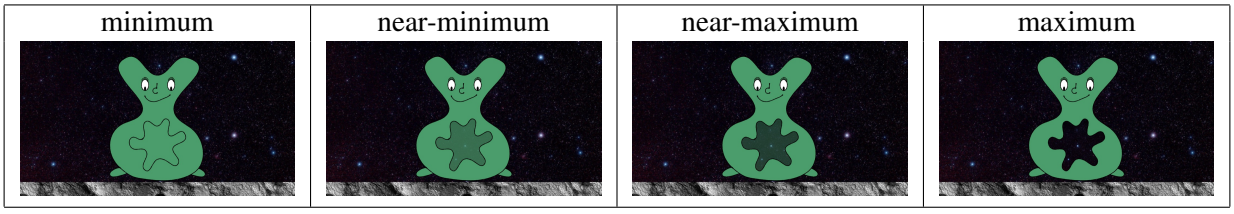
Generalizating NPIs to positive uses in an Artificial Language Jeremy Kuhn and Mora Maldonado

Overview Negative Polarity Items (NPIs) are characterized by a polarity-sensitive use, restricted to downward entailing (DE) environments. However, in many languages, the very same lexical items also have positive uses that appear in upward entailing (UE) environments. For example, under negation, the NPI *any* has an existential meaning (*I didn't talk to anybody* = *I didn't talk to a single person*). But certain UE environments allow *any* to appear with a universal (or free choice) meaning (*I talked to anyone was interested* = *I talked to everyone who was interested*). Similar positive uses can be found for other NPIs in English (*ever*, *yet*, *anymore*) and other languages (e.g. French *encore*).

Ladusaw (1979) observes that the positive and negative uses of NPIs are often systematically related: they are *logical duals*. If an NPI is licensed by (and scoping under) negation, its positive counterpart carries the meaning the word would need to receive if it were interpreted as scoping above the negation, in order to derive the same sentential meaning. For *any*, existential force under negation becomes universal force (since $\neg\exists = \forall\neg$). This observation offers a potential diachronic explanation of the systematic ambiguity. When the syntactic distribution of a logical item is restricted so that it always appears in the presence of negation, the meaning of the item is ambiguous between two denotations depending on its scope relative to negation: A¬ or ¬B. The typological data above can then be explained by the hypothesis that when the use of an NPI is extended to new, UE environments, an attractive interpretation is the wide-scope dual meaning (A).

Using an artificial language learning paradigm, we test how learners generalize the meaning of NPIs when they appear in positive environments. We teach English speaking participants an artificial language which includes a negative marker *em* ('not') and a degree modifier *tup*, roughly equivalent to English 'at all'. During training, participants are exposed to sentences in which *tup* is restricted to negative sentences (i.e. *tup* never occurs without *em*). At test, participants are asked to interpret sentences where the degree modifier appears on its own, without negation. We evaluate whether learners are more likely to assign a universal meaning (as attested in the typological data) or an existential meaning to this sentence.

Methods Participants were taught a miniature language consisting of four predicates, four proper nouns, one negative marker and one degree modifier. All predicates denote gradable properties with closed scales (e.g., transparent/opaque). For each property, we define four possible scale points: 'minimum', 'near-minimum', 'near-maximum', 'maximum,' as shown below for the noun *Greenie* and the predicate *pleet*:



Participants were first trained on the following non-target sentences: (i) *simple positive (SP)* (e.g., 'Greenie pleet'), used when the predicate applies to a maximum or near-maximum degree; (ii) *simple negative (SN)* (e.g., 'Greenie em pleet'), used for minimum or near-minimum degrees; and (iii) *negative NPI (Neg-NPI)* (e.g., 'Greenie em pleet tup'), used only when the predicate applies to a minimum degree. Crucially, participants had no evidence of the use of the degree modifier in absence of negation. At test, participants were asked to interpret these *positive NPI (Pos-NPI)* held-out sentences (e.g., 'Greenie pleet tup'). Participants had to decide whether the Pos-NPI sentence can be used when the noun applies to the predicate to a near-minimum degree or to a maximal degree. These two choices correspond to the two dual meanings that could be posited for the NPI: existential and the universal, respectively. After test, we asked subjects for translations of all four sentence types. (This experiment was preregistered here.)

Results 49 English speaking participants were recruited on Prolific and successfully trained on non-target sentences (i.e. accuracy rates above 75%). Fig. 1 (left side) shows the proportion of trials on which participants chose the 'maximum-degree' meaning for Pos-NPI sentences during the test phase. A logit mixed-effects model showed that the proportion of responses compatible with these maximum meanings is significantly below chance ($\beta = -1.6; p = .0137$), revealing an overall preference for 'near-minimum' meanings. However, a visual inspection of Fig. 1 reveals two clusters of participants. While approximately 50% of participants consistently derive 'near-minimum' meanings (driving the statistical effect reported above), a second group, which corresponds to 25% of our sample, systematically select the 'maximum-degree' meaning (binomial test: $p < .05$). This suggests the existence of two populations who generalize in different directions.

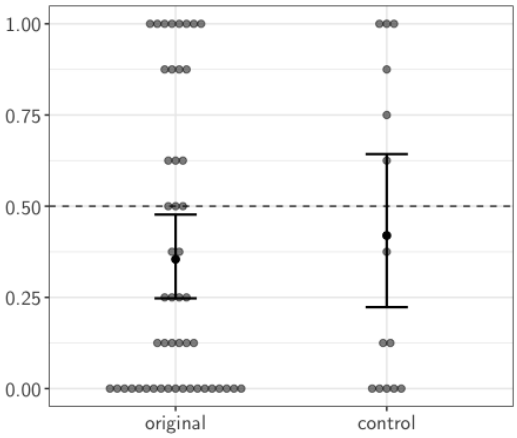


Figure 1: Proportion of 'maximum' responses

Translations provided at the end of the experiment give further insight into the make-up of these two groups. 'Maximum degree' responders systematically translate Pos-NPI sentences as using universal degree modifiers like 'completely' or 'very,' and Neg-NPI sentences as involving either the same words (e.g. 'Greenie is very transparent') or 'at all' (e.g. 'Greenie is not transparent at all'). Among the 'near-minimum degree' responders, translations of Pos-NPI sentences are less consistent. Notably, though, no subjects translated the meaning using existential degree modifiers like 'a bit' or 'somewhat'. On the other hand, a number of participants gave the sentence a *negative* meaning: 'X tup' is translated as 'not X.' For such subjects, the 'near-minimum' meaning is presumably chosen as the one that is comparatively closer to the minimum.

Control To investigate whether 'near-minimum' responses arise from a negative interpretation of the Pos-NPI sentences, we modified the original experiment, replacing the 'near-minimum' choice with a 'minimum' choice in test trials. Fig.1 (right side) shows pilot results for 14 participants. While these results are preliminary, the existence of a group ($\sim$50% of participants) that consistently derive minimum interpretations supports the hypothesis that in both experiments, non-'maximum' responders interpret the NPI as negation.

Discussion The results reported here show that several different strategies are adopted when extending the meaning of NPI items to contexts without a licensor. These strategies correspond to two meaning shifts attested in diachronic typology. First, one group of participants assign a 'maximum degree' interpretation to Pos-NPI sentences, thus displaying a pattern of generalizing the NPI meaning to its wide-scope dual. In the appropriate sociolinguistic contexts, such a population could explain the emergence of positive *any*, *ever*, *yet*, and *anymore*. A second group of participants assigns a 'near-minimum' degree interpretation to Pos-NPI sentences. Translations and a control experiment suggest that this is not due to an existential interpretation, but rather due to a negative interpretation of the NPI, possibly due to a repair strategy with reconstructed negation. Interestingly, this generalization corresponds to Jespersen's cycle (1917), in which a minimizing NPI is reinterpreted as contributing negation itself.

Further syntactic and semantic factors may influence the generalization strategy adopted by participants. Syntactically, a word order that privileges a specific scopal configuration may make a wide-scope dual interpretation more or less accessible. Semantically, properties of the predicates may also affect generalization preferences. In future work, we intend to use the present paradigm to test the strength of these factors, which may make specific predictions about the kind of diachronic change a given language is likely to undergo.

Refs. Jespersen 1917. *Negation in English and other languages*. • Ladusaw 1979. *Polarity sensitivity as inherent scope relations*.