Learnability and constraints on the semantics of clause-embedding predicates

Learnability and constraints on the semantics of clause-embedding predicates

Mora Maldonado (mora.maldonado@univ-nantes.fr)
Laboratoire de Linguistique de Nantes
CNRS-Université de Nantes

Jennifer Culbertson (jennifer.culbertson@ed.ac.uk)
School of Philosophy, Psychology, and Language Sciences,
University of Edinburgh

Wataru Uegaki (w.uegaki@ed.ac.uk)
School of Philosophy, Psychology, and Language Sciences,
University of Edinburgh

Abstract

Responsive predicates are clause-embedding predicates like English *know* and *guess* that can take both declarative and interrogative clausal complements. The meanings of responsive predicates when they take a declarative complement and when they take an interrogative complement are hypothesized to be constrained in systematic ways across languages, suggesting that these constraints represent semantic universals. We report an artificial language learning experiment showing that one of these proposed constraints is indeed reflected in the inferences participants make while learning a novel responsive predicate. Our results add support to a growing body of evidence linking semantic universals to learning.

Keywords: semantic universals; clause-embedding predicates; responsive predicates; artificial-language learning; semantics

Introduction

Despite the immense diversity of human languages, researchers have identified a number of common properties which recur across languages. Of particular interest here are commonalities in the meaning of lexical items, for example colour terms, kinship terms, and quantifiers (e.g. Berlin & Kay, 1969; Murdock, 1949; Barwise & Cooper, 1981; Keenan & Stavi, 1986). These so-called lexical-semantic universals are argued to provide a window into some of the core properties of natural language semantic systems. Understanding their nature is therefore a fundamental goal in semantics. One prominent hypothesis concerning the nature of lexical-semantic universals is that they arise from mechanisms active during **learning**. That is, universal semantic properties of lexical items arise because meanings that possess these properties are easier to learn or generalize, and are therefore more likely to get lexicalised. Indeed, several recent studies have provided substantial evidence for this hypothesis with respect to a number of semantic universals. These include universals in the domain of quantifiers (Chemla, Buccola, & Dautriche, 2019; cf. Hunter & Lidz, 2013), colour terms (Steinert-Threlkeld & Szymanik, 2020), personal pronouns (Maldonado & Culbertson, 2021) and evidentiality (Saratsli, Bartell, & Papafragou, 2020).

In this study, we extend this line of investigation to a new domain: **clause-embedding predicates**—verbs like *know* and *guess* in English, which can be immediately followed by a clause as in ⌜*I know that it's raining out*⌝. Recent developments in formal semantics suggest that clause-embedding predicates behave in a unified way. In particular, several potential lexical-semantic universals have been argued to hold in this domain (Spector & Égré, 2015; Theiler, Roelofsen, & Aloni, 2018; Uegaki, 2019; Roelofsen & Uegaki, 2020). Here, we use an **artificial language learning paradigm** to investigate whether one such constraint—clausal-distributivity—is reflected during learning. To preview, we find that adults learning a novel clause-embedding predicate in the lab infer that Clausal-distributivity holds without explicit evidence.

Theoretical background

Some clause-embedding predicates, like English *know* and *guess*, can take both declarative and interrogative clausal complements. Following Lahiri (2002), we call these predicates **responsive predicates**. The behavior of responsive predicates is exemplified in (1) and (2) with *know* and *guess*, where the predicates take a declarative complement headed by *that* in (1a), (2a) and an interrogative complement headed by *whether* in (1b), (2b).

(1) a. Jo knows *that* it was raining outside.
     b. Jo knows *whether* it was raining outside.

(2) a. Jo guessed *that* it was raining outside.
     b. Jo guessed *whether* it was raining outside.

Since Karttunen (1977), a major question for the semantics of clause-embedding predicates is the relationship between the interpretation of a given responsive predicate when it embeds a **declarative** complement (as in (1a),(2a)) and when it embeds an **interrogative** complement (as in (1b),(2b)). That is, how to state the interpretation of "V-wh" in terms of "V-that" and vice versa. We can narrow down hypotheses concerning this relationship in terms of a *constraint* on the meaning of responsive predicates. In this section, we will discuss two such constraints: **Veridicality uniformity** and **Clausal-distributivity**.

Veridicality uniformity

Veridicality uniformity, due to Spector and Égré (2015), intuitively states that a responsive predicate meaning either 'refers to truth' with respect to *both* declarative and interrogative complements, or it does not refer to truth under *either* type of complement. For example, roughly, *know* refers to

237

truth both in (1a) and in (1b): (1a) entails that it is in fact raining, and, (1b) entails that Jo knows the correct answer to the question of whether it is raining (if it's in fact raining, (1b) entails that Jo knows that it's raining; if it's not raining, (1b) entails that Jo knows that it's not raining). By contrast, *guess* does not refer to truth either in (2a) or in (2b): (2a) does not entail that it is raining, and (2b) does not entail that Jo guesses the correct answer to the question of whether it is raining.[1]

Compare these predicates to a fictitious predicate *shknow*, which means 'know' when it takes a declarative complement and 'wonder' when it takes an interrogative complement (due to Spector & Égré, 2015). This predicate would not satisfy Veridicality uniformity, since it refers to truth when it takes a declarative (just like *know*), but not when it takes an interrogative complement (similar to *guess*).

Formally, Veridicality uniformity can be defined in terms of the notion of veridicality, as follows:[2]

(3) **Definition: veridicality**

   a. A predicate V is **veridical w.r.t. declarative complements** iff (for every x and every declarative complement p) $\ulcorner x\ Vs\ that\ p \urcorner$ entails $\ulcorner p \urcorner$.

   b. A predicate V is **veridical w.r.t. interrogative complements** iff (for every x, every interrogative complement Q, and every declarative complement p corresponding to Q's true answer) $\ulcorner x\ Vs\ Q \urcorner$ together with $\ulcorner p \urcorner$ entails $\ulcorner x\ Vs\ that\ p \urcorner$.

(4) A predicate is **Veridicality uniform** iff it is either veridical w.r.t. both declarative and interrogative complements, or it is non-veridical w.r.t. both declarative and interrogative complements.

If Veridicality uniformity is a cross-linguistic constraint on responsive predicate meanings, then we do not expect to find predicates like *shknow* in any natural language (Spector & Égré, 2015; Steinert-Threlkeld, 2020).

Clausal distributivity

Theiler et al. (2018) propose another constraint on responsive predicate meanings, **Clausal distributivity** (henceforth, **C-distributivity**), which also picks out predicates like *know* and *guess* as possible, but *shknow* as impossible. For predicates that satisfy C-distributivity, the same relationship between the predicate and the complement holds for both declarative and interrogative embedding contexts. In particular, the meaning

of a sentence in which the predicate embeds an interrogative can always be paraphrased in terms of a specific answer to the question denoted by the interrogative complement. This can be formally stated as follows:

(5) A predicate V is **C-distributive** iff (for every x and every interrogative complement Q) $\ulcorner x\ Vs\ Q \urcorner \Leftrightarrow$ there is an answer p to Q such that $\ulcorner x\ Vs\ p \urcorner$

For example, the predicate *know* satisfies C-distributivity because (1b) means that there is some answer such that Jo *knows* that answer (i.e., (1a)). The predicate *guess* also satisfies this constraint because (2b) means that there is some answer such that Jo *guesses* that answer (i.e., (2a)).

By contrast, *shknow* does not satisfy C-distributivity because 'Jo wonders whether it is raining' does not mean the same thing as there is some answer such that Jo *knows* that answer (i.e., that it is raining or that it's not raining, whichever is true). Thus, if we restrict our attention to these three predicates, C-distributivity makes the same prediction as Veridical uniformity. In the next section, we review two existing studies that aim to systematically evaluate the empirical validity of these two constraints.

Existing empirical studies

Compared to the rich theoretical literature on these constraints (e.g., Spector and Égré, 2015; Theiler et al., 2018), relatively few attempts have been made to assess their validity on empirical grounds. However, there are two notable exceptions: Roelofsen & Uegaki (2021) and Steinert-Threlkeld (2020).

Roelofsen & Uegaki (2021) investigated the cross-linguistic validity of several constraints including Veridicality uniformity and C-distributivity based on a survey of semantic descriptions in the existing literature. Their survey shows that both constraints make correct predictions for a large class of responsive predicates across languages. Nevertheless, there are several apparent counterexamples, suggesting that the constraints may not be categorical but rather probabilistic.

Steinert-Threlkeld (2019) tested Veridicality uniformity in learnability experiments using a simple feed-forward neural network. He defined the lexical semantics of four predicates: *know*, *be certain*, *wondows*, and *knopinion*, where the former two satisfy Veridicality uniformity and the latter two are fictitious predicates that violate it. The network was trained on sentence-truth value pairs for each of the predicates. The results showed that the model learned the two predicates that satisfied Veridicality uniformity faster than the two that didn't. Assuming that the behavior of the learning model tracks human learners, the study suggests that Veridicality uniformity may hold across languages because predicates satisfying it are easier to learn.

While these two strands of empirical work represents an important first step, two important issue remain. First, neural network models are idealized learners, and their behavior does *not* always track human behavior (e.g., see Linzen &

---

[1]See Tsohatzidis (1993) and Spector and Égré (2015) for detailed empirical arguments for this claim. In particular, they show that, even though (2b) sometimes imply that Jo made a correct guess, this is not a mandatory semantic entailment, as evidenced by the felicitous continuation *...but, she turned out to be wrong*.

[2]See Theiler et al. (2018) for more precise definitions and arguments for adopting the particular definitions. Specifically, technically, the interrogative complement Q in (3b) has to be one that is 'exhaustivity-neutral' (i.e., admits only one true answer). This is to avoid a complication that arises from cases of the so-called 'mention-some' reading (e.g., $\ulcorner$*Jo knows where she can buy an Italian Newspaper*$\urcorner$ can be true even if she knows only one of many places that sell Italian newspapers).

Baroni, 2021; McCoy, Pavlick, & Linzen, 2019). Second, existing evidence does not tease apart the validity of Veridicality uniformity and that of C-distributivity. Roelofsen & Uegaki's survey does not provide strong evidence in favour of one constraint over the other as a cross-linguistic universal. Steinert-Threlkeld's results are not conclusive in this respect, either, as his fictitious predicates, *wondows* and *knopinion* violate C-distributivity as well as Veridicality uniformity.

Current study

Our main goal is to provide behavioral data from human subjects that serves to identify constraints on responsive predicates. Specifically, we assess the hypothesis that responsive predicates satisfy C-distributivity (and not necessarily Veridicality uniformity). From this hypothesis, we derive a novel learning-based prediction: when learning a new responsive predicate, learners will infer that it is C-distributive.

To test this hypothesis, we use an artificial language learning paradigm where learners have to generalize (i.e. *extrapolate*) from ambiguous evidence (a.k.a 'Poverty of Stimulus' method; Wilson, 2006; Culbertson & Adger, 2014; Maldonado & Culbertson, 2020). Participants in our experiment are first taught the meaning of two declarative-embedding predicates, both unattested in English as lexical items:

(6) a. FALSEBEL: *x falsebel that p* 'x falsely believes p'
   b. KNOWFALSE: *x knowfalse that p* 'x has correct knowledge that p is false'.

Participants are then required to assign a meaning to the interrogative-embedding version of the predicate. That is, to *extrapolate* its interrogative-embedding meaning from what they learned about its declarative-embedding meaning. The question is whether they extrapolate the meanings in line with C-distributivity.

If FALSEBEL is C-distributive, ⌜*Jo falsebel whether it's raining*⌝ is true only in situations where Jo believes a false answer to the question of whether it's raining. This is so because, assuming C-distributivity, ⌜*Jo falsebel whether it's raining*⌝ is true iff Jo falsely believes that it's raining or she falsely believes that it's not raining. By contrast, if KNOWFALSE is C-distributive, *Jo knowfalse whether it's raining* is true only in situations where Jo believes a *true answer*. This is so because, assuming C-distributivity, *Jo knowfalse whether it's raining* is true iff Jo has a correct belief that it is not raining or she has a correct belief that it is raining. Thus C-distributivity predict distinct patterns of response (as outlined in more detail below) for these two predicates.

Note that Veridicality uniformity does not make any prediction about the interrogative-embedding meanings for FALSEBEL and KNOWFALSE. This is so because the predicates, as defined in (6), are non-veridical w.r.t. interrogative complements (and thus satisfy Veridicality uniformity) no matter what their interrogative-embedding meanings turn out to be. To see this, consider whether ⌜*x falsebel Q*⌝ together with ⌜*p*⌝ entails ⌜*x falsebel p*⌝. Given the meaning

of ⌜*x falsebel p*⌝ as defined in (6a), we know that this entailment doesn't hold because ⌜*x falsebel p*⌝ is false if ⌜*p*⌝ is true, regardless of the meaning of ⌜*x falsebel Q*⌝. The same reasoning holds for ⌜*x knowfalse Q*⌝. More intuitively, since the two predicates are defined to refer to false answers, ⌜*Jo falsebel/knowfalse whether it is raining*⌝ will never entail ⌜*Jo falsebel/knowfalse the **true** answer*⌝, because the latter cannot be the case regardless of what the true answer is. Hence, our study allows us to isolate the empirical predictions of the C-distributivity constraint. If C-distributivity holds, then we predict two distinct response patterns for the two predicates. If only Veridicality uniformity holds, then no such prediction is made.

Methods

This experiment, including predictions, design, and analysis was preregistered here.

Design and materials

Participants were randomly assigned to one of two conditions. In both conditions, they learned a new verb *lem*. During training, this verb was combined with declarative complements like ⌜*Jo lems that is it raining*⌝. Conditions differed in the meaning of this declarative-embedding. In one condition, the meaning of *lem* corresponded to KNOWFALSE, and in the other condition to FALSEBEL (see (6)). After learning this declarative-embedding meaning of *lem*, participants in both conditions were required to assign an interpretation to the interrogative-embedding version of the predicate.

All trials consisted of a sentence together with a scenario. Scenarios depicted (a) a fictional character Jo, who is choosing what to wear among three possible options: sun, rain and snow equipment (right-side of pictures in Figure 1); and (b) the actual weather outside (left-side of pictures in Figure 1). Jo's choices were designed to be indicative of her beliefs about the weather.

Sentences could involve declarative or interrogative complements. Declarative-embedding sentences had the form ⌜*Jo lems that p*⌝, where p is one of [*it's raining outside*, *it's sunny outside*, *it's snowing outside*].[3] The meaning of the predicate was conveyed by showing participants whether these sentences could be used in 5 different types of scenarios, summarized in Table 1.

Interrogative-embedding sentences—seen only during testing—had the form ⌜*Jo lems Q*⌝, where Q is *what the weather is like*.[4] Participants had to decide whether the sentence ⌜*Jo lems Q*⌝ could be used in the following three sce-

---

[3]To maximize naturalness, declarative sentences were introduced by the connective 'and' in the FALSEBEL (as Figure 1) and by the connective 'but' in the KNOWFALSE condition.

[4]We opted to use a constituent *wh*-clause ⌜*what the weather is like*⌝ rather than a *whether*-clause like ⌜*whether it is raining*⌝ in order to maximise the surface syntactic difference between declarative and interrogative clauses, and thus to rule out the possibility that participants mistake an interrogative complements for a declarative clause (and vice versa) during the testing phrase (but see Discussion).
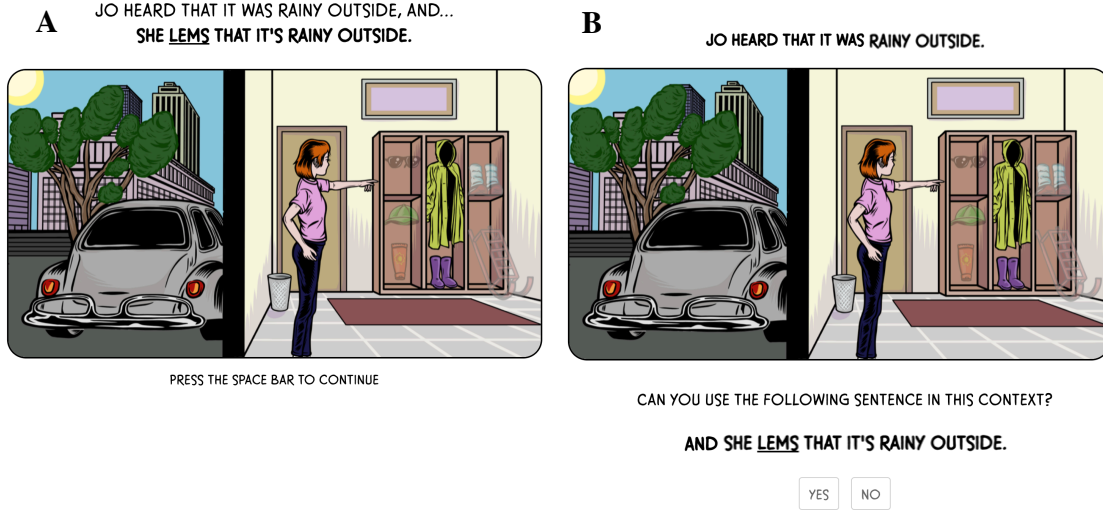
Figure 1: Example trials for Exposure (**A**) and Acceptability (**B**) in the FALSEBEL condition.

Scenarios	FALSEBEL	KNOWFALSE
¬p and Jo believes that p	✓	✗
¬p and Jo believes that ¬p	✗	✓
¬p/p and Jo has no belief wrt p	✗	✗
p and Jo believes that ¬p	✗	✗
p and Jo believes that p	✗	✗

Table 1: Scenarios used to convey the declarative-embedding meaning of the sentence ⌜*Jo lems that p*⌝. The ✓ sign indicates that the sentence could be used in the scenario; ✗ that it could not.

narios: (i) When Jo believes a true answer to *Q* (True answer); (ii) When Jo believes a false answer to *Q* (False answer); and (iii) When Jo has no belief (No answer). Examples of these scenarios are provided in Figure 2. Learners who infer that *lem* is C-distributive in the each condition were expected to accept the sentence ⌜*Jo lems Q*⌝ in different scenarios, as illustrated in Table 2.

Scenario	FALSEBEL	KNOWFALSE
(i) True answer to Q	✗	✓
(ii) False answer to Q	✓	✗
(iii) No answer to Q	✗	✗

Table 2: Responses satisfying C-distributivity by condition. A response in the FALSEBEL condition is compatible with C-distributivity if ⌜*Jo lems Q*⌝ is accepted (✓) in a situation where Jo believes a false answer to Q and rejected elsewhere (✗). In the KNOWFALSE condition, compatibility with C-distributivity occurs when ⌜*Jo lems Q*⌝ is accepted (✓) in situation where she believes a true answer to Q and rejected elsewhere (✗).

Procedure

Participants were instructed that they would be learning a new predicate *lem* by observing situations in which it could be used. They were told that all situations involve Jo, who is getting ready to go outside and may or may not have correct beliefs about the weather. They were given a hint that *lem* does not mean 'to know' or 'to think'.

The experiment had two phases: the training phase and the testing phase. In the training phase, participants were taught the declarative-embedding meaning of *lem*. The training consisted of (a) an exposure block, where participants were presented with scenarios together with the sentence ⌜*Jo lems that p*⌝ (positive evidence only; Fig.1A); and (b) an acceptability block, where participants were shown the six scenarios in Table 1 and were asked to decide whether the sentence ⌜*Jo lems that p*⌝ could be used truthfully (Fig.1B). There were 12 exposure and 56 acceptability trials. Participants were given feedback on their answers during acceptability trials, i.e., they were given both positive and negative evidence.

In the testing phase, participants were told that they were going to meet Jo's mother, Pam, who would ask yes/no questions about Jo using the verb *lem*. Participants had to answer Pam's questions on the basis of the scenarios. Crucially, these questions included cases where *lem* takes both declarative and interrogative complements (Fig.2). There were 27 testing trials (15 involving interrogative-embedding *lem*).

Participants were finally asked to complete a short questionnaire asking (i) whether they had a strategy for completing the experiment, (ii) what they thought the novel verb meant, and (iii) what languages they have experience with.

Participants

196 English-speaking adults were recruited via Prolific (FALSEBEL= 85; KNOWFALSE= 111). Participants were paid 9 GBP/hour for their participation which lasted on average 15 minutes. Per our pre-registration, we only included in the analysis participants who successfully learned the declarative-embedding meaning of the predicate. Specifically, only participants who achieved a mean accuracy of at

Figure 2: Example of testing trials in (**A**) True answer , (**B**) False answer and (**C**) No answer scenarios (see Table 2 for reference).

least 66% on each type of scenario during training were included in the analysis. This is because if participants have not learned the declarative-embedding meanings, then they cannot meaningfully extrapolate to interrogative-embedding contexts. This resulted in the analysis of 40 participants in FALSEBEL and 30 in KNOWFALSE.[5]

Results

Recall that participants in our experiment were exposed to a novel predicate *lem*, which could mean either FALSEBEL or KNOWFALSE. Participants were taught the declarative-embedding meaning of this predicate, but were provided with no evidence of its interrogative-embedding meaning. At test, participants were asked to interpret sentences where the predicate *lem* took an interrogative complement. We were interested in whether the meaning that participants assigned to these sentences satisfied C-distributivity.

Fig.3 shows the proportion of responses compatible with C-distributivity by condition. Responses were considered compatible with C-distributivity as a function of condition and specific scenario, as indicated in Table 2. A visual inspection of Fig.3 suggests that participants in both conditions tended to infer that the predicate satisfies C-distributivity, although participants in the FALSEBEL condition appear to infer C-distributivity more consistently than participants in the KNOWFALSE condition.

Following our pre-registered plan, we evaluated this pattern statistically.[6] A logistic mixed-effects regression model, including random intercepts per subject (nested by condition) and scenario, revealed that the proportion of trials in which *lem* was treated as satisfying C-distributivity is significantly above chance (β =3.502; p =.0024).[7]

---

[5]Independent of the question of C-distributivity, KNOWFALSE turned out to be very difficult to learn even just w.r.t. declarative complements. For this reason we were not able to collect our pre-registered sample size. In the Discussion, we discuss possible reasons why KNOWFALSE turned out to be difficult to learn.

[6]Analyses were performed using the lme4 package in R (R Core Team, 2018)

[7]Given that we did not predict an effect of condition, we did not include this categorical predictor in our confirmatory analysis.
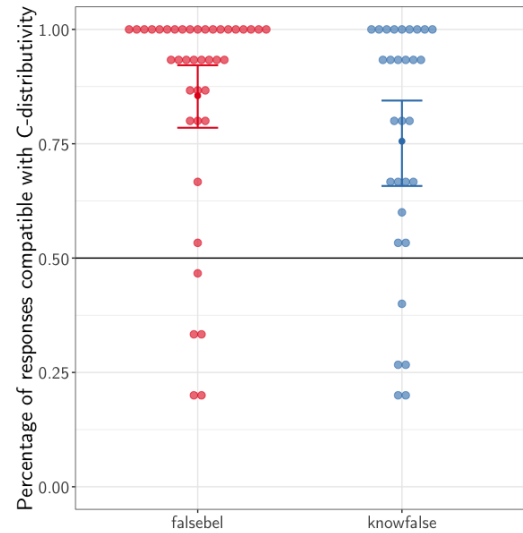

Figure 3: Proportion of responses compatible with C-distributivity by condition. Dots represent individual participant means. Error bars represent standard error on by-participant mean.

Although we did not predict an effect of condition, we conducted a secondary analysis to evaluate whether this finding was stronger in the FALSEBEL condition (as suggested by Fig.3). To do this, we fit a model predicting responses by Condition. The effect of Condition was only marginally significant (p = 0.063), indicating that a preference for meanings that satisfy C-distributivity holds regardless of the specific predicate (i.e., FALSEBEL or KNOWFALSE).

Discussion

In this experiment, we used an artificial language learning experiment to investigate learners' implicit assumptions about semantic constraints on responsive predicates. Our aim was to test whether C-distributity, which constrains the relation between declarative- and interrogative-embedding meanings of responsive predicates (Theiler et al., 2018), is at play during learning. We trained learners on the declarative-embedding meaning of a novel response predicate, and then tested their inferences about the corresponding interrogative-

embedding meaning. We found a strong tendency for learners to infer this meaning in a way that satisfies C-distributivity. This held across the two predicates we tested (meaning 'falsely believe' or 'correctly disbelieve'), suggesting that a bias for C-distributivity does not depend on the specific meaning of the predicate.

Moreover, compatibility with C-distributivity predicts different response patterns for each of these meanings (see Table 2): under C-distributivity, ⌜*Jo falsebel what the weather is like*⌝ is true only if Jo believes a wrong answer to the question. The pattern is the opposite for ⌜*Jo knowfalse what the weather is like*⌝.[8] The fact that a preference for C-distributive meanings is found in both conditions suggests that participants' inferences are dependent on training, and not on some general strategy (e.g., similarity with a native predicate).

Notably, our results cannot be straightforwardly accounted for by an alternative constraint argued to hold for responsive predicates, Veridicality uniformity. As noted above, both FALSEBEL and KNOWFALSE predicates satisfy this constraint. Consequently, Veridicality uniformity doesn't make any prediction about what participants' inferences should be in our experiment. In particular, Veridicality uniformity does not predict different response patterns between conditions: the interrogative-embedding meaning of both predicates is non-veridical, so there is no reason why they should be different. However, while our findings can only be fully accounted by C-distributivity, they do not suggest that a Veridicality uniformity constraint is not *also* at play. Showing this would require, for example, a follow up experiment testing predicates for which Veridicality uniformity makes distinct predictions and C-distributivity does not.

Finally, it is worth returning to the difference in learnability of the two declarative-embedding meanings. Both of the novel predicates were quite hard to learn, as evinced by high exclusion rates in both conditions (50% in FALSEBEL and 70% in KNOWFALSE). This is not in principle problematic for our conclusion, since we were not testing learnability of the predicates. Rather, we were interested in how participants extrapolate to a novel interrogative-embedding meaning *once they have learned* the declarative-embedding meaning. However, there seems to be an extra cost associated with KNOWFALSE, and at the same time, at marginally weaker preference for inferring C-distributivity. It is worth considering

whether the higher exclusion rates for the KNOWFALSE condition could potentially confound the interpretation of our extrapolation results, at least for this condition.

We analyzed the debriefing questionnaires, in which participants reported what they believed the predicates meant. Responses revealed that most participants who were excluded in the FALSEBEL condition treated the predicate as meaning 'believe' or 'know'. This is further reflected in their errors during training: they accept the use of the predicate in every scenario where Jo believes p, regardless of whether p is true or false. By contrast, excluded participants in the KNOWFALSE condition reported thinking that the predicate meant 'disbelieve' or 'disagree', and accordingly they accepted sentences such as ⌜*Jo knowfalse that it's raining outside*⌝ in scenarios where it's raining but Jo thinks it's not raining.[9]

This suggest that participants may find it hard to draw the inference that the complement of the predicate is false (i.e. anti-factivity). This difficulty was more pronounced for KNOWFALSE than for FALSEBEL. This may be due to the different availability of similar meanings in English. In many contexts, the English predicate 'to believe' triggers the pragmatic inference that the complement is false. For example, ⌜*Jo believes that I have a sister*⌝ often implies that ⌜*I don't have a sister*⌝ (Chemla, 2008). While this is not the lexical meaning of 'believe' (i.e., ⌜*Jo believes that p*⌝ is compatible with *p* being true), the existence of such inference suggests that this anti-factive interpretation is accessible to English speakers. In contract, there is no English predicate that licenses an inference equivalent to KNOWFALSE, possibly explaining the asymmetry between conditions. There is no reason to believe that this difficulty would somehow lead participants to make C-distributive responses in the testing phase for a spurious reason. Rather, it might explain why more participants in the KNOWFALSE produced non-C-distributive response patterns.

Conclusion

A growing body of research has been interested in investigating the link between learning and universal features of lexical semantics. Here we have extended this to a novel class of lexical-semantic universal—constraints on embedding predicates—which have been hypothesized based on cross-linguistic data. We tested whether one such universal, Clausal distributivity, can be explained by a bias active during learning. Clausal distributivity is a constraint on the relation between declarative- and interrogative-embedding meanings of responsive predicates (e.g., English *know* and *guess*). We taught English-speaking participants novel response predicates in a declarative-embedding context, and tested how they extrapolated the declarative-embedding meaning to interrogative-embedding contexts. Our results show that learners assume, in the absence of any explicit evidence,

---

[8]Our training is compatible with two slightly different interpretations of KNOWFALSE: (a) '*x* has correct knowledge that *p* is false'; and (b) '*p* is false and *x* believes that *p* is false'. In the main text, we have assumed interpretation (a). Under interpretation (b), e.g., ⌜*Jo knowfalse that it is rainy*⌝ is compatible with a situation where it is in fact sunny and Jo believes that it is snowy. Under this interpretation of KNOWFALSE, C-distributivity would predict that ⌜*Jo knowfalse what the weather is like*⌝ is true iff Jo has a correct belief about one of the possible answers to *Q* that it is false (but not for *all* possible answers). In the experiment, this would make the interrogative-embedding sentence compatible with both correct and wrong answers to *Q*. No participant in our experiment accepted interrogative-embedding sentences in both scenarios. We assume that this is due to the fact that most participants have the interpretation (a) of KNOWFALSE (and not (b)).

[9]Most participants who did pass the training provided correct paraphrases for the meaning of *lem*. Debriefing results can be found in the OSF repository.

that Clausal distributivity holds. This finding suggests that a Clausal-distributivity constraint might drive inferences during natural language acquisition, thus providing a mechanism explaining this cross-linguistic tendency. In future research, we hope to extend this approach to speakers of other languages to show that this pattern emerges independently from learners' native language.

Acknowledgments

References

Barwise, J., & Cooper, R. (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, *4*(2), 159–219. doi: 10.1007/978-94-009-2727-8_10

Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Oakland, CA: University of California Press.

Chemla, E. (2008). An epistemic step for antipresuppositions. *Journal of Semantics*, *25*(2), 141–173.

Chemla, E., Buccola, B., & Dautriche, I. (2019). Connecting content and logical words. *Journal of Semantics*, *36*(3), 531-547. doi: 10.1093/jos/ffz001

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*(16), 5842–5847. doi: 10.1073/pnas.1320525111

Hunter, T., & Lidz, J. (2013). Conservativity and Learnability of Determiners. *Journal of Semantics*, *30*(3), 315–334. doi: 10.1093/jos/ffs014

Karttunen, L. (1977). Syntax and Semantics of Questions. *Linguistics and Philosophy*, *1*(1), 3–44. doi: 10.1007/BF00351935

Keenan, E. L., & Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, *9*(3), 253–326.

Lahiri, U. (2002). *Questions and Answers in Embedded Contexts*. Oxford: Oxford University Press.

Linzen, T., & Baroni, M. (2021). Syntactic structure from deep learning. *Annual Review of Linguistics*, *7*, 195–212.

Maldonado, M., & Culbertson, J. (2020). Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry*, 1-42.

Maldonado, M., & Culbertson, J. (2021). Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry*, 1–71. doi: 10.1162/ling_a_00406

McCoy, T., Pavlick, E., & Linzen, T. (2019). Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3428–3448).

Murdock, G. P. (1949). *Social structure*. New York: Macmillan.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Roelofsen, F., & Uegaki, W. (2020). Searching for a universal constraint on the denotation of clause-embedding predicates. In *Proceedings of semantics and linguistic theory 30* (pp. 542–561).

Saratsli, D., Bartell, S., & Papafragou, A. (2020). Cross-linguistic frequency and the learnability of semantics: Artificial language learning studies of evidentiality. *Cognition*, *197*, 104194.

Spector, B., & Égré, P. (2015). A uniform semantics for embedded interrogatives: *An* answer, not necessarily *the* answer. *Synthese*, *192*(6), 1729–1784. doi: 10.1007/s11229-015-0722-4

Steinert-Threlkeld, S. (2020). An explanation of the veridical uniformity universal. *Journal of Semantics*, *37*(1), 129–144.

Steinert-Threlkeld, S., & Szymanik, J. (2020). Ease of learning explains semantic universals. *Cognition*, *195*, 104076.

Theiler, N., Roelofsen, F., & Aloni, M. (2018). A uniform semantics for declarative and interrogative complements. *Journal of Semantics*, *35*(3), 409–466.

Tsohatzidis, S. L. (1993). Speaking of truth-telling: The view from *wh*-complements. *Journal of Pragmatics*, *19*, 271–279.

Uegaki, W. (2019). The semantics of question-embedding predicates. *Language and Linguistics Compass*, *13*(1), 1–17.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982.