Exploring scalar diversity through priming: A lexical decision study with adjectives

Radim Lacina (radim.lacina@uni-osnabrueck.de)

Nicole Gotzner (nicole.gotzner@uni-osnabrueck.de)
Institute of Cognitive Science, Osnabrück University
Wachsbleiche 27, 49090 Osnabrück, Germany

Abstract

When someone says *My soup was warm*, they are often understood as saying that it was warm, but not hot. This is assumed to arise via a scalar implicature. According to the standard assumption, *warm* and *hot* are in competition and by saying *warm*, we reason that the speaker did not intend to convey *hot*. This exclusion of alternatives should apply uniformly to any expression that can be ordered on a scale. Yet there are substantial differences in the endorsement rates of the strengthened meaning between various scales. These could be due to the availability of expressions or to the underlying semantic structure. We use priming to measure how active in the mind lexical expressions are. Contrary to the standard assumption, the more an expression was primed, the less likely a scalar implicature was endorsed. We discuss how the semantic structure of adjectives can support pragmatic reasoning without lexical alternatives.

Keywords: scalar diversity; lexical decision task; priming; adjectives; boundedness

Introduction

Speakers often use words whose meaning can be mapped onto a scale. A prime example of such words are gradable adjectives (Kennedy & McNally, 2005; Kennedy, 2007):

(1) The chicken soup I was served was warm.

In this example, the hearer needs to map the meaning of *warm* onto a temperature scale which will determine its *lower bound*, i.e. the minimum degree of temperature that counts as warm. This is the basic semantic meaning of *warm*.

When hearing (1) the hearer might also come to the conclusion that while the soup was warm, it was not hot. They may only make this inference if they interpret the utterance as *upper bounded*. This, however, could be cancelled:

(2) The chicken soup I was served was warm. In fact, it was hot.

In the case of (2), this is not a contradiction as the meaning of *warm* also covers the temperature range of *hot*. *Hot* is said to be the informationally stronger term to *warm* on the scale of (upwardly looking) temperature (Horn, 1972), called a *Horn scale*. The interpretation where the stronger term is negated has come to be known as a *scalar implicature* (e.g., Levinson, 1983).

The upper-bounded meaning of scalar terms is standardly assumed to arise via the negation of alternatives. The hearer reasons that the speaker could have said *hot* but since they chose to utter the weaker term *warm*, *hot* should be excluded (Grice, 1975; Horn, 1972). Yet as shown above, the basic meaning of gradable adjectives can be understood in terms of degree scales. Here, we investigate to what extent semantic features underpinning the meaning of adjectives and the availability of lexical expressions explain scalar implicature computation.

Scalar diversity

Scalar implicatures have been studied and discussed extensively both in the theoretical pragmatic and semantic literature and in experimental research in linguistics and psychology. Most of the work has been devoted to studying quantifier scales both in the theoretical (e.g., Horn, 1972; Chierchia, 2004; Fox & Katzir, 2011) and experimental (e.g., Bott & Noveck, 2004; Huang & Snedeker, 2009; Grodner et al., 2010) literature. That *some* is strengthened to mean *some but not all* has served as the founding block for the study of implicatures (Horn, 1972).

This literature tacitly assumes that the mechanism for deriving a scalar implicature is *uniform* and would generalize from quantifiers to other expressions (Doran et al., 2009; Degen, 2015; Van Tiel et al., 2016). Yet the probability of a hearer endorsing a scalar implicature depends on the particular meaning scale in question (Doran et al., 2009, 2012; Van Tiel et al., 2016). Compare the following sentence to (1):

(3) My classmate Sally is intelligent.

Here, it would be improbable that the hearer would derive the scalar implicature that Sally is intelligent, but not brilliant. Van Tiel et al. (2016) have found exactly this in their study of various scales. The <*intelligent*, *brilliant*> scale was reported to have an endorsement rate of the strengthened meaning of less than 10%, compared to the <*warm*, *hot*> scale in (1), for which the rate was over 60%. This phenomenon of variability in scalar implicature has since been dubbed *scalar diversity*. The discovery of this variation in the rates of implicatures has led to much research devoted to explaining the phenomenon. It has been found that there are general differences between various broad types of scalar words (Van Tiel et al., 2016; Sun et al., 2018; Gotzner et al., 2018; van Tiel et al., 2019; Ronai & Xiang, 2021a, 2022; Sun et al., 2023).

There are two classes of explanations that have been put forward. The first one relates to the *availability of alternatives*. Assuming that listeners activate and negate alternatives

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

in scalar implicature reasoning, they should only endorse a scalar implicature if the strong expression is highly active in their mind. Various experimental results have pointed towards the crucial role of the availability of alternatives in scalar implicature processing (e.g., Barner et al., 2011 for acquisition; Gotzner, 2019, Bott & Frisson, 2022, note however, that Van Tiel et al., 2016 did not find an effect of availability on the scalar diversity under their operationalisation). In line with this view, Ronai & Xiang (2021b) found that the scalar diversity effect is smaller when the strong term is presented in a contextual question (QUD). Similarly, Hu et al. (2022) present modeling evidence that the scalar diversity effect can be explained via uncertainty about alternatives. They found that the higher the entropy in their measure of uncertainty for a particular scale was, the lower the implicature endorsement rate was.

Rees & Bott (2018) formalised this availability claim and proposed the *Salience model* for implicature derivation. It consists of three stages, the Alternative stage, the Usage mechanism and the Implicature derived stage. The inputs during the first stage consist of all the variables that might influence the derivation of an alternative, such as the question-under-discussion (QUD), the semantic and socio-pragmatic context as well as speaker knowledge. Through all these things, the activation of the alternative might be higher or lower. There is a threshold of activation between the Alternative and Usage mechanism stages that determines whether the process of implicature derivation starts. It is only when the above-mentioned factors cause the alternative to become activated, or salient, enough that it crosses this threshold and the comprehender moves to the second stage, which ends in the implicature being derived. During this Usage Mechanism stage, the alternative is negated and combined with the literal meaning of the sentence.

The second type of explanation for the scalar diversity effect relates to the *scale structure* underlying the semantics of adjectives. Gotzner et al. (2018) found that the endorsement of scalar implicature varied systematically for different types of adjectives. They tested 70 adjectival pairs and annotated them for the following features: boundedness and extremeness of the strong term, the type of standard of the weak term (minimum or maximum standard, relative), and the polarity of the scale.

Boundedness refers to cases in which on a particular scale, the strong term denotes an endpoint (Kennedy, 2007). For example, the adjective *full* represents the endpoint of a scale, since something cannot be filled to a larger extent than to the one described by *full*. An effect of boundedness was already found in the study of Van Tiel et al. (2016) testing scales of different grammatical categories and bounded adjectives have been found to be associated with higher endorsement rates in Gotzner et al. (2018).

Contrary to boundedness, strong expressions that are extreme like *brilliant* are less likely to be used in scalar implicature reasoning (see also Beltrama & Xiang, 2013). Extreme adjectives are said to be terms for which the degree lies beyond the usual contextual range (Morzycki, 2012).[1] Gotzner et al. (2018) also found that extremeness predicted lower rates of implicature endorsement.

Semantic distance was measured by Gotzner et al. (2018) in a rating study where participants had to indicate on a Likert scale how much stronger they felt a statement involving the strong scalar term was compared to one that included the weak one. Gotzner et al. (2018) and Van Tiel et al. (2016) both found that the higher the distance, the higher the proportion of implicatures for a particular scale in question.

Finally, standard type was either relative or absolute with the latter category dividing into minimum standard and maximum standard adjectives (Kennedy & McNally, 2005; Kennedy, 2007). This was annotated for the weak scale-mate rather than the strong. Gotzner et al. (2018) reported that maximum standard adjectives exhibited lower inference rates compared to relative ones.

Overall, these findings provide evidence that the scale structure underlying the meaning of adjectives contributes to scalar implicature reasoning. While some features make implicature derivation more likely, some hinder the appearance of pragmatic reasoning. There is now a question that arises—how can these influences of semantic structure be explained? One possibility that we aim to test in the current study is the connection to the availability of alternatives and the Salience model. It could be that scales with particular structures, such as those that are bounded, evoke the lexical stronger term more compared to unbounded ones; in other words make it more available during comprehension. If the stronger scale-mate is in turn more available, it is more likely to cross the activation threshold with the process of implicature derivation starting. If this is not the case, the role of scale structure in scalar implicature may result from other aspects such as the way the semantics of different expressions is computed and how it maps onto degree scales (see Gotzner et al., 2018).

This leads to two predictions, namely that (a) the higher availability of the strong term ought to result in higher rates of implicatures for the scale in question, and (b) that the semantic features associated with higher inference rates should be positively associated with increased availability in processing, according to alternative-based accounts. This in turn raises the question of how availability in processing might be operationalised in order to be testable. We propose to use the activation of the strong term in the mind of the comprehender as measured in a lexical decision task (priming). Below, we discuss the literature on the priming of meaning alternatives in implicature processing.

The priming of alternatives

Recently, researchers have attempted to directly test the presence of scalar alternatives in online implicature derivation.

[1] Whether or not the strong term for a particular scale was extreme or not was tested for by means of compatibility with modifiers such as *downright*. For example, it is acceptable to say *downright excellent*, but not *downright good*.

De Carvalho et al. (2016) ran the first study aiming to examine the activation of strong scalar terms. What they found was that weak terms (*some*) activated their stronger scale mates (*all*) more than the strong ones did for the weak. They interpreted these findings as evidence for the psychological reality of lexical scales as envisaged in the theoretical work of Horn (1972). While the research of De Carvalho et al. (2016) did reveal that informational strength relations play a role in real-time comprehension, their experimental design presented these scalar words outside of a context where they could give rise to implicatures.

Ronai & Xiang (2023) addressed this issue in their experiments aiming to see whether weak scalar terms activated their stronger counterparts within sentential contexts and outside of them. They exposed their native English participants to sentences such as *Zack's carpet was dirty/patterned*. Following 650ms after the final, a critical word, which was either a weak scalar (*dirty*) or an unrelated one (*patterned*), the participants saw the target word *filthy*, which was the informationally stronger scale-mate of the related prime. They found that these stronger terms were recognised as existent English words quicker when the preceding sentence contained the weak scalar prime compared to an unrelated prime. This was taken as evidence for the weak scale-mate activating the stronger one. In order to test whether this activation was specific to scalar implicature derivation processes, Ronai & Xiang (2023) also tested the same combination of prime and target words when presented as isolated lexical items (i.e. without any sentential context). In this experiment, they found no evidence of priming. This, they argued, suggested that the activation seen in the sentential experiment was indeed due to implicature processes.

What we may take from these studies is that strong scalar terms are active during comprehension when preceding stimuli include their weaker scale-mates, that this effect is specific to contexts that can support an implicature, and that there is asymmetric priming based on informational strength. All of these findings suggest that we may use the activation of the strong term by its weaker scale-mate as compared to an unrelated baseline as a measure of availability.

The question of whether this activation of the strong scalar term (measured by priming) can be linked to the rate at which a given scale gives rise to implicature has already been raised by Ronai & Xiang (2023) but the study did not find a significant correlation between priming and inference rates. One potential reason is that the studies used varying sentence frames, which might have led to certain scales getting more contextual support for inference derivation compared to others.

The current study

In the present study, we aimed at addressing the question of whether the differences in the semantic features of various adjectival scales influence the degree to which weak adjectives activate their stronger scale mates in the course of comprehension and whether they are directly linked to whether or not comprehenders derive a scalar implicature. We investigate (a) whether the semantic features of scales predict the availability of strong terms by their weak scale-mates operationalised by priming and (b) whether this activation strength for each scale predicts its observed endorsement rate of the implicature meaning. We use a similar methodology to Ronai & Xiang (2023) but we keep the sentence frames constant and we focus on one grammatical class, that is adjectives.

We are interested here in whether the property of boundedness and of the extremeness of the stronger term play a role. We also examine whether the type of adjective (relative vs. minimum or maximum standard) and the semantic distance between the weak and strong terms have an effect on the level of activation of the strong term, as measured by the size of priming in lexical decision. We then test whether those same scales that show priming to a larger extent are also the scales that exhibit higher rates of the endorsement of implicatures, as predicted by the Salience model.

In order to achieve this, we conducted a web-based lexical decision experiment with the adjectival scales studied in the research of Gotzner et al. (2018), which we report below. In our analysis, we take their semantic annotations as well as recorded implicature endorsement rates and combine these with our collected response time data.

Hypotheses and predictions

We firstly predict a general priming effect to occur, given previous results. Next, we hypothesise that the semantic features will have the following effect on priming strength via our link to the Salience model: (1) Adjectives on bounded scales are primed to a larger degree than those on unbounded ones; (2) extreme strong terms are primed to a smaller degree than non-extreme ones; (3) strong absolute adjectives are more strongly primed than strong relative ones; (4) semantic distance should affect the amount of priming (if the direction of this effect is positive, this indicates that strong terms are more strongly activated). As for the relationship between the strength of priming and inference rates, we predict the following: The strength of priming of the strong term is positively correlated with the probability of a scalar implicature being drawn.

The priming experiment

In the current section, we report a web-based lexical decision experiment that used the rapid serial visual presentation method to expose native English participants to simple sentences containing weak scalar terms (or unrelated words) followed by strong scalar probes of the equivalent scales. This experiment was run in order to gain the data for the correlation analysis testing whether the strength of priming depends on the semantic features of different scales. It was pre-registered on OSF (linked here).

Method

Participants On Prolific, we recruited 150 native English speakers based in the US. The average age of the participants

was 28 (SD = 4.57), while the gender distribution was: 85 women, 58 men, 6 people of diverse gender and 1 person who chose not to answer the gender question. The experiment took approximately 13 minutes. The participants were paid £1.95 as compensation for their time.

Materials We constructed 64 items, each of which had two variants. We took the weak and strong adjectives researched in the study of Gotzner et al. (2018) as our starting point. The weak scale-mates were used within sentential frames to form the RELATED condition. To construct the UNRELATED conditions, we used the words from the study of Ronai & Xiang (2023). Additional unrelated words were selected to fit the other adjectives that were not tested in Ronai & Xiang (2023). We collected latent-semantic analysis scores (Landauer & Dumais, 1997) for the similarity between the target words (strong scalars) and the UNRELATED items, making sure that our UNRELATED words were as little related to the targets as possible. We used these unrelated primes as our control condition to provide a floor for the baseline accessibility of our target words in an unsupporting context. We also made sure that the related and unrelated primes were matched in letter-length and log-frequency (see our project entry on OSF for more information: https://osf.io/dbzpq/). The target word was always the strong scalar term to the related condition (here *hot*). Take the following example:

(4) It is warm/lucky. [RELATED/UNRELATED]

Additionally, we created 64 filler items, which had the same sentence frames as the experimental ones, but their associated probes were English non-words. Therefore, the overall experimental item to filler ratio was 1:1. This also meant that the correct answer ratio was the same, since all our experimental items contained existent words for probes and all the fillers non-words.

Procedure The experiment was run on the PCIbex platform (Zehr & Schwarz, 2018). The participants were told they would read sentences that some person might utter and to focus on their meaning. The sentences were presented using the RSVP method (Potter, 2018); in this paradigm, words are presented individually one after another in the centre of a computer screen for a fixed amount of time. Words were presented individually for 350ms each. Following each sentence, 650ms elapsed before the prime word was shown in uppercase and in red lettering. The task of the participants was to indicate whether the sequence shown was a word of English or not. The *J* key was assigned to *YES* and the *F* key to *NO*. The time-out limit for responding was 3000ms. After seven practice items, the experiment itself started. Participants saw a variant of all the 64 critical items determined by the Latin Square design and every 64 filler in a randomised order.

Analysis First, we excluded the data from 20 participants for failing to reach the accuracy threshold of 90% on the combined set of experimental and filler items. We only entered the trials with correct responses into the analyses. Next, we excluded all trials where the recorded response time was under 150ms or larger than $SD = 2.5$. We also excluded the data from two items (*happy/ecstatic* and *pretty/beautiful*), since these were erroneously given different strong terms compared to the study of Gotzner et al. (2018), making their further analysis invalid.

Next, we ran a linear mixed-effects model using the *lm4* and *lmerTest* packages (Kuznetsova et al., 2017) in R (R Core Team, 2022) on log-RTs. This was our BASE MODEL. We entered RELATEDNESS as the sole fixed factor. RELATED trials were coded as 1 and UNRELATED ones as 0. This means that any effect of priming, that is faster RTs to targets following RELATED primes, would be indicated by *negative* slopes. We included the full random effects structure Barr et al. (2013) with intercepts for participants and items as well as random slopes for RELATEDNESS.

Following this, we extracted the random effects structure from the BASE MODEL in order to further analyse the slope terms for the condition of RELATEDNESS for items. Using the slope terms allowed us to examine the variability between items in terms of the effect of RELATEDNESS on them. We chose not to use raw RTs, since the slopes are arguably more representative of the variation of the strength of the effect within our sample of adjectives from a centred baseline. Having made the slope term as the dependent variable, we ran a linear model (the DIVERSITY MODEL) which included the semantic structure variables obtained from the annotations reported in Gotzner et al. (2018). We used these annotations for the sake of comparability. These predictors were BOUNDEDNESS, EXTREMENESS, STANDARD TYPE and SEMANTIC DISTANCE. The way we conducted hypothesis testing was by means of model comparison. We gradually added predictors in the order above and ran a sequence of models. These were then compared using the *anova()* function in R to see whether adding these predictors significantly increased model fit.

Finally, we ran the INFERENCE MODEL, in which we tested whether the strength of priming predicted the rate at which participants endorsed the strengthened meaning as reported in Gotzner et al. (2018). We constructed a linear model with endorsement rates as the dependent variable and the slope terms for the condition of RELATEDNESS extracted from the BASE MODEL as the independent variable. We also calculated the Pearson correlation value for the relationship between the two variables.

Results

Our BASE MODEL revealed a significant main effect of RELATEDNESS ($\beta = -0.038$, $SE = 0.006$, $df = 64.293$, $t = -6.838$, $p < 0.001$). As for the random effects values for items that are of further interest, the intercept had the variance of 0.007 ($SD = 0.086$), random slope for RELATEDNESS showed 0.001 for variance with the standard deviation of 0.025. The correlation term between the intercept and the random slope was $r = 0.17$. The results mean that the strong scalar target words were recognised faster when the prime in

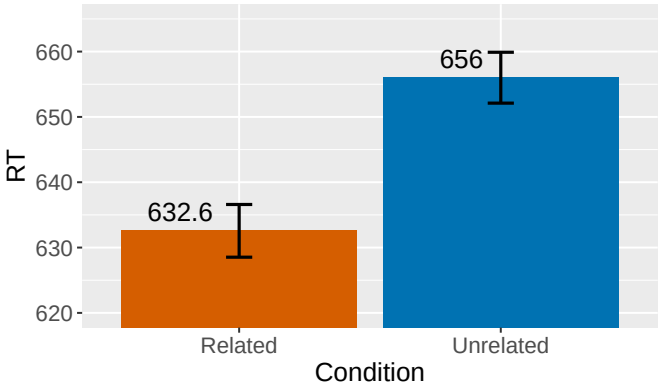

Figure 1: Mean RTs per condition with associated SEs.

the preceding sentence was a weak scale-mate compared to an unrelated word. This means that in the aggregate, weak scale-mates activated their stronger counterparts during comprehension. The reader may consult a graphical representation of these results in Figure 1.

In the DIVERSITY MODEL, the ANOVA test comparing the model fit of added predictors revealed that only the addition of the BOUNDEDNESS variable significantly approved model fit. None of our other variables (EXTREMENESS, STANDARD TYPE, and SEMANTIC DISTANCE) were significant predictors of priming rates. The values for the factor of BOUNDEDNESS were: $\beta = 0.011$, $SE = 0.004$, $df = 60$, $t = 2.629$, $p = 0.011$. In Figure 2, we show the variability found within the items. We draw the attention of the reader to the fact that *negative* values indicate a larger priming effect.

The INFERENCE MODEL revealed a significant effect of priming predicting the inference rates obtained from Gotzner et al. (2018). The resulting values were the following: $\beta = 2.500$, $SE = 1.035$, $df = 60$, $t = 2.415$, $p = 0.019$. The calculated correlation between the two variables was $r = 0.298$. Note that the more negative the slope is, the stronger the priming effect is. This means that a positive effect in the model suggests that the less priming a particular scale has, the higher its inference rate is. The reader may consult a graphical representation of the correlation between them with the slope of the INFERENCE MODEL included in Figure 3. The effect remained significant when the two most extreme outliers were removed (items 8 *calm/unflappable* and 55 *thin/skinny*), see the OSF entry for more details.

Discussion

In the current study, we aimed to test whether the semantic structure of scalar adjectives influences their priming in contexts which could give rise to scalar implicatures. Specifically, we were interested in whether boundedness, extremeness, standard type, and semantic distance plays a role in modulating the activation of the strong scale mate by the weak counterpart present in the stimuli sentence. We also tested whether priming strength predicted comprehenders'

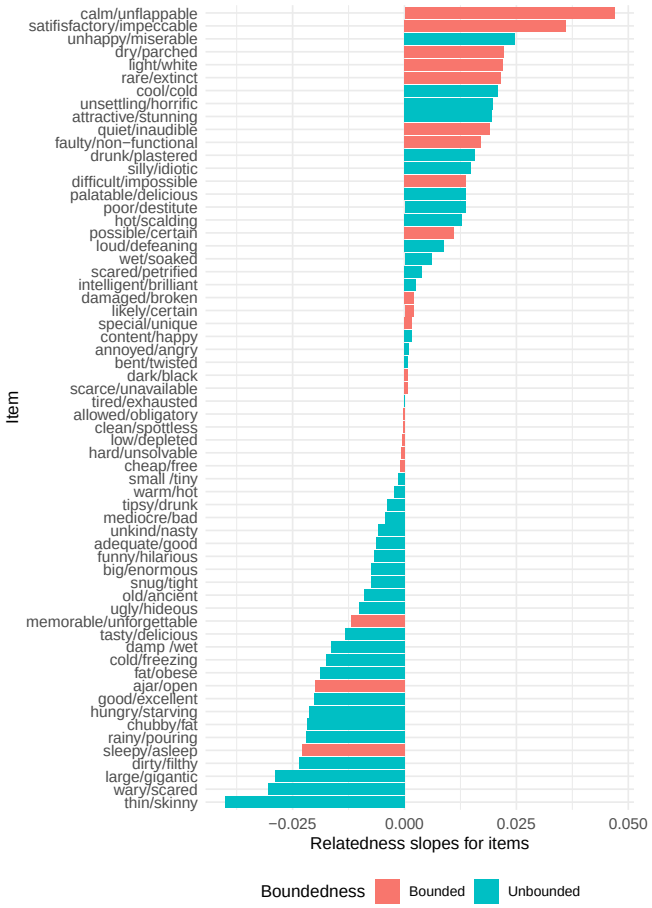

Figure 2: The random slopes for items for the factor of RELATEDNESS extracted from the BASE MODEL.

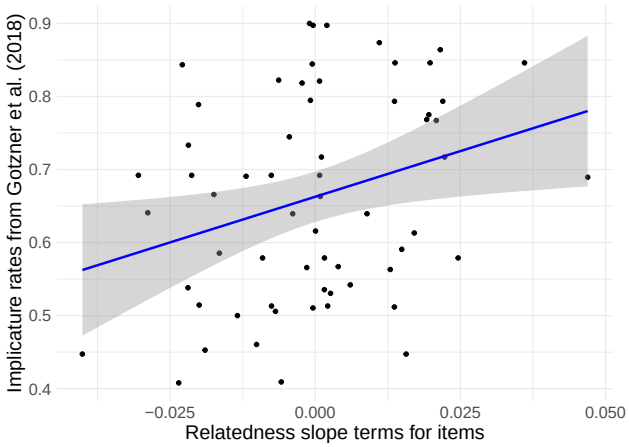

Figure 3: Relatedness slope terms for adjective pairs and their implicature rates from Gotzner et al. (2018), and the associated regression line. For the slopes, positive values indicate a *weaker* priming effect than average, while negative ones a *stronger* effect.

rates of endorsing implicatures.

Firstly, the overall results of our experiment as analysed by the condition of RELATEDNESS successfully replicated the earlier findings of Ronai & Xiang (2023) in that the activation of the strong term (*hot*) was higher when the preceding sentence included the corresponding weaker scale-mate (*warm*) compared to an unrelated word. Out of all the semantic features that we tested, it was only boundedness that was found to be linked to the strength of priming of strong scalar adjectives. However, the effect was in the opposite direction than predicted—bounded scales exhibited weaker priming effects compared to those that were not bounded. Extremeness, standard type, and semantic distance were found to have no effects on whether the strong scale in question was more or less primed by its weaker scale-mate.

Our study reveals that there is a relationship between the activation of the strong term and implicature endorsement rates. Yet again, however, the effect is in the opposite direction. It was those scales that were primed to a lesser extent or even exhibited interference that were associated with higher implicature endorsement rates. Firstly, this contrasts with the results of Ronai & Xiang (2023), who found an insignificant correlation of $r = 0.004$ between their priming data and inference rates, whereas the correlation in our study was significant and its value was $r = 0.298$, which albeit weak, is a markedly higher one suggesting that there is indeed a link. Our data, therefore, provide the first evidence linking an online processing measure with an offline measure of the eventual product of pragmatic comprehension. This finding is crucial for the uniformity assumption as well as for the Salience model of implicature processing proposed by Rees & Bott (2018). *Lower* levels of activation were associated with higher implicature endorsement rates. This goes directly against the model's predictions, which were expecting the effect in the opposite direction. This suggests that the simple account linking increased availability of the strong term with the probability of deriving an implicature is inadequate.

Our data suggest that the long-standing assumption that comprehenders need to access and negate the stronger alternative, conceptualised as a lexical item (Horn, 1972), in order to derive an implicature might be false. Rather, it might be that comprehenders' knowledge of the scale structure alone, for example its boundedness, is sufficient for the derivation. In other words, it is enough to know where the lower bound of an expressions lies on the relevant degree scale and whether that scale has an end-point (without reasoning about the lexical item corresponding to that end-point). This approach has recently been advanced as the Measurement Mechanism model by Gotzner & Lacina (2023), which suggests that comprehenders reason about intervals on an underlying meaning scale rather than about lexical alternatives[2].

Furthermore, our findings challenge the uniformity assumption, since a link between availability, inference rates and scale structure was found, suggesting that bounded and unbounded scales might have different derivations of scalar implicatures. One way to understand the differences between bounded and unbounded scales is via the distinct role of comparison classes in the meaning of these expressions. Unbounded scales need a comparison class to resolve their basic meaning, for example tall basketball players are taller than the average man Kennedy & McNally (2005). The meaning of bounded scales does not vary contextually in the same way. This distinction also plays a role in scalar implicature computation. In an incremental decision task, Alexandropoulou et al. (2022) found evidence that relative adjectives only trigger scalar implicatures when the context instantiates a relevant comparison class. Minimum standard adjectives in turn invoke a lower bound via their basic meaning and they triggered scalar implicatures independently of the context.

There are other possible explanations for our finding of the negative impact of priming on inference rates. It might be that what our experiment is tapping into is a process that occurs after the Activation threshold was crossed and further processing is already under way. Since we probed our targets at 650ms after the prime, this is plausible (for comparison see Husband & Ferreira, 2016). At this point, the strong term may already be negated, which could cause its suppression, consistent with the findings on the role of polarity and negation in the slow-downs associated with scalar implicature derivation (Bott & Noveck, 2004; Van Tiel & Pankratz, 2021). Future research is needed to measure the role of alternatives during incremental processing of scalar implicatures and at more immediate time points.

Conclusion

This research explored the phenomenon of scalar diversity in the derivation of implicatures through the lens of priming during online sentence comprehension. We tested the Salience model and its underlying uniformity assumption by means of linking online priming measures and offline implicature rates. We asked two questions, firstly, whether the strength of activation of the strong term caused by the presence of the corresponding weaker term as the prime could be predicted by the semantic features of the scale. Secondly, we asked whether priming strength predicted implicature endorsement rates. In a web-based lexical decision experiment, we found that only the boundedness of scales predicted priming strength. Contrary to our predictions, we found that unbounded scales exhibited stronger priming effects. Priming strength predicted the endorsement rates, but also in the opposite direction—it was the less primed scales that showed higher rates of comprehenders endorsing the strengthened meaning. We interpret these results as going against the predictions of the Salience model, challenging the uniformity assumption and suggesting that as opposed to the common assumption in the literature, implicatures may not always need the lexical stronger alternative available for their derivation.

[2]See Buccola et al. (2022) for another proposal for scalar implicatures in general that does away with lexical alternatives.

Acknowledgements

We would like to thank our research assistants Berit Reise and Maria Stella Villa Avila for their help with stimuli creation and the experimental script. Both authors were supported by the German Research Foundation (Deutsche Forschungsgemeinschaft) through an Emmy-Noether grant awarded to the second author (Nr. GO 3378/1-1).

References

Alexandropoulou, S., Herb, M., Discher, H., & Gotzner, N. (2022). Incremental pragmatic interpretation of gradable adjectives: The role of standards of comparison. In *Semantics and linguistic theory* (pp. 481–497).

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Beltrama, A., & Xiang, M. (2013). Is 'good' better than 'excellent'? an experimental investigation on scalar implicatures and gradable adjectives. In *Proceedings of sinn und bedeutung* (Vol. 17, pp. 81–98).

Bott, L., & Frisson, S. (2022). Salient alternatives facilitate implicatures. *Plos one*, *17*(3), e0265781.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*(3), 437–457.

Buccola, B., Križ, M., & Chemla, E. (2022). Conceptual alternatives: Competition in language and beyond. *Linguistics and Philosophy*, *45*(2), 265–291.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In *Structures and beyond* (Vol. 3, pp. 39–103). Oxford, UK.

De Carvalho, A., Reboul, A. C., Van der Henst, J.-B., Cheylus, A., & Nazir, T. (2016). Scalar implicatures: The psychological reality of scales. *Frontiers in psychology*, *7*, 1500.

Degen, J. (2015). Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*, 11–1.

Doran, R., Baker, R. E., McNabb, Y., Larson, M., & Ward, G. (2009). On the non-unified nature of scalar implicature: an empirical investigation. *International Review of Pragmatics*, *1*, 1-38.

Doran, R., Ward, G., Larson, M., McNabb, Y., & Baker, R. E. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language*, 124–154.

Fox, D., & Katzir, R. (2011). On the characterization of alternatives. *Natural language semantics*, *19*(1), 87–107.

Gotzner, N. (2019). The role of focus intonation in implicature computation: a comparison with only and also. *Natural Language Semantics*, *27*(3), 189–226.

Gotzner, N., & Lacina, R. (2023). Generating and selecting alternatives for scalar implicature computation: The Alternative Activation Account and other theories. *ResearchGate*. doi: 10.13140/RG.2.2.30246.09282

Gotzner, N., Solt, S., & Benz, A. (2018). Scalar diversity, negative strengthening, and adjectival semantics. *Frontiers in psychology*, *9*, 1659.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55.

Horn, L. R. (1972). *On the semantic properties of logical operators in english*. University of California, Los Angeles.

Hu, J., Levy, R., & Schuster, S. (2022). Predicting scalar diversity with context-driven uncertainty over alternatives. In *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 68–74).

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, *58*(3), 376–415.

Husband, E. M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience*, *31*(2), 217–235.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy*, *30*(1), 1–45.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 345–381.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmertest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1–26.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211.

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Morzycki, M. (2012). Adjectival extremeness: Degree modification and contextually restricted scales. *Natural Language & Linguistic Theory*, 567–609.

Potter, M. C. (2018). Rapid serial visual presentation (rsvp): A method for studying language processing. In *New methods in reading comprehension research* (pp. 91–118). Routledge.

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rees, A., & Bott, L. (2018). The role of alternative salience in the derivation of scalar implicatures. *Cognition*, *176*, 1–14.

Ronai, E., & Xiang, M. (2021a). Exploring the connection between question under discussion and scalar diversity. *Proceedings of the Linguistic Society of America*, *6*(1), 649–662.

Ronai, E., & Xiang, M. (2021b). Pragmatic inferences are qud-sensitive: an experimental study. *Journal of Linguistics*, *57*(4), 841–870.

Ronai, E., & Xiang, M. (2022). Three factors in explaining scalar diversity. In *Proceedings of sinn und bedeutung* (Vol. 26, pp. 716–733).

Ronai, E., & Xiang, M. (2023). Tracking the activation of scalar alternatives with semantic priming. In (Vol. 2, pp. 229–240). (https://doi.org/10.3765/elm.2.5371)

Sun, C., Tian, Y., & Breheny, R. (2018). A link between local enrichment and scalar diversity. *Frontiers in Psychology*, *9*, 2092.

Sun, C., Tian, Y., & Breheny, R. (2023). A corpus-based examination of scalar diversity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Van Tiel, B., & Pankratz, E. (2021). Adjectival polarity and the processing of scalar inferences. *Glossa: a journal of general linguistics*, *6*(1).

van Tiel, B., Pankratz, E., & Sun, C. (2019). Scales and scalarity: Processing scalar inferences. *Journal of Memory and Language*, *105*, 93–107.

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of semantics*, *33*(1), 137–175.

Zehr, J., & Schwarz, F. (2018). *PennController for internet based experiments (IBEX).* (https://doi.org/10.17605/OSF.IO/MD832)