An experimental assessment of the nall lexical gap

Mora Maldonado, Benjamin Spector, Ruizhe Zhou, Marie-Léa Le Clainche

Across languages, the concept *not all* (*nall*) is never lexicalized. Existing accounts for this gap fall into two main categories. In *Cognitive Markedness* accounts (*CM*), *nall* is not lexicalized because it expresses a *cognitively unnatural concept* (Katzir & Singh 2013). In *Communicational Efficiency* accounts (*CE*), *nall* is not cognitively marked, but is not very useful in communication, and its absence reflects a pressure towards communicative efficiency. (Enguehard & Spector 2021, Uegaki 2024).

Recently, CM accounts for other lexical constraints (e.g., convexity.) have been tested in studies involving *rule learning* or *artificial language learning* (Chemla et al 2019, Maldonado & Culbertson 2021). The underlying assumption is that words or rules involving cognitively unnatural meanings are harder to learn than those involving natural ones. In this study, we aim to disentangle the two accounts by conducting two artificial language learning experiments that test whether learners are more likely to infer that a novel word means *nall* rather than *some*. The CM account predicts that learners prefer to lexicalize meanings akin to 'some' rather than meanings akin to 'nall'. The CE account predicts that learners are as likely to adopt a system that lexicalizes 'nall' as one that lexicalizes 'some'.

We use an extrapolation paradigm. Learners are first taught two novel words: one is compatible with *all* scenarios, and one is compatible with *none* situations. At test, learners must decide which of the two novel words they would use to describe a 'mixed', *some-but-not-all* (*sbna*) scenario. If participants prefer to have a word that means 'some', they will use the same word for *all* and *sbna* scenarios. Conversely, if they prefer to lexicalize 'not all', they will use the same word for *no* and *sbna* scenarios.

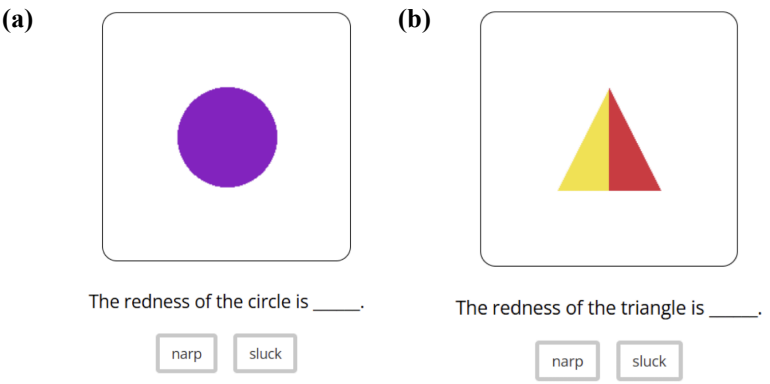

Fig 1. Training and testing (Adj. condition). (a) This is a *seen* trial involving a *none* situation, as the circle is not red. (b) This is a *target* trial, involving a *snba* situation, as some part of the triangle is red.

Method English-speaking participants were told they would learn two words used to describe colored shapes. We manipulated whether novel words were determiners, adjectives or verbs. Sentences were constructed as function of the condition: (a) Determiner condition: [<NOVEL DETERMINER> of the <SHAPE> is <COLOR>] (e.g. *Idho of the triangle is red*), (b) Adjective condition: [The <COLOR>-ness of the <SHAPE> is <NOVEL ADJECTIVE>] (e.g. *The redness of the triangle is narp*), and (c) Verb condition : [The <SHAPE> <NOVEL VERB> <COLOR>] (e.g. *The circle pakes blue*). The two novel words were randomly selected for each participant. The experiment consisted of a training and a testing phase. During training, participants were taught to use each novel word to describe one of two scenarios: (1) one in which the shape mentioned is entirely filled with the specified color (*all* scenario) and (2) one in which the shape is entirely filled with a different color (*none* scenario; see Fig. 1a). Feedback was provided throughout the training, and shapes and colors varied trial to trial. The testing phase was similar to training but without feedback. It included trials involving *all* and *none* scenarios, identical to those used during training (henceforth called '*seen* trials'). Crucially, we also included *target* trials involving a *some-but-not-all* scenario, where the mentioned shapes were half-filled with the specified color and half-filled with a different color (Fig. 1b).

The prediction for the target trials is that if *nall* is cognitively marked, participants would prefer, on average, the word which was compatible with *all* scenarios during training. More specifically, this would suggest that learners prefer a system that lexicalizes a *some* and a *no* meaning, rather than one that lexicalizes a *nall* and an *all* meaning. These preferences are expected to hold regardless of the grammatical category of the novel words.

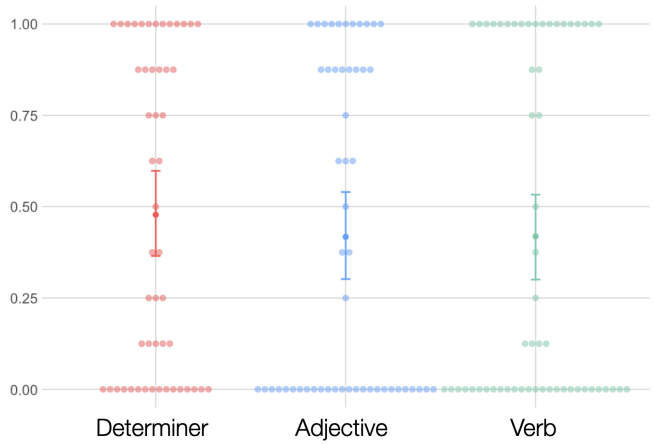

Fig 2. Proportion of responses compatible with a *some* meaning in target trials during testing.

Results 180 English-speaking participants were randomly assigned to one of the three conditions (Det., Adj., Verb). Participants were excluded if they had fewer than 11/16 correct trials for either *all* or *none* situations during training and testing. This ensures that participants have correctly learned the situations compatible with each word. Fig. 2 shows the frequency with which participants in each condition selected for *sbna* situations the word they had learned for *all* situations. Recall that this response pattern suggests that participants are treating one of the novel words as meaning 'some.' We ran a Bayesian binomial mixed-effects model to evaluate whether the proportion of responses compatible with a *some* meaning is above what one would expect by chance. The model also included as a fixed effect the category of the novel words (Det vs. Adj vs. Verb). We included random intercepts for participants and by item. This model shows that participants preferred to assign a *nall* rather than a *some* meaning to a novel word (intercept =-1.087, 90% CI = [-2.09, -0.107], SE = 0.612, P(intercept > 0) = 0.03). When looking at the difference between conditions, we find moderate to strong evidence that the use of a novel determiner brings the overall mean up (P(beta > 0)=0.78), and moderate evidence that the use of a novel adjective or verb brings the general mean down (P(beta < 0)=0.74).

Discussion Across all three conditions, participants showed no preference for a *some* meaning, challenging the cognitive markedness account of quantifier (non-)lexicalization. Surprisingly, if anything, we found the opposite: a general preference for a *nall* meaning, although the origins of this preference remain unclear. Interestingly, this preference was weaker when the novel words were determiners rather than adjectives or verbs. This slight variation likely reflects the influence of participants' native language. For instance, in English, the most natural completion for *The redness of the circle is ...* might involve the adjectives 'complete' and 'incomplete', which could explain the observed preference for *nall*. However, such a preference was also noted in the verb condition, where the target words could be interpreted as meaning 'is' or 'is not' (as suggested by some participants during the debriefing). It is not straightforward why the negated copula should be preferred for *sbna* situations. Even more surprising are the results in the determiner condition. Here, participants could interpret one of the novel words as meaning 'some' (e.g., *Some of the circle is red*), leading to more *some*-consistent responses. However, the preference for *some*-responses is only slightly higher than in the adjective and verb conditions. Overall, these results argue against the Cognitive Markedness hypothesis. **References**

Chemla, Buccola & Dautriche, I. (2019). Connecting content and logical words. *J. of Semantics*, *36*(3), 531-547. | Enguehard & Spector (2021). Explaining gaps in the logical lexicon of natural languages: A decision theoretic perspective on the square of Aristotle. *Semantics & Pragmatics*, 14(5), 1-31. | Katzir & Singh. (2013). Constraints on the lexicalization of logical operators. *L&P*, 36, 1–29.| Maldonado & Culbertson (2022). Person of interest: Experimental investigations into the learnability of person systems. *LI 53*(2), 295-336. | Uegaki (2024). The informativeness/complexity trade-off in the domain of Boolean connectives. *LI*, 55(1), 174–1