining 20% of the learners never observe *every* in an ETC in their input. For such learners, both G_+ and G_{-every} are compatible with their input data. As we will see, assuming that they are Bayesian learners, they will take their input data to overwhelmingly favor G_{-every}.

Before introducing the technical details of Bayesian learning, let us start with a simple example that helps illustrate the intuition behind it. Suppose you can observe random samples (with replacement) from an unknown set of single-digit numbers. For simplicity, let us assume that you know the set either contains all the digits (i.e., from 0 to 9; call it H_1) or only contains the even numbers (call it H_2). Now, suppose you observe the following sequence of samples (20).

(20) 2, 4, 8, 4, 0, 6, 2, ...

How would you weigh the two hypotheses as you observe more and more samples? Presumably, when you observe the first sample, you will not form a very strong preference for either hypothesis. However, as you observe more and more samples, e.g., 7 even numbers in a row in the case above, you will probably think that the evidence strongly favors H_2 (according to which the set only contains even numbers), even though technically your observations are still compatible with both hypotheses. Intuitively, the reason is that if the set contained both even and odd numbers, it would be quite surprising that you observed 7 even numbers in a row, whereas if the set only contained even numbers, then 7 even numbers in a row would be totally expected. Formally, we can use Bayes' theorem to calculate the *posterior probability* of a hypothesis H given observation O as follows (21).

(21) $P(H \mid O) = \frac{1}{P(O)} \cdot P(H) \cdot P(O \mid H).$

Here, $P(H)$ is the *prior* probability of the hypothesis, $P(O \mid H)$ is the *likelihood* of the observation under the hypothesis, and $P(O)$ is the probability of the observation. When we compare two hypotheses, we can use (21) to calculate the odds between their posterior probabilities as follows (22).

(22) $\frac{P(H_2|O)}{P(H_1|O)} = \frac{P(H_2)}{P(H_1)} \cdot \frac{P(O|H_2)}{P(O|H_1)}$

On the right-hand side of the equation, the first term $\frac{P(H_2)}{P(H_1)}$ is the ratio between the prior probabilities of the two hypotheses, which corresponds to our initial preference between the two hypotheses. The second term $\frac{P(O|H_2)}{P(O|H_1)}$ is the ratio between the likelihoods of the observation under the two hypotheses, which reflects the extent to which the observation O provides evidence for H_2 over H_1: the higher this ratio, the stronger the observation favors H_2 over H_1.

For instance, let O_1 be the initial observation of a single sample 2. The probability of observing O_1 is $1/10$ under H_1, and $1/5$ under H_2. The strength of the evidence $\frac{P(O_1|H_2)}{P(O_1|H_1)}$ is $\frac{1/5}{1/10} = 2$, which only slightly favors H_2 over H_1.

In contrast, let O_7 be the observation of the 7 samples in (20). The probability of observing O_7 is $(1/10)^7$ under H_1, and $(1/5)^7$ under H_2. The strength of the evidence $\frac{P(O_7|H_2)}{P(O_7|H_1)}$ is $\frac{(1/5)^7}{(1/10)^7} = 2^7 = 128$, which strongly favors H_2 over H_1. Even if we have a rather strong initial preference for H_1, e.g., we initially believe that H_1 is 10 times as likely as H_2, after observing O_7, the ratio between the posterior probabilities is $\frac{P(H_2|O_7)}{P(H_1|O_7)} = \frac{P(H_2)}{P(H_1)} \cdot \frac{P(O_7|H_2)}{P(O_7|H_1)} = \frac{1}{10} \cdot 128 = 12.8$. That is, now we instead believe that H_2 is over 10 times more likely than H_1.

The discussion above illustrates how Bayesian learning can capture our intuition about different hypotheses as evidence accumulates in this simple example. More generally, Bayesian learning has been applied to concept learning and word learning (e.g., Xu and Tenenbaum, 2007). A distinctive feature of Bayesian learning is that learners can take lack of occurrences of certain instances in the input to be evidence that such instances are not possible. This allows them to switch between hypotheses even when both hypotheses are compatible with the input.

Now we return to the case of ETCs. We assume that, similar to the case above, learners treat their input as random samples (with replacement) from the underlying grammar. For those learners who never observe *every* among the 40 instances of ETCs in their input, the likelihood is $(1/3)^{40}$ under G_+ and $(1/2)^{40}$ under $G_{-\text{every}}$.[7] We can then calculate the ratio between the

[7] In general, the sentences generated by a grammar are not assumed to be drawn with equal probabilities, but here a uniform distribution is a reasonable first approximation given that the relevant sentences are similar enough to

posterior probabilities as follows (23).

(23) $$\frac{P(G_{-\text{every}}|O)}{P(G_{+}|O)} = \frac{P(G_{-\text{every}})}{P(G_{+})} \cdot \frac{P(O|G_{-\text{every}})}{P(O|G_{+})} = \frac{P(G_{-\text{every}})}{P(G_{+})} \cdot \frac{(1/2)^{40}}{(1/3)^{40}} \approx \frac{P(G_{-\text{every}})}{P(G_{+})} * (1.1 * 10^{7}).$$

We can see that the second term $\frac{P(O|G_{-\text{every}})}{P(G_{+})}$, whose value is $1.1 * 10^{7}$, overwhelmingly favors $G_{-\text{every}}$ over G_{+}. As a result, even if the learners have a high initial preference for G_{+}, practically all the learners will switch to $G_{-\text{every}}$. For instance, suppose the learners initially believe G_{+} is 1000 times as likely as $G_{-\text{every}}$, they will end up believing instead that $G_{-\text{every}}$ is over 10000 times more likely than G_{+}.

In sum, among Generation 1 learners, about 20% of them never encounter *every* in an ETC in their input. Assuming that they are Bayesian learners, they will end up inducing a grammar that rules out such expressions. As discussed above, when such learners become speakers and produce ETCs that will serve as input to learners in the next generation, they will never produce *every* in an ETC since it is simply not part of their grammar. Consequently, learners in the next generation are more likely to observe no instances of *every* in an ETC and induce $G_{-\text{every}}$. As this process iterates over generations, $G_{-\text{every}}$ will take over within the population.

5. Grammatical triviality

In the previous section, we illustrated how meaning-driven ungrammaticality can arise through iterated learning. However, recall that there are semantically trivial sentences that are nevertheless grammatical, such as the ones below (24).

(24) a. Alice is vegetarian and is not vegetarian.
b. It is raining or it is not raining.

Given that such sentences are semantically trivial, pragmatic speakers are also highly unlikely to produce them, and therefore it is quite likely that learners in the next generation will never encounter them in their input. This naturally raises important questions: Why do learners still consider such sentences grammatical? More generally, how does the iterated learning analysis account for the contrast between semantically trivial sentences that are grammatical and those that are ungrammatical?

The key difference, we suggest, lies in the prior probability distribution over the space of grammars that learners entertain. To see the role of prior probabilities in learning, let us return to the earlier example where one observes a sequence of digits and tries to figure out the set from which these numbers were randomly sampled. This time we drop the simplifying assumption that the set either contains all the digits or only contains even numbers. That is, in principle the set can be any subset of the set of digits. In this more general setting, if you observe the following sequence (25), which is the same as before (20) and has 7 even numbers in a row, presumably you would still think that the set is most likely the set of even digits. It is easy to check that this hypothesis is the one that maximizes the likelihood of the observation, and therefore it is unsurprising that we also intuitively find this hypothesis the most plausible.

(25) 2, 4, 8, 4, 0, 6, 2, ... (repeated from (20))

---

each other, which only differ in terms of the quantifier used in the DP.

However, suppose you observe the following sequence (26) instead.

(26) 1, 9, 8, 9, 3, 4, 1, ...

Presumably, you are unlikely to think that the set is $\{1,3,4,8,9\}$, i.e., it contains exactly the five digits that you observe. Note that this set also maximizes the likelihood of the observation, just as the set of even digits does in the previous example. What accounts for the difference between the two cases? Intuitively, this is because in the latter case the set $\{1,3,4,8,9\}$, unlike the set of even digits, is a much more unnatural possibility to consider. Formally, this corresponds to a much lower prior probability for the hypothesis $H_{\{1,3,4,8,9\}}$ than the hypothesis $H_{\{0,2,4,6,8\}}$. As a result, more evidence is required before we begin to favor the hypothesis $H_{\{1,3,4,8,9\}}$.

The main takeaway from the discussion above is that even though Bayesian learners may take the lack of occurrences of certain instances in the input to be evidence for the hypothesis that such instances are impossible, they will not necessarily consider such a hypothesis the most likely. If a hypothesis is considered highly unlikely a priori, or in the extreme case it is simply not part of the space of hypotheses learners entertain, then learners will not conclude that the missing instances in the observation are impossible.

Therefore, by placing constraints on the grammars in the learners' hypothesis space, our approach can also account for the existence of semantically trivial sentences that are presumably absent in the learners' input but nevertheless grammatical. Specifically, to account for cases such as (27a), we suggest that the candidate grammars in the learners' hypothesis space do not distinguish between items from an open class. As a result, given that *vegetarian*, *tired*, and *hungry* all belong to the open class of adjectives, when learners acquire a grammar that generates attested non-trivial sentences such as (27b), it will also generate (27a).

(27) a. Alice is vegetarian and is not vegetarian.
b. Alice is tired and is not hungry.

A couple of remarks are in order. First, our approach need not assume a sharp distinction between open- and closed-class items. The difference can be more gradient in that learners have higher prior probabilities over grammars that distinguish between closed-class items than between open-class ones. And such a difference may be further derivable from the difference in the size of the class. When a class has k items that can be distinguished in terms of whether they are ruled in or not by a grammar, there are 2^k possibilities. For open-class items, the number of actual members is very large and there can be even more members added to the class. It would be intractable to consider all the possible fine-grained distinctions and therefore learners are more likely to assume that the grammar does not distinguish between items of an open class.

Second, we do not claim that the contrast between open- and closed-class items can account for all cases of grammatical constructions that presumably are not in a learner's input. Consider parasitic gaps such as (28). They are very rare: Pearl and Sprouse (2013) found no attested examples among nine child-directed speech corpora that contain approximately $675,000$ words.

(28) Which book did you judge _ before reading _?

While strictly speaking it is still an open empirical question whether parasitic gaps are completely absent in children's input and they still manage to deem such constructions grammatical, for the sake of the argument let us assume that this is indeed the case. This cannot be captured by the contrast between open- and closed-class items, but our approach is compatible with there being additional constraints on the learners' hypothesis space such that they may be able to induce a grammar that rules in parasitic gaps based on observations of other grammatical constructions attested in their input. For instance, Pearl and Sprouse (2013) suggest that Across-The-Board constructions such as (29), which do occur in children's input, may facilitate the acquisition of parasitic gaps.

(29) What did you read _ and then review _?

6. Comparison with the Logicality approach

In terms of the theoretical assumption about the architecture of natural language grammar, our proposal is crucially different from the logicality approach in that it maintains the classic separation of syntax and semantics. According to our approach, ungrammatical trivial sentences are simply *syntactically ill-formed* according to the synchronic grammar. This straightforwardly accounts for the empirical fact that their predicted trivial meanings are opaque to non-specialist speakers (30), because strictly speaking, such expressions cannot be interpreted according to the synchronic grammar.

(30) **Opacity of the trivial meanings:** The predicted trivial meanings of ungrammatical sentences are not intuitively accessible by non-specialist speakers.

For instance, after hearing a grammatical tautological sentence, one can use *duh* to express that it is trivially true (31a). However, it is infelicitous to use *duh* in response to *every* in an existential *there* construction (31b).

(31) a. A: Alice is vegetarian or not vegetarian.
B: Duh.
b. A: *There is every smiling cat.
B: #Duh.

In contrast, according to the logicality approach, ungrammatical trivial sentences are syntactically well-formed and interpretable according to the synchronic grammar. Their unacceptability is due to natural logic filtering out certain sentences with trivial meanings. To account for the opacity of the trivial meanings of such sentences, proponents of natural logic will have to stipulate that this module operates at the subconscious level that blocks intuitive access to the trivial meanings. Note that the issue is not about the *effability* of the meanings of the relevant expressions. It is not at all problematic to assume that non-specialist native speakers understand the meaning of *every smiling cat* as a generalized quantifier and the meanings of *or* and *not* as Boolean functions at the subconscious level, without being able to articulate what such expressions mean. The crucial point is that in the case of grammatical sentences such as *every smiling cat left* and *Alice is vegetarian or not vegetarian*, speakers are able to intuitively access the truth conditions that allow them to form truth value judgments and identify the trivial truth conditions the latter has. This is not the case for ungrammatical trivial sentences such as **there is every smiling cat*. The truth conditions of such sentences are opaque to the speakers. Under the logicality view, it is unclear why we should expect this correlation between logicality (however it is to be characterized) and opacity to hold.[8] We do not take this to be a knockdown

[8]Relatedly, the rescaling-based account needs to resort to the possibility of a RESCALE-based interpretation to

argument against the logicality approach, but we point out that the correlation between opacity and logicality is largely left unexplained.[9]

Despite this major difference, our iterated-learning analysis also shares similar assumptions with the logicality approach. In particular, to account for the contrast between grammatical and ungrammatical sentences with trivial meanings, we make use of the distinction between open- and closed-class expressions. This distinction is also made by Del Pinal (2019) and is closely related to the distinction between logical and non-logical vocabulary in the logical-skeleton-based approach. However, as mentioned earlier, our approach is amenable to a more gradient view of the open- and closed-class contrast.

Finally, our approach can go beyond cases that involve strict propositional triviality (cf. Schwarz and Simonenko, 2018). The iterated learning process can filter out a construction as long as (i) it is highly unlikely to be used by pragmatic speakers so that learners in the next generation are likely to encounter no instances in their input, and (ii) the hypothesis space of the learners' grammars and the prior probabilities are such that Bayesian learners will conclude from the lack of occurrences of the construction that the construction is ungrammatical. A construction can satisfy these conditions without being strictly speaking semantically tautological or contradictory. For instance, Theiler et al.'s (2019) meaning-driven explanation of the incompatibility between neg-raising predicates and interrogative complements (8b) is based on the triviality of the propositional content of the construction relative to its *Excluded Middle* presupposition. In order to apply the logicality approach, they need to make strong theoretical assumptions about the nature of the excluded middle presupposition, including postulation of a null logical term contributing the presupposition (Theiler et al., 2019: 108-110). However, under the iterated learning approach, the core insight of their analysis can be maintained without such commitments. Regardless of the nature of the excluded middle inference triggered by neg-raising predicates, i.e., whether or not it is a presupposition and whether or not it is instantiated by a null logical term, it is in tension with the propositional content when neg-raisers take interrogative clauses by making it uninformative. This is enough to make such constructions highly unlikely to be produced by pragmatic speakers, and consequently they end up being ungrammatical after the iterated learning process.

7. Further discussion

7.1. Assumptions about the initial stage of iterated learning

The iterated learning model presented in §4 assumes that semantically trivial sentences started out grammatical in the initial population and gradually became ungrammatical through iterated learning. That is, the model assumes that there was a point in the past where sentences such as **there is every smiling cat* were considered grammatical by the English speakers back then, and only became ungrammatical for speakers at later stages through language change. This

account for some cases of grammatical triviality, even if it is not the intuitive interpretation of the sentence. Our account does not require computation of such a non-actualized alternative interpretation to account for grammatical triviality. They are simply grammatical according to the synchronic grammar.

[9] Chierchia (2013: 53) attempts to explain the correlation by appealing to the complexity of the relevant expressions. However, the correlation between complexity and opacity is tentatively suggested and remains to be worked out in full detail.

assumption can feel quite implausible, since intuitively it seems unlikely that there was a stage where *all* English speakers would consider such sentences grammatical with a trivial meaning.

However, this assumption is for expository purpose only and not a crucial commitment. The iterated learning approach is compatible with different assumptions about the initial stage. For instance, we may instead assume that in the initial stage, existential *there*-constructions had not been introduced to the English grammar of the vast majority in the population, and a small group of speakers started the linguistic innovation by using various DPs in such constructions, with the following meaning (32), repeated from (3).

(32) $[\![\textit{there be det NP}]\!] \Leftrightarrow [\![\textit{det}]\!]([\![\textit{NP}]\!])(D)$ (D: the domain of individuals)

Given that determiners such as *a* and *no* would lead to informative utterances that facilitate successful communication, such constructions would be used a lot and gradually picked up by the rest of the population. In contrast, given that *every* in an existential *there* construction has a trivial meaning, it was unlikely to be used. Consequently, it would be unlikely to be picked up by the rest of the population. And even if a small fraction of speakers did pick it up, it would eventually die out through the same iterated learning process in §4.

7.2. Phenomena of meaning-driven unacceptability may be heterogeneous

The iterated learning approach crucially assumes that the target combinatorial restrictions can be accounted for in terms of syntax in the synchronic grammar. In some cases, e.g., the selectional restrictions of clause-embedding predicates, it is straightforward to spell out the relevant syntactic rules. However, this might not be a reasonable assumption for some of the phenomena discussed in §2. For example, it is broadly agreed in the linguistic community that the distribution of NPIs is semantic/pragmatic in nature (e.g., Kadmon and Landman, 1993; Krifka, 1994; Chierchia, 2013).

We consider it possible that the phenomena of meaning-driven ungrammaticality are heterogeneous and our iterated learning approach is applicable to only a subset of such cases.

8. Conclusion

In this paper, we proposed a new analysis of the link from semantic triviality to ungrammaticality, which models the iterated learning process in diachronic change. The model involves two main ingredients. First, pragmatic speakers are unlikely to produce semantically trivial sentences, and as a result such sentences are lacking in the linguistic input for the learners in the next generation. Second, Bayesian syntax learners may induce a grammar that rules out certain trivial sentences from their lack of occurrences—although this is not always the case due to constraints on the hypothesis space of grammars, allowing for grammatical sentences with trivial meanings. The effect amplifies through generations.

Our account does away with natural logic as an additional grammatical module assumed in the logicality hypothesis. It straightfowardly accounts for the opacity of the predicted trivial meanings of the ungrammatical constructions. The account may be further extended to allow for a more gradient view of the open- and closed-class contrast and apply to grammatical restrictions that concerning sentences that are not strictly trivial.

References

Abrusán, M. (2014). *Weak island semantics*, Volume 3. OUP Oxford.

Abrusán, M. (2019). Semantic Anomaly, Pragmatic Infelicity, and Ungrammaticality. *Annual Review of Linguistics 5*(1), 329–351.

Barwise, J. and R. Cooper (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy 4*(2), 159–219.

Bonnay, D. (2008). Logicality and invariance. *Bulletin of Symbolic Logic 14*(1), 29–68.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry 37*(4), 535–590.

Chierchia, G. (2013). *Logic in Grammar: Polarity, Free Choice, and Intervention*. Oxford: Oxford University Press.

Culbertson, J. and S. Kirby (2016, January). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in Psychology 6*.

Del Pinal, G. (2019). The Logicality of Language: A new take on Triviality, "ungrammaticality", and Logical Form. *Noûs 53*(4), 785–818.

Fox, D. and M. Hackl (2006). The universal density of measurement. *Linguistics and Philosophy 29*, 537–586.

Gajewski, J. (2002). On Analyticity in Natural Language. Ms., Massachusetts Institute of Technology.

Gajewski, J. (2008). More on quantifiers in comparative clauses. In *Semantics and linguistic theory*, pp. 340–357.

Gajewski, J. (2009). L-triviality and grammar. Handout of a talk delivered at the UConn Logic Group.

Goodman, N. D. and M. C. Frank (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences 20*(11), 818–829.

Grimshaw, J. (1979). Complement selection and the lexicon. *Linguistic Inquiry 10*(2), 279–326.

Kadmon, N. and F. Landman (1993). Any. *Linguistics and Philosophy 16*, 353–422.

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), 10681–10686.

Kirby, S. and J. R. Hurford (2002). The emergence of linguistic structure: An overview of the iterated learning model. *Simulating the evolution of language*, 121–147.

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

Krifka, M. (1994). The semantics and pragmatics of weak and strong polarity items in assertions. In *Semantics and linguistic theory*, pp. 195–219.

Lahiri, U. (1998). Focus and negative polarity in hindi. *Natural language semantics 6*(1), 57–123.

Mayr, C. (2019). Triviality and interrogative embedding: context sensitivity, factivity, and neg-raising. *Natural language semantics 27*, 227–278.

Milsark, G. (1977). Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis 3*(1), 1–30.

Oshima, D. Y. (2007). On factive islands: pragmatic anomaly vs. pragmatic infelicity. In *Annual Conference of the Japanese Society for Artificial Intelligence*, pp. 147–161. Springer.

Pearl, L. and J. Sprouse (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition 20*(1), 23–68.

Pistoia-Reda, S. and U. Sauerland (2021). Analyticity and modulation: Broadening the rescale perspective on language logicality. *International Review of Pragmatics 13*(1), 1–13.

Qing, C. and W. Uegaki (in prep.). Meaning-driven ungrammaticality through iterated learning.

Schwarz, B. and A. Simonenko (2018, December). Factive islands and meaning-driven unacceptability. *Natural Language Semantics 26*(3-4), 253–279.

Smith, K. and E. Wonnacott (2010). Eliminating unpredictable variation through iterated learning. *Cognition 116*(3), 444–449.

Theiler, N., F. Roelofsen, and M. Aloni (2019, June). Picky predicates: Why believe doesn't like interrogative complements, and other puzzles. *Natural Language Semantics 27*(2), 95–134.

Uegaki, W. (2023). *Question-Orientedness and the Semantics of Clausal Complementation*. Springer.

Uegaki, W. and Y. Sudo (2019). The **hope-wh* puzzle. *Natural Language Semantics 27*(4), 323–356.

van Benthem, J. (1989). Logical constants across types. *Notre Dame Journal of Formal Logic 30*(3), 315–342.

van Benthem, J. (2002). Logical constants, the variable fortunes of an elusive notion. In W. Sieg, R. Sommer, and C. Talcott (Eds.), *Reflections on the foundations of mathematics: Essays in honor of Solomon Feferman*, pp. 420–440. AK Peters, Ltd Natick, MA.

von Fintel, K. (1993). Exceptive Constructions. *Natural Language Semantics 1*(2), 123–148.

Xu, F. and J. B. Tenenbaum (2007). Word learning as Bayesian inference. *Psychological Review 114*(2), 245–272.

Zuber, R. (1982). Semantic restrictions on certain complementizers. In *Proceedings of the 12th International Congress of Linguists, Tokyo*, pp. 434–436.