# On Being Certain that Presuppositions don't Project Universally

Maik Thalmann
University of Göttingen
Göttingen, Germany
maik.thalmann@gmail.com

Andrea Matticchio
UMass Amherst
Amherst MA, USA
amatticchio@umass.edu

**Abstract**

In this paper, we present new experimental evidence that presupposition projection out of quantified environments can be modeled in a framework that uses a trivalent logic like Strong Kleene. Assuming a quantificational analysis, our experiment focuses on projection out of attitude reports, which are understudied from the empirical point of view. Theories that assume generalized universal projection, which are viewed as standard, cannot capture our data, which are instead compatible with the idea that presupposition failures can take part in the meaning computation. In the data analysis, we show how modeling standard deviations using a distributional model allows for new insights by enabling the detection of the squeamishness characteristic of presupposition failure in participants' response behavior.

## 1 Introduction

The projection problem for presuppositions is the problem of predicting what presupposition a complex sentence will impose on the context given the presupposition triggers in its constituents. A long line of research has addressed this problem as part of compositional semantics. Descriptively, logical operators have traditionally been associated with specific behaviors with respect to presupposition projection, stipulated as part of the information that comes with the lexical meaning of the operator (as in the classification of *holes*, *plugs*, and *filters* in Karttunen 1973).

Instead, we here pursue the hypothesis that the projection behavior of an operator does not have to be stipulated, but follows from the logical contribution of the operator and a logic dealing with truth-value gaps. We hypothesize that this logic is Strong Kleene for all domains of quantification. In this paper, we focus on presupposition projection from the scope of quantifiers because there is, as yet, no consensus about what empirical picture a theory of projection should derive for quantifiers over individuals. To see this, consider the variants of the sentence in (1), where the argument of the presupposition trigger *stop* contains a variable bound by a quantifier.

(1) All/Some/None of the students $\lambda x$ [ $x$ stopped smoking. ]

At first glance, it seems reasonable to assume that the sentence in (1) with *all* presupposes that all of the students used to smoke, a universal presupposition. However, for existentially quantified presuppositions, as in (1) with *some*, Karttunen and Peters (1979) already suggest that a universal presupposition is too strong. A first proposal for these cases was fully spelled out by Heim (1983): for all the sentences above, the presupposition should carry universal quantificational force, but additionally, different strategies can apply and weaken the presupposition.

Chierchia (1995) suggests that the situation is more complex and that universal and existential presuppositions can follow from different logical forms of the same sentence. Later, Sudo et al. (2012) report a large degree of inter-speaker variation. Finally, work by Fox (2013) (building on George 2010) argues that a system based on the trivalent Strong Kleene logic allows for an adequate model of presupposition projection for (1) that is in line with the experimental results in Chemla (2009). Again, the predictions are qualified by the availability of mechanisms that can weaken (or sometimes strengthen) presuppositions—for example, *local accommodation*.

(2) Given a domain of quantification $D$, the truth value of a formula $\forall x\ \phi(x)$ is:
   a. `T` if $\phi(x) =$ `T` for all $x \in D$;
   b. `F` if there is an $x \in D$ such that $\phi(x) =$ `F`;
   c. `#` otherwise (if there is an $x \in D$ such that $\phi(x) =$ `#` and for no $x \in D$, $\phi(x) =$ `F`).

In (2), we report how truth values are assigned to a universally quantified formula according to the Strong Kleene logic, `#` being *neither true nor false*. Notably, importing Strong Kleene into natural language will not exactly predict presupposition projection, but rather trivalent truth conditions for a sentence.[1] The result is that for sentences with quantifiers, different requirements on the context might be predicted when the sentence is asserted or negated.

A proposal by Charlow (2009) further complicates the picture. According to his intuitions, *soft* and *hard* presupposition triggers (Abusch 2010; Abrusán 2016) lead to different projection patterns: the presuppositions of hard triggers (like *regret*, *too*, and *again*), unlike those of soft triggers (like *discover*, *win*, and *stop*), always project universally out of quantifiers. We will consider this potential difference between soft and hard triggers in our experimental design.

While most of the research on projection out of quantifiers has concentrated on quantifiers over individuals, projection out of modal domains has been at least partially neglected, to our knowledge. In remedy, here we focus on attitude reports. Seminal work by Heim (1992), grounded in dynamic semantics, derives that attitude predicates like *think* behave as plugs.

(3) Peter thinks that Jan stopped smoking.
   *Presupposes:* Peter thinks that Jan used to smoke. *(according to Heim 1992)*

Apart from approaches assuming that projection out of attitudes has a different profile altogether (as in Geurts 1999), the desideratum in (3) has not been questioned. More recent frameworks, like those by Schlenker (2009) or Mandelkern (2024), are still designed to meet this intuition. Instead, a very natural extension of the Strong Kleene approach to modality would predict that attitude predicates, or at least those with a quantificational semantics, behave just like quantifiers over individuals (translating (2) to a modal domain) for presupposition projection.

Before moving on to the predictions of this hypothesis, it is important to note that our prototypical attitude predicate *think* is a neg-raiser: this means that the falsity conditions of a *think* report might not be the Boolean complement of its truth conditions. For this reason, we will use the predicate *be certain*, which is not a neg-raiser and can give rise to indirect scalar implicatures like a universal quantifier when embedded under negation, as shown in (4).

(4) Peter is not certain that Jan smokes.
   $\not\rightsquigarrow$ Peter is certain that Jan doesn't smoke.
   $\rightsquigarrow$ Peter considers it possible that Jan smokes.

Assuming for the non-quantificational case that a sentence like *Jan canoed again* presupposes that *Jan canoed last time*, we give, in (5), the predictions of Strong Kleene for a *certain*-report embedding the presupposition trigger *again*. Being different from the predictions of a generalized Universal Projection hypothesis, in (6), this yields the starting point for our experiment.

(5) Peter is certain that Jan canoed again. *(Strong Kleene)*
   a. `T` if Peter is certain that Jan canoed last time and this time.
   b. `F` if Peter considers it possible that Jan canoed last time and didn't this time.
   c. `#` if Peter considers it possible that Jan didn't canoe last time, but he is certain that if Jan canoed last time, he canoed this time, too.

[1] Fox (2013) then adopts *Stalnaker's Bridge Principle* to conclude that the presupposition of a quantified formula is a disjunctive proposition; in our case, the negation of (2-c).

(6) Peter is certain that Jan canoed again. *(Universal Projection)*
   a. `T` if Peter is certain that Jan canoed last time and this time.
   b. `F` if Peter is certain that Jan canoed last time and considers it possible he didn't this time.
   c. `#` if Peter considers it possible that Jan didn't canoe last time.

## 2 Experiment: Quantifiers over worlds

### 2.1 Design and Materials

The experiment was set up as in (7) and carried out via PCIbex (Zehr and Schwarz 2022), with the `true`, `false`, and `undefined` scenarios serving as controls since for those contexts, both Strong Kleene and the Universal Projection hypothesis make the same predictions.

(7) $2 \times 4$ design (within-within; 48 items; 34 participants; $\overline{\text{age}} = 23.8$, SD $= 6.2$)
   a. NEGATION: without vs. with matrix negation
   b. SCENARIO: `true` vs. `false` vs. `undefined` vs. `critical`
   c. Presupposition TRIGGER as a pseudo-factor: *again* vs. *stop* (betw. items, 24 each)

Participants were asked to judge the truth of an attitude report in a context (the SCENARIO manipulation) using the scale in (8); this resulted in a continuous trivalent truth-value judgment task (see Križ and Chemla 2015 for the non-continuous version of this experimental task).

(8) (completely false) (neither completely true nor completely false) (completely true)

The items were structured as follows: a character expresses their doxastic state about a past and a present situation; in an (informationally redundant) image presented alongside the utterance, past and present situations are represented by pictures connected by an arrow. The partitioned cartoon bubble represents what the character considers possible. Before starting, participants were familiarized with the characters and the meaning of the arrows and the clouds. In (9), we list a sample item (translated from German) with *again*; the images are shown in Figure 1. For space reasons, we must leave out a detailed presentation of the *stop* items, though we show the images for one in Figure 1. The inclusion of the soft trigger *stop*, alongside the hard trigger *again*, allows us to see whether speakers differentiate between trigger types (cf. Charlow 2009).

(9) Peter is (not) certain that Jan canoed again.
   `true` "I'm certain Jan canoed last time, and I'm certain Jan canoed this time."
   `false` "I'm certain Jan canoed last time, but I have no idea if this time, Jan canoed or not."
   `undefined` "I have no idea if last time, Jan canoed or not, but I am certain that Jan canoed this time."
   `critical` "I have no idea if last time, Jan canoed or not, and I have no idea if this time, Jan canoed or not."

In the `critical` condition, the doxastic state of the character allows the possibility that the presupposition is not met, and also that the presupposition is met and the embedded proposition is false. If presuppositions project universally, as in Heim (1992), the possibility that the presupposition is not met should make the sentence undefined. If, on the other hand, truth values are as in (5), the possibility that the embedded proposition is false should make the sentence false, and its negation should be true. The predictions are summarized in Figure 2.

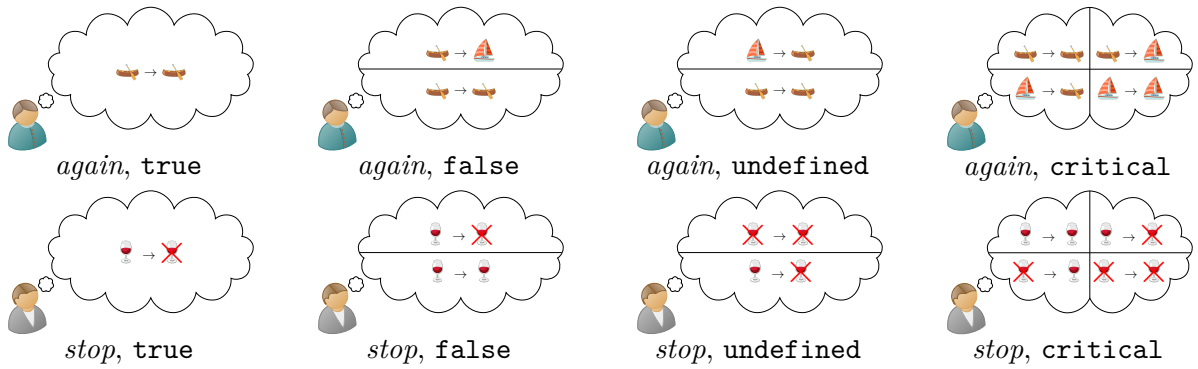


Figure 1: Images redundantly encoding the SCENARIO manipulation for the translated reports "Peter is (not) certain that Jan canoed again" (*top*) and "Markus is (not) certain that Sonja stopped drinking wine" (*bottom*).

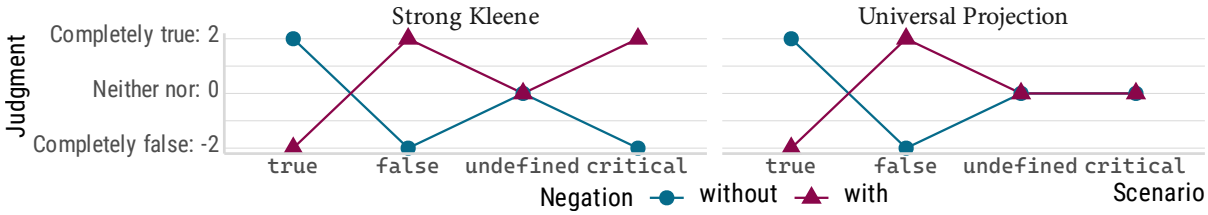


Figure 2: Predictions about presupposition projection from *be certain* for the two hypotheses for each SCENARIO.

This design crucially relies on the NEGATION factor to detect judgments of intermediate truth: if a sentence is neither true nor false, adding matrix negation should not affect the judgment.

We noted in (4) that *be certain* under negation gives rise to an indirect scalar implicature. To ensure that this inference does not affect our results, the character's doxastic state in every item allows for the possibility that the embedded proposition is defined and true. This way, no infelicity can arise because a scalar implicature is violated.

## 2.2 Results

We excluded 4.35% of the data due to the submission times. First, all observations where submission times exceeded 60 s, and then, all observations whose submission times exceeded the new mean ±3 standard deviations. The descriptive results after exclusion are shown in Figure 3.

We fit a Bayesian mixed distributed linear regression (using `brms` from Bürkner 2021 in R 4.4.1, R Core Team 2024) with 8 chains of 10,000 iterations each (5000 for warm-up). As is evident from the rating distribution in Figure 3, our data involve unequal variances between conditions: especially the `undefined` scenario shows large variances, while the `true` and `false`

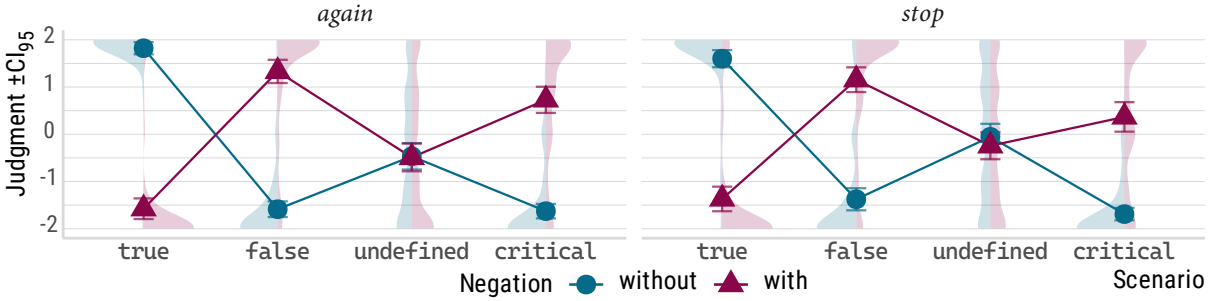


Figure 3: Results. The shaded areas indicate the raw ratings; the confidence intervals involve no by-participant clustering. The lower end of the scale ('completely false') was coded −2, the upper end ('completely true') as 2.

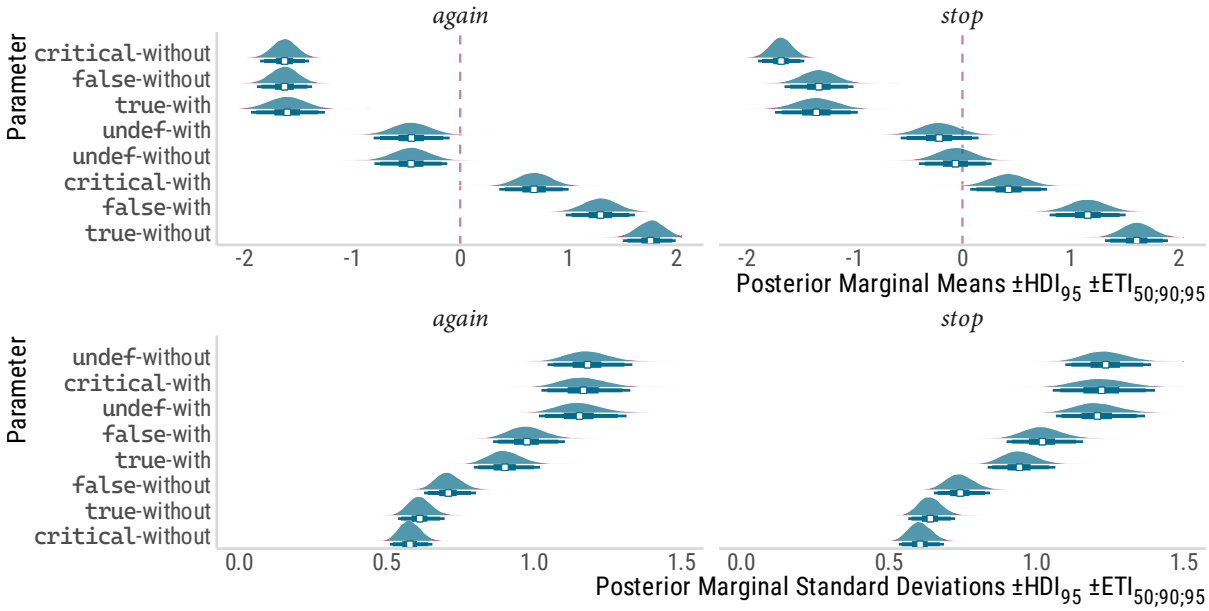


Figure 4: Posterior marginal effects for the location (*top*) and the scale parameters (*bottom*) in the model.

scenarios display much clearer rating clusters. For this reason, we estimated not just the location, as is standard, but also the scale parameter (the standard deviation) for each predictor, resulting in what is called a distributional model. This allows us to capture the unequal variances instead of violating the homogeneity of variance assumption (see, e.g., Kruschke 2015, Ch. 19). In (10), we show the formula for each sub-model, with the maximal random effect structure for the means (Barr et al. 2013). The standard deviations, being strictly positive, are modeled on the log scale, though we will present the back-transformed standard deviations on the response scale.

(10) a. $Y \sim \text{NEGATION} * \text{SCENARIO} * \text{TRIGGER} + (1 + \text{SCENARIO} * \text{TRIGGER} * \text{NEGATION} \mid \text{PARTICIPANT}) + (1 + \text{SCENARIO} * \text{NEGATION} \mid \text{ITEM})$

b. $\log \sigma \sim \text{NEGATION} * \text{SCENARIO} + \text{TRIGGER}$

The priors for model parameters (except those for correlations) were $N(0, x)$, with $x$ ranging between .25 (for the log standard deviation) and 1 (for the location slopes). For the intercept and the slopes in the location sub-model, we also added lower and upper bounds to the priors to reflect the bounded dependent variable at $-2$ ('completely false') and 2 ('completely true').

The posterior marginal effects, which we use to show posterior predictions rather than relying on coefficients, were computed on the posterior draws using the `emmeans` package (Lenth 2024) and are shown in Figure 4 for both the means and the (back-transformed) standard deviations.

The controls all behave as expected, with the scenarios `true` and `false` showing an effect of negation in the expected direction, and `undefined` being insensitive to the polarity manipulation. As for the `critical` scenario, we see a clear effect of matrix negation and an overall similarity with the `false` scenario, suggesting that participants do not judge the items in these scenarios as involving presupposition failures. These results are supported by a Bayes Factor analysis using Regions of Practical Significance (ROPE) in Figure 5 (for an introduction, see Kruschke 2018). If the prior and posterior odds are the same (and there is no evidence for either difference or equivalence), then the (log) Bayes Factor is 0. If the posterior odds are higher than the prior odds (and the data thus support a difference hypothesis), the Bayes Factor will be larger than 0.

Interestingly, we also find stark contrasts with the standard deviations, Figure 4. Here, `true`

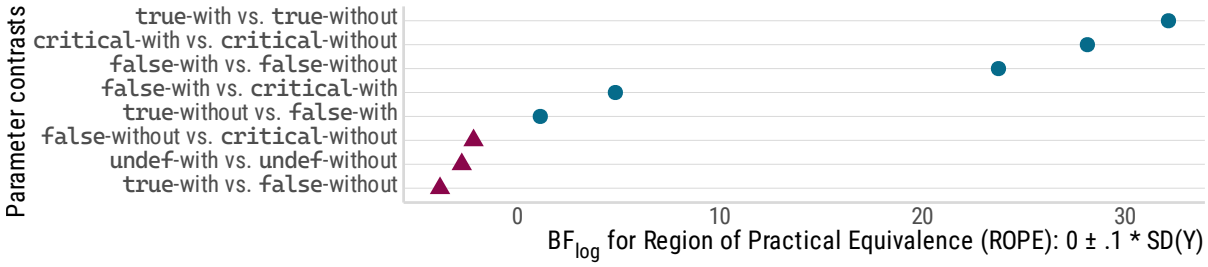


Figure 5: Bayes Factors. Comparisons where equivalence hypotheses are supported shown in red, others in blue.

and `false` scenarios show relatively small dispersion (with an increase due to negation), while the `undefined` scenario shows large estimates irrespective of negation. We take this to index presupposition failure-induced 'squeamishness' (Strawson 1964). By contrast, the `critical` scenario patterns with `true` and `false` without negation, though negation impacts dispersion.

## 3 Discussion and conclusion

In accordance with a presuppositional view of *stop* and *again*, the `undefined` scenario does not show an effect of negation. In the `critical` scenario, negation has an effect, suggesting no presupposition failure. This is in line with the generalized Strong Kleene hypothesis, but clashes with the Universal Projection hypothesis. Additionally, there is no evidence that the two types of triggers show different projection patterns (contra Charlow 2009). More generally, we take our results to support a view where presupposition failures can take part in the meaning composition, and presupposition projection is determined by the contribution of each logical operator—including attitude predicates—provided a suitable trivalent logic like Strong Kleene.

To our knowledge, we provided the first application of distributional models to detect intuitions of presupposition failure via standard deviations. In our data, presupposition failures increased standard deviations more than negation, thus modeling the well-known squeamish reaction (Strawson 1964; von Fintel 2004). This result opens up new analytical avenues for experiments on truth-value gaps, even when the dependent variable does not target truth values.

There is an unexpected difference between the `critical` and `false` scenarios with negation, both in terms of means and standard deviations. In speculation about this pattern, we offer two possible explanations. It is a fact that the `critical` scenario with negation is the most complex condition of the design: it involves the highest number of partitions in the character's doxastic state and requires extra computational cost because of negation. It is possible that judgments become less sharp and variance increases as a result of participants' uncertainty or cognitive load. Alternatively, it might be the case that the participants, who are fully informed about the character's doxastic state, judge the negated report in the `critical` condition as underinformative, even if true. Consider the sentence in (11) together with its scalar implicature:

(11) Peter is not certain that Jan canoed again.
   a. *Entails:* Peter considers it possible that Jan didn't canoe again.
   b. *Implicates:* Peter considers it possible that Jan canoed again.

The entailment and the implicature of the sentence address only two out of four of Peter's doxastic possibilities in the `critical` scenario. Instead, these two doxastic possibilities are all that features in the `false` scenario, where the sentence is judged as true with little uncertainty.[2]

[2]Clemens Mayr (p.c.) suggests that embedded polar interrogatives, which are felicitous with negated *be certain* (Mayr 2019), might compete with our items. A trivalent theory of embedded questions is needed to explore this.

**Acknowledgements.** This project was carried out within the RTG 2636 "Form-Meaning Mismatches," funded by the *Deutsche Forschungsgemeinschaft*, project number: 429844083.

## References

Abrusán, Márta (2016). "Presupposition cancellation: Explaining the 'soft–hard' trigger distinction". In: *Natural Language Semantics* 24.2, pp. 165–202. DOI: 10.1007/s11050-016-9122-7.

Abusch, Dorit (2010). "Presupposition triggering from alternatives". In: *Journal of Semantics* 27.1, pp. 37–80. DOI: 10.1093/jos/ffp009.

Barr, Dale J. et al. (2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal". In: *Journal of Memory and Language* 68.3, pp. 255–278. DOI: 10.1016/j.jml.2012.11.001.

Bürkner, Paul-Christian (2021). "Bayesian Item Response Modeling in R with brms and Stan". In: *Journal of Statistical Software* 100.5, pp. 1–54. DOI: 10.18637/jss.v100.i05.

Charlow, Simon (2009). "'Strong' predicative presuppositional objects". In: *Proceedings of ESSLLI 2009*.

Chemla, Emmanuel (2009). "Presuppositions of quantified sentences: Experimental data". In: *Natural Language Semantics* 17.4, pp. 299–340. DOI: 10.1007/s11050-009-9043-9.

Chierchia, Gennaro (1995). *Dynamics of meaning: Anaphora, presupposition, and the theory of grammar*. Chicago: University of Chicago Press.

von Fintel, Kai (2004). "Would you believe it? The King of France is back! (Presuppositions and truth-value intuitions)". In: *Descriptions and Beyond*. Ed. by Marga Reimer and Anne Bezuidenhout. Oxford: Oxford University Press, pp. 315–341.

Fox, Danny (2013). "Presupposition projection from quantificational sentences: Trivalence, local accommodation, and presupposition strengthening". In: *From Grammar to Meaning*. Ed. by Ivano Caponigro and Carlo Cecchetto. Cambridge University Press, pp. 201–232. DOI: 10.1017/cbo9781139519328.011.

George, B. (2010). "A new case for an old logic: Reviving Strong Kleene approaches to presupposition projection". Unpublished ms. UCLA.

Geurts, Bart (1999). *Presuppositions and Pronouns*. Amsterdam: Elsevier.

Heim, Irene (1983). "On the projection problem for presuppositions". In: *WCCFL 2: Second Annual West Coast Conference on Formal Linguistics*. Ed. by Michael Barlow, Daniel P. Flickinger, and Michael T. Wescoat, pp. 114–125.

— (1992). "Presupposition projection and the semantics of attitude verbs". In: *Journal of Semantics* 9, pp. 183–221. DOI: 10.1093/jos/9.3.183.

Karttunen, Lauri (1973). "Presuppositions of compound sentences". In: *Linguistic Inquiry* 4.2, pp. 169–193.

Karttunen, Lauri and Stanley Peters (1979). "Conventional Implicature". In: *Presupposition*. Ed. by David A. Dineen and Choon-Kyu Oh. New York: Academic Press, pp. 1–56.

Križ, Manuel and Emmanuel Chemla (2015). "Two methods to find truth-value gaps and their application to the projection problem of homogeneity". In: *Natural Language Semantics* 23.3, pp. 205–248. DOI: 10.1007/s11050-015-9114-z.

Kruschke, John K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Oxford: Academic Press.

— (2018). "Rejecting or Accepting Parameter Values in Bayesian Estimation". In: *Advances in Methods and Practices in Psychological Science* 1.2, pp. 270–280. DOI: 10.1177/2515245918771304.

Lenth, Russell V. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.0. URL: https://CRAN.R-project.org/package=emmeans.

Mandelkern, Matthew (2024). *Bounded meaning: The dynamics of interpretation*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780192870049.001.0001.

Mayr, Clemens (2019). "Triviality and interrogative embedding: context sensitivity, factivity, and neg-raising". In: *Natural Language Semantics* 27.3, pp. 227–278. DOI: 10.1007/s11050-019-09153-8.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Schlenker, Philippe (2009). "Local Contexts". In: *Semantics & Pragmatics* 2. DOI: 10.3765/sp.2.3.

Strawson, P. F. (1964). "Identifying reference and truth-values". In: *Theoria* 30.2, pp. 96–118. DOI: 10.1111/j.1755-2567.1964.tb00404.x. URL: https://doi.org/10.1111%2Fj.1755-2567.1964.tb00404.x.

Sudo, Yasutada et al. (2012). "Presupposition projection out of quantified sentences: Strengthening, local accommodation and inter-speaker variation". In: *Logic, Language and Meaning. 18th Amsterdam Colloquium*. Ed. by Maria Aloni et al. Berlin: Springer, pp. 210–219. DOI: 10.1007/978-3-642-31482-7_22.

Zehr, Jérémy and Florian Schwarz (2022). *PennController for Internet Based Experiments (IBEX)*. DOI: 10.17605/OSF.IO/MD832.