

Homogeneity and enrichability affect scalar processing

Chao Sun (University College London), Ye Tian (Université Paris Diderot), Richard Breheny (University College London)

Recent experimental studies found scalar expressions of different lexical categories generate scalar inferences (SIs) at different rates [1-3]. [3] employed an inference paradigm to test participants' interpretation of statements containing scalar expressions. Several classes of scalar expressions were examined including quantifiers (e.g. <all, some>), modals (<certainly, possibly>), adjectives (<beautiful, pretty>) and verbs (<dislike, loathe>). Participants read a statement, e.g. 'John says: This student is intelligent'. Then they were asked whether or not the speaker implied the negation of the stronger statement in which scalar expression was replaced by its stronger scale mate, e.g. 'Would you conclude from this that, according to John, she is not brilliant?'. They found that a significant variation in the derivation rates of SIs across different supposedly scalar expressions. Quantifiers and modal expressions generate SIs more frequently than adjectives and verbs. Moreover, there was much greater variability within adjectives and verbs than quantifiers and modals. [3] explored a wide range of explanations, and found that semantic distance, measuring the difference in the perceived strengths between the pair of scalar terms accounted for a small but significant amount of the variance in inference rates. Additionally, 'boundedness', related to whether the stronger scalar term referred to an endpoint accounted for over three times more variance than did semantic distance. Although it is clear that both factors played a role, together they only accounted for 22% of the variation found in the inference task, leaving a large amount of variation unexplained. Here we present studies that explore two additional factors that may affect the derivation rates for the items in [3] - scale homogeneity and propensity for local enrichment.

Scale homogeneity relates to the property of expressions to be somewhat underspecified or polysemous in their core meaning(s). For example, 'brilliant' can be understood in terms of different extended senses, such as high degree of intelligences, or charitableness, resourcefulness. Therefore, scalemates of 'brilliant' could vary in terms of which extended sense is evoked. It is possible that scalar pairs in [3], e.g. <brilliant, intelligent>, <beautiful, pretty>, <tight, snug> are not always construed as being on the same entailment scale. In contrast, scalar pairs, e.g. <finish, start>, <all, some> have fairly homogeneous core senses across uses, it would be difficult to construe these terms as not being on the same entailment scale. An important assumption for SIs derivation is that the scalemates must form an entailment relation. We hypothesize that other things being equal, the more homogeneous the sense of the items in a pair, the higher the rate of SI derivation.

The propensity of a scalar term to be 'locally' enriched during utterance comprehension is another factor that may explain the variation. We consider local enrichment as a pragmatic process that strengthens the meaning encoded in a particular lexical item. To some extent, it is a separate mechanism from deriving Gricean SIs. One important driver of local enrichment concerns a principle that the explicit proposition expressed by an utterance is itself relevant [4]. On our view, if the literal meaning of an utterance is not sufficiently relevant for the context, listeners will enrich the literal meaning to yield an adequately relevant interpretation. In [3], sentences are presented out of context, leaving participants wondering how sentences with these weak scalar terms might be relevant. For weak scalar quantifiers and modal expressions (e.g. 'some' and 'possible'), the unenriched meaning is very weak. In contrast, the unenriched meanings for adjectives and verbs (e.g. 'intelligent', 'start') tend to be more specific in terms of where on the scale to fix the interpretation. We conjecture that scalar

expressions differ in their susceptibility to local enrichment; quantifiers and modals are more likely to be enriched compared to adjectives and verbs.

Experiments: In exp. 1, we used the same scalar expressions from [3] to obtain the derivation rate for each scale pair. Participants were asked to rate on a 0-100 scale to indicate to what extent they could infer from the speaker's statement that the speaker does not believe the stronger alternative. We replicated [3]'s general results that the ratings of SI for quantifiers ($M=76.03$) and modals ($M=64.35$) were significantly higher than for adjectives ($M=34.95$, $p=.001$) and verbs ($M=35.30$, $p=.004$); No significant difference was found between quantifiers and modals, or between adjectives and verbs. To investigate whether scale homogeneity or local enrichability can account for some of the variation, we developed the 'but' task (exp. 2) and the 'so' task (exp.3) to operationalise these ideas.

Exp. 2: 'but'-task: we used the same scalar pairs investigated in [3] to construct 43 experimental sentences¹ with the form "X but not Y". 20 participants rated the naturalness of these sentences (each judging a total of 50 sentences), e.g. (1a-b):

1. a. The student is brilliant but not intelligent. <brilliant, intelligent>
- b. The weather is hot but not warm. <hot, warm>

'but' has a denial-of-expectation conventional implicature. Thus a sentence, 'X but not Y' is felicitous to the extent that X can be construed to not strictly entail Y but normally or often to imply Y. A scale with high homogeneity is one where the stronger term is interpreted to entail the weaker term. Entailment relations require that if X entails Y, whenever X holds, Y must hold. Thus these 'X but not Y' sentences should be unnatural if the contrasting predicates X and Y are on the same entailment scale. So if naturalness rating for 'but' sentences is low, it suggests a higher degree of homogeneity for the given scale; whereas if the rating is high, then the degree of homogeneity is relatively low. We predicted that the naturalness rating for scalar expressions in our 'but' task should negatively correlate with the inference task results. As shown in fig.1, the naturalness of the 'but' sentences correlated negatively with the inference task results ($r=-.341$, $p=.025$). This confirmed prediction on scale homogeneity that the higher the degree of scale homogeneity, the higher the rate of SI derivation.

Exp. 2: 'so'-task: we measured the degree to which each of the weak scalar terms is liable to undergo local enrichment. We constructed 43 sentences of the form 'X so not Y'. Another 20 participants rated the naturalness for these sentences e.g. (2a-b).

2. a. The student is brilliant so not intelligent. <brilliant, intelligent>
- b. The weather is hot so not warm. <hot, warm>

'So' implies that the second segment follows in some way from the first. 'X so not Y' suggests that one might expect not Y, given X. Thus, X so not Y sentences should be more coherent to the extent that the weaker scalar expression can be locally enriched to have an upper bound meaning. For example, to understand (2b) as felicitous, 'warm' must have its meaning locally enriched to just warm. In our task, if the naturalness rating for 'so' sentences is low, it suggests that the expression is not liable to be locally enriched; whereas if the rating is high, then it is liable to be locally enriched. We predicted that the naturalness rating for scalar expressions in the 'so' task should positively correlate with the inference task results. As shown in fig.2, the naturalness rating of 'so' task positively correlated with the inference task results ($r=.417$, $p=.005$).

¹ There were 7 filler sentences, which contained clearly felicitous (e.g. The banker is rich but not happy) and clearly infelicitous items (e.g. The man left the party but he never came).

Multiple linear regression analyses were conducted to predict the rating of scalar inferences in exp.1 from scale homogeneity degree, propensity for local enrichment, semantic distance, and ‘boundedness’. The last two factors are the only ones that accounted for significant amount of variances in [3]. We found that when only the scale homogeneity degree used as a predictor, it accounted for 11% of the variance. When propensity for local enrichment was included, two predictors together accounted for 30% of the variance. Propensity for local enrichment accounted for an additional 19% of the variance. Including two factors from [3], the regression model accounted for 66% of the variance. In this model, scale homogeneity degree did not significantly explain the variance because semantic distance scores and ‘but’ task scores are highly correlated.

On-going corpus study: although the scalar diversity effect was shown in [3] is interesting and requires an explanation, their data do not represent how frequently these items trigger an SI, or indeed how diverse they are, in *real use*. For example, the item “some” is rated to strongly trigger SI readings. Therefore the explanations were selected to explain *this* effect. However, [5] has shown in a corpus study that “some” does not frequently trigger an SI reading. We argue that a balanced, large-scale corpus study is a better approach for studying scalar diversity. In a pilot study, we randomly selected 10 items from [3], and extracted sections of dialogue containing the target item from the Switchboard corpus. Using a similar paradigm as in [5], participants rated how similar the statement with the negation of the stronger term to the statement with the weak scalar expression. The pilot study results do not significantly correlate with the inference rating in [3], suggesting that results in [3] do not reflect frequencies of SI readings.

Conclusion: we argue that local enrichability (LE) independently accounts some of the scalar diversity previously reported. We discuss the still open issue what lies behind the lack of independence between scale homogeneity (SH) and semantic distance. To the extent that LE and SH are independent factors, they motivate a deflationary account whereby diversity vis a vis global pragmatic inference is less than previously assumed. By the same token, our results support the notion that LE and quantity implicature arise via separate mechanisms. For future directions, we argue for a large-scale balanced corpus approach for a better understanding of scalar diversity in real use.

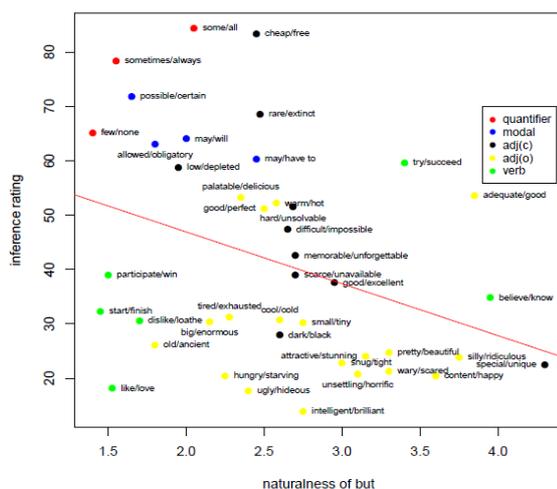


Fig.1 Negative correlation between the absence of homogeneity and inference rate

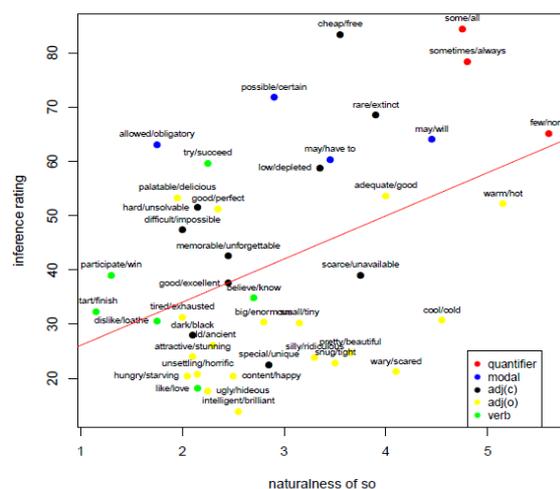


Fig.2 Positive correlation between the propensity of local enrichment and the inference rate