

## The rise and particularly fall of presuppositions: Evidence from duality in universals

Remus Gergel, Maike Puhl, Simon Dampfhofer & Edgar Onea\*

**Abstract.** At the center of this paper is the question whether presuppositions are more likely to be gained or lost in the process of language change. We offer a new experimental method that aims at ascertaining the re-learning speed of potentially presuppositional items based on nonce words and which integrates certain factors of change such as social prestige in an artificial but clearly contextualized set-up. The meaning targeted is of a quantifier meaning ‘both’ with speakers of German and the initial results point to higher ease of losing rather than incorporating the presupposition, but with an interesting resilience after a critical questioning of presuppositional status.

**Keywords.** diachronic semantics; presuppositional quantifiers; reinterpretation times; semantic language processing

**1. Introduction.** In the theoretical literature on language change and historical linguistics, one fundamental question pertains to the ways in which different layers of meaning interact and are transformed, or even vanish. The classical view is, for instance, that implicatures conventionalize and thus become lexical entries, while disappearing as implicatures (see, e.g., Traugott & Dasher 2002, Eckardt 2006 for discussions). Other central aspects of meaning at the semantics-pragmatics interface have, however, been less studied. Crucially for the current enterprise, the question how lexical items come to be presupposition triggers or lose this property has been much less discussed on a principled basis, although case studies on presuppositional items certainly exist (see Beck, Berezovskaya & Pflugfelder 2009, Schwenter & Waltereit 2010, Los & Komen 2012, Beck & Gergel 2015, Gergel, Blümel & Kopf 2016, Gergel & Kopf-Giammanco 2021, Carlier & Lamiroy 2018, among others). Essentially, there are two types of approaches to our central question that have been discussed, even if they have been developed on a very different set of examples. On the one hand, Eckardt (2009) proposes a principle of Avoid Pragmatic Overload which incorporates presuppositions and essentially states that in certain critical situations in which there are too many side-messages (including presuppositions and implicatures) speakers will drop some. This can be understood as presuppositional *loss*. Gergel (2020), on the other hand, focuses on the increased presence of the triggers of presuppositions across different contexts over time and suggests a diachronic corollary of the synchronic principle of Maximize Presuppositions, which comes down to an *increase* as far as the signaling of presuppositional meaning components goes.

However, even if historical linguistics can elucidate the historical path of development for any individual presupposition trigger, the question how and why, in terms of cognitive underpinning, such shifts come about still remains an open and important one. Any theory put forth by historical linguistics on this matter would come with particular cognitive predictions that would need to be borne out generally, irrespective of the peculiarities of the historical processes. Thus, one would

---

\*We thank the audience of ELM2 for their feedback. We also thank Alexandre Cremers for helpful methodological discussion regarding the statistical analysis of our results at a workshop in Vilnius financed by the ARQUS network. Authors: Remus Gergel, Saarland University ([remus.gergel@uni-saarland.de](mailto:remus.gergel@uni-saarland.de)), Maike Puhl, Saarland University ([maike.puhl@uni-saarland.de](mailto:maike.puhl@uni-saarland.de)), Simon Dampfhofer, University of Graz ([simon.dampfhofer@edu.uni-graz.at](mailto:simon.dampfhofer@edu.uni-graz.at)) & Edgar Onea, University of Graz ([edgar.onea-gaspar@uni-graz.at](mailto:edgar.onea-gaspar@uni-graz.at)).

expect to be able to reproduce and test mechanisms of language change in the lab to at least some extent. Such experimental systematic research is non-existent to our knowledge for the classical areas of presupposition triggers; however, some studies that can be considered comparable in the broader sense are available. This line of research falls largely, of course, into the classical Labovian idea of experimentally exploiting present reactions of speakers to uncover processes that can be relevant for language change more generally. And just as naturally, our approach shares more with recent attempts to explain paths of change in the area of meaning (rather than sound change or morphosyntax as in Labovian studies), such as Zhang, Piñango & Deo (2018), Fedzechkina & Roberts (2020), Fuchs, Deo & Piñango (2020), Gergel, Kopf-Giammanco & Puhl (2021), Puhl & Gergel (2022). Returning to the general question regarding presuppositions, our paper addresses it by considering one single case of a presupposition trigger, the lexical item corresponding to the meaning of the quantifier *both*. On the classical view (see Heim & Kratzer 1998 for discussion), this quantifier is similar to a universal quantifier such as *all*, but it has the presuppositional restriction that it can only be used appropriately in cases in which the restrictor set has the cardinality two.

The question we ask is this: how easy is it for speakers to start with the presuppositional meaning of *both* and learn (and adapt to) a new meaning in an experimental scenario that emulates language change in certain ways, corresponding to the meaning of *all*, as compared to starting with the non-presuppositional item *all* and learning and adapting to a new presuppositional meaning in the experimental setup corresponding to the meaning of *both*? Thus, this would give us a direct way to compare the ease at which an item can pick up or lose a presupposition in a language change situation. In what follows, we describe our key experiment (Section 2) together with a relevant replication (Section 3) with their respective methods and results, before returning to a more general discussion in Section 4.

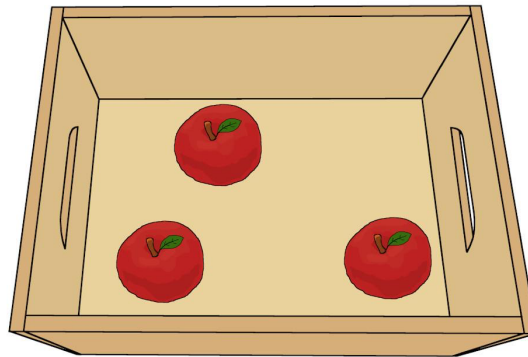
**2. Experiment 1.** Experiment 1 was designed to investigate whether it would be easier to reinterpret BOTH as ALL (notation: BOTH  $\rightarrow$  ALL) or vice versa (ALL  $\rightarrow$  BOTH). To avoid any influence of previous knowledge of the actual words *both* or *all* (i.e., their German variants), participants were taught a nonce word, *gure*, instead, which would mean either BOTH or ALL.

During the experiment, participants were asked to imagine visiting a fictitious community (German speaking diaspora in the US). They were exposed to two types (roughly correlating with generations) of native speakers. Older speakers would use *gure* in its original meaning, while younger and up-to-date speakers would use the reinterpreted meaning (opposite quantifier; e.g. BOTH instead of the original ALL).

**2.1 METHOD.** We recruited 25 native speakers of German by advertising on university newsletters and on university related social media groups (11m/14f; mean age 23.1, SD 3.2). They were financially remunerated for their participation. The experiment was conducted on-site at the University of Graz. Participants were separated into two groups; one group would learn that *gure* originally had the meaning of BOTH (13 participants), the other learnt that *gure* originally had the meaning of ALL (12 participants). In addition to *gure*, participants were taught two filler nonce-words whose meaning had no presuppositional component.

The experiment was split into a training phase and a test phase. During the training phase, participants were told that they should imagine being accompanied by an older native speaker as well as a young non-native friend. They were then shown images on a computer screen and heard sentences containing a nonce word, produced by the non-native person describing the situation. After this, the old person would tell participants whether the sentence was true in the situation presented. If the sentence was not true, the old person would, in addition, provide a reason why it

was false. The older speaker didn't adhere to a maximize presupposition principle; in situations where two out of two items were relevant, they accepted both the meaning of BOTH and ALL as true. Spoken stimuli of native speakers were produced in a version of the Saarland dialects (a remote variant of Mosel-Franconian), which would sound exotic in the South-Eastern Austrian region of Graz where the study was conducted. The non-native friend was a speaker from Graz who spoke a rather neutral dialect similar to Standard German. A sample item is given in Figure 1, with an added English translation. There the contribution of the old speaker differed in the ALL→BOTH and BOTH→ALL conditions respectively.



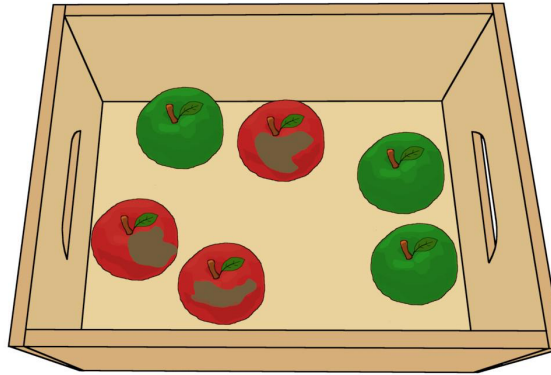
- |                   |   |            |
|-------------------|---|------------|
| Alter Sprecher:   | “Sag mal, Sarah, was siehst du in dem Korb?”            |            |
| Junge Sprecherin: | “Gure Äpfel sind rot.”                                  |            |
| Alter Sprecher:   | “Sehr gut.”   | (ALL→BOTH) |
|                   | “Nein, das stimmt nicht. Es sind mehr als zwei Äpfel.”  | (BOTH→ALL) |
| Old speaker:      | “Tell me, Sarah, what do you see in this basket?”       |            |
| Young speaker:    | “Gure apples are red.”                                  |            |
| Old speaker:      | “Very good.”  | (ALL→BOTH) |
|                   | “No, that's not right. There are more than two apples.” | (BOTH→ALL) |

Figure 1: Sample item, training phase

After three training items each, participants were asked to rate the truth of five sentences on a binary scale. This was done in order to verify the success of the training. In addition to judgments, we also measured the reaction times of participants. Participants received written feedback after each judgment whether their choice was correct, as well as an explanation of their mistake in case they were wrong. Afterwards, they were exposed to additional blocks consisting of training items and test judgments. In total, the training phase consisted of 5 *gure* blocks and 2 filler blocks. All training items were presented in a fixed order.

In the second phase of the experiment (test phase), participants were asked to imagine visiting a reunion of younger members of the community. The young native speakers would use *gure* in its reinterpreted meaning, i.e., BOTH →ALL or ALL→BOTH. Participants were introduced to a friend F who had been abroad for some time and was thus not up to date with current language developments, and a high prestige competent local speaker S. They were then shown images and heard sentences. Participants were told that they didn't know which of the people attending the reunion had produced the sentences (which means that the speaker could either have high or low competence regarding the reinterpreted meaning). After hearing the sentences, participants were asked to rate their agreement for the sentence on a scale from 1 to 10. After each item, participants

then read a short dialogue between F and S commenting on the situation. Both would mention if they thought that the speaker had made a mistake but would not explain why they thought the sentence was wrong. Participants were shown 18 items containing *gure* and 6 filler items which were separated into five blocks and randomized within the blocks. The meaning of some fillers had changed compared to the meaning used by the older speaker during training, while the meaning of others had not. A sample item is given in Figure 2.



Sprecher: “Gure roten Äpfel sind faul.”  
 BOTH→ALL: F: “Hast du den dritten faulen Apfel gesehen?”  
 S: “Wieso sagst du das, Lara? Das stimmt doch, er hat ‘gure’ gesagt.”  
 ALL→BOTH: F: “Mist. Wie viele grüne haben wir?”  
 S: “Das passt aber nicht. “Gure” klingt nach etwas, was hier höchstens meine Oma sagen würde!”

Speaker: “Gure red apples are rotten.”  
 BOTH→ALL: F: “Did you see the third rotten apple?”  
 S: “Why do you say that, Lara? He said ‘gure’, he was right.”  
 ALL→BOTH: F: “Damn. How many green ones do we have?”  
 S: “That’s not right. ‘Gure’ sounds like something my grandma would say here!”

Figure 2: Sample item, experiment phase

2.2. RESULTS. All participants completed the training successfully, meaning that they answered with the expected judgments in the last blocks. Also, judgments of items containing those fillers whose meaning had not changed compared to the training phase were as expected, i.e., they did not change significantly compared to judgments during training. We analyzed reaction times and judgments of the test phase by fitting data for all relevant items (i.e., items where either *both* would be correct but *all* would not be, or vice versa) with linear mixed models in R (R Core Team 2021) using the package *lme4* (Bates, Mächler, Bolker & Walker 2015).<sup>2</sup> In particular, we used the order

<sup>2</sup> There is a methodological question whether our 10-point Likert scales fulfill the requirements for applying the *lmer* function from the *lme4* package. After all, one could argue that our data could in principle exhibit ceiling effects and thus should rather be analyzed with the cumulative link mixed model such as the *clmm* function from the ordinal package (Christensen 2019). We generally believe that a 10-point Likert scale, on which the extremes were rarely reached at all in the experiment, does qualify as discrete rather than ordinal data much more than in the case of a 4- or 7-point Likert scale. Thus, treating our results as involving ordinal data would lead to loss of information. We did perform cumulative link models with the same parameters in all cases and found similar trends, but generally less

of presentation of the items, the group variable (reflecting the BOTH→ALL vs. ALL→BOTH) conditions, and random slopes and intercepts for items for both reaction times and judgments, as reflected by the formulae in (1) and (2) below:

- (1)  $\text{reactiontime} \sim \text{order} * \text{group} + (1 + \text{group} | \text{item}) + (1 + \text{order} | \text{item})$
- (2)  $\text{judgment} \sim \text{order} * \text{group} + (1 + \text{group} | \text{item}) + (1 + \text{order} | \text{item})$

Results are shown in Figure 3 and Table 1 for reaction times and in Figure 4 and Table 2 for judgements (using R packages Lüdtke 2021 for plotting and Fox, Price & Weisberg 2019 for Anova). As can be seen, we found a significant main effect for order in both cases, and an interaction between order and group for reaction times, and a main effect for groups in the case of judgments.

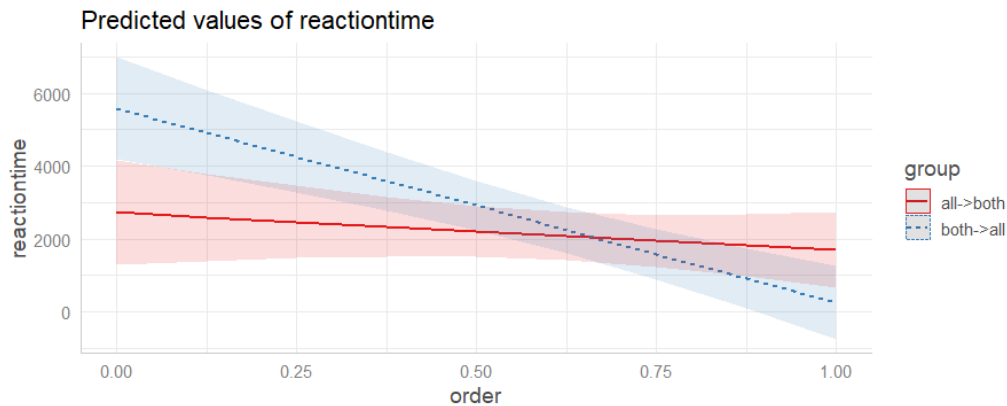


Figure 3: Results Reaction Time, Experiment 1

	Chisq	Df	Pr(>Chisq)	
order	14.4387	1	0.0001448	***
group	3.3893	1	0.0656210	.
order:group	13.0456	1	0.0003037	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 1: Anova for reaction time model, Experiment 1

In the plots, “reactiontime” indicates the time between the end of the audio stimulus (i.e. heard sentence) and the participants reaction (button press). “judgment” represents acceptance of a sentence given a situation on a scale from 1 to 10. For better comparability, we inverted the scale of judgments for the ALL→BOTH-group so that it goes from 10 to 1 instead of 1 to 10. This way, low judgment values for both groups indicate judgments according to the original meaning of *gure* learnt during training, while high values indicate acceptance of the reinterpreted meaning. On the

significant effects. In the case of experiment 1, the cumulative link mixed model shows a trend of order ( $p = 0.051$ ), no significant effect of group, and no significant interaction; the dataset is too small to support both main effects in addition to interaction. In contrast, a cumulative link mixed model taking into account order and group but assuming no interaction does show significant effects of order ( $p = 0.00608$ ) as well as group ( $p = 0.00233$ ). While we maintain that we deem linear mixed models to be a more appropriate method of analysis for our data, we acknowledge this to be a debatable issue.

x-axis, order represents participants’ progress during the experiment phase (order 0 is the start of the experiment, and order 1 the end).

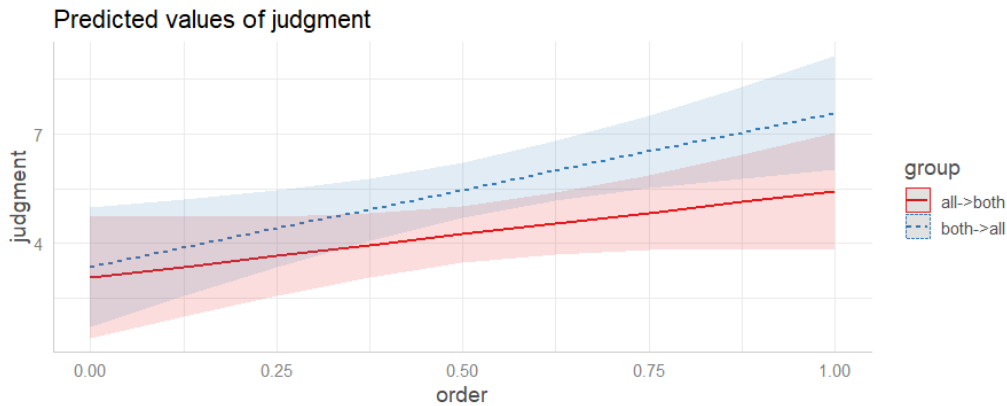


Figure 4: Results Judgments, Experiment 1

	Chisq	Df	Pr(>Chisq)	
order	7.0845	1	0.007775	**
group	6.6758	1	0.009773	**
order:group	1.6200	1	0.203098	

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 2: Anova for judgment model, Experiment 1

2.3 DISCUSSION. Looking at the judgments data in more detail, we can observe that some participants would accept the reinterpreted meaning rather quickly while others wouldn’t accept the changed meaning at all. The result is a split between participants with some responding with high values of judgment towards the end of the experiment and some with low values. This trend was not group specific; both groups showed a similar split. It is likely due to a general reluctance of some people to accept deviations from a meaning they had been explicitly taught before.

More importantly, we take the combination of the analysis of reaction times and judgments to have an added value regarding the interpretation of the results. While reaction times only show an interaction suggesting that participants got quicker over time in the BOTH→ALL group more than in the other group, and the judgments only show that BOTH→ALL was also judged better overall (with only a marginal tendency of an interaction), we believe that these findings provide strong evidence for a quicker and better adaptation to language change in the BOTH→ALL direction.

**3. Experiment 2.** To get more reliable information on whether participants actually understood the meaning of *gure* as a presupposition (as opposed to truth-conditional meaning), we conducted follow-up experiments which replicated the method of Experiment 1 but added presupposition tests (Family of Sentences tests). In addition, replicating Experiment 1 appeared to be a useful step overall, given that our findings in the first experiment can be widely considered explorative and thus less reliable despite statistical significance.

3.1 METHOD. We recruited 24 native speakers of German, again by advertising on university newsletters and on university related Facebook groups (16f/8m; mean age 24.4, SD 4.2). They were financially remunerated for their participation. Due to a technical error, two participants’ data

were unusable, resulting in 22 usable datasets (16f/6m; mean age 23.8, SD 3.3), 11 per group. The experiment was conducted on-site at the University of Graz.

The setup of the training and test phases was identical to Experiment 1. In addition, we added presupposition test surveys both after the training phase and after the test phase. In each survey, participants were shown 15 items consisting of a sentence from the Family of Sentences paradigm as well as a question about that sentence. Five out of those questions tested the presupposition of there being exactly two relevant items, while the other 10 questions tested other conditions. Participants were then asked to answer the question on a binary scale (“yes” or “no”). Assuming *S* to be any sentence, we used five different types of embedding: *if-S*, *not-S*, *maybe-S*, *thinks-that-S*, and *says-that-S*. An example for an *if-S* embedding together with a translation is shown in Figure 5.

Jessica sagt: „Wenn gure Milchpackungen leer sind, müssen wir neue kaufen.“  
 Kannst du dann davon ausgehen, dass Jessica der Meinung ist, dass es genau zwei Milchpackungen gibt?  
 Jessica says: “If gure cartons of milk are empty, we have to buy new ones.”  
 Can you then assume that Jessica thinks that there are exactly two cartons of milk?

Figure 5: Example item presupposition test, Experiment 2

3.2 RESULTS. Again, all participants completed the training successfully. The statistical analysis we used for the two experiments were identical.

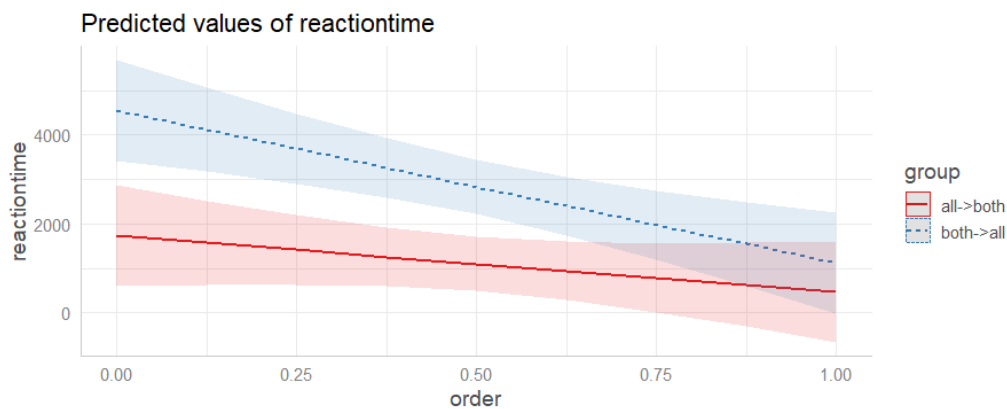


Figure 6: Reaction time results, Experiment 2

	Chisq	Df	Pr(>Chisq)	
order	11.6035	1	0.0006583	***
group	15.2507	1	9.414e-05	***
order:group	2.4792	1	0.1153595	

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 3: Anova for reaction time model, Experiment 2

Judgment and reaction time results roughly mirror the results from Experiment 1, which was expected due to the analogous method; these are shown in Figure 6 and Figure 7 graphically and

the significance tests for the linear models in Table 3 and Table 4, respectively. This time, however, for both reaction times and judgments, we observed significant main effects of group and order but no significant interaction.<sup>3</sup>

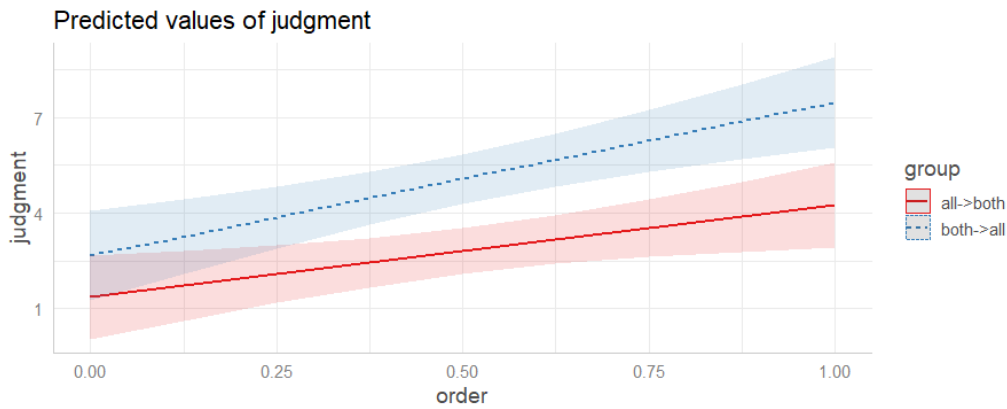


Figure 7: Judgment results, Experiment 2

	Chisq	Df	Pr(>Chisq)	
order	20.3314	1	6.512e-06	***
group	18.4323	1	1.760e-05	***
order:group	1.3985	1	0.237	

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Table 4: Anova for judgment model, Experiment 2

We analyzed the presupposition tests by comparing the mean results of both surveys for each group. The results are shown in Figure 8, which represents the numerical values also summarized in Table 5.

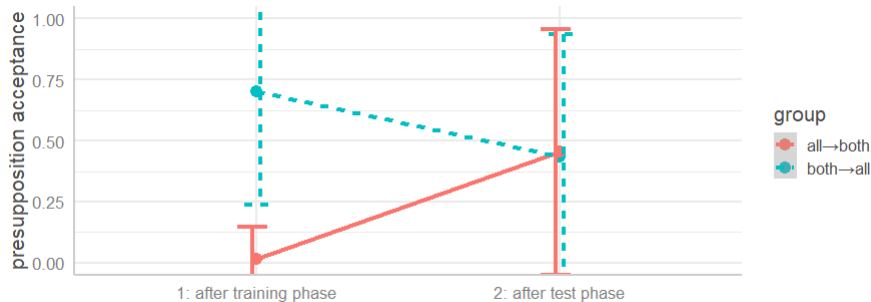


Figure 8: Presupposition test acceptance, Experiment 2

As expected, there is a considerable difference between the two groups after the training phase, since only the *both*-group actually acquired a presupposition, whereas the results strongly overlap after the test phase. Even if visually obvious, we checked for statistical significance using a

<sup>3</sup> Unlike Experiment 1, here the cumulative link mixed model taking into account both main effects as well as interaction showed significant effects of order ( $p = 0.00886$ ) and group ( $p = 0.03826$ ), and no significant interaction.

generalized linear mixed model with group and phase as predictors and random intercepts for items. This revealed a highly significant interaction, as witnessed in Table 6.

	after training	after test
all→both	0.02±0.13	0.45±0.50
both→all	0.70±0.46	0.44±0.50

Table 5: Numerical results for presupposition tests

	Chisq	Df	Pr(>Chisq)	
phase	0.0040	1	0.9492653	
group	5.3978	1	0.0201618	*
phase:group	15.1168	1	0.0001011	***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Table 6: Anova for presupposition test model, Experiment 2

3.3 DISCUSSION. Experiment 2 confirmed the findings of Experiment 1 and additionally showed that participants acquired the presuppositional meaning of *gure* in the training phase and lost it during the test phase for the BOTH→ALL group, whereas in the other condition they started without a presuppositional meaning for *gure* after the training phase and acquired it to some degree during test phase. The high standard deviation for the presupposition test after the test phase can be explained by the observed split in participants’ acceptance of the reinterpreted meaning. This would result in mean judgments near the middle of the total range with high standard deviations, though there were clear groups of participants who achieved a very high level of learning in both groups.

**4. Overall discussion.** Our results indicate that it is easier to perform reinterpretation in a contextualized language change situation in such a way as to lose rather than gain a presupposition. This is witnessed by a quicker and better learning process in terms of adaptation to ongoing language change in the case of presuppositional loss as compared to acquiring a presuppositional item. This has been ascertained in Experiment 1 in an exploratory manner and confirmed in Experiment 2. In addition, Experiment 2 also confirmed that the acceptabilities measured in both experiments are due to presuppositional effects proper and not some alternative prototypicality effects. However, our results are limited to one single presuppositional item. Clearly, such preliminary results must be refined and validated further, especially using further types of presupposition triggers.

From an experimental perspective, the paradigm we have set up invites both extensions and clarifications. While we sought contextualization, including with respect to the simulation of sociolinguistic factors that are relevant for change, it should be obvious that an artificial set-up can always be improved in multiple ways. A further factor that is inherent to most presuppositional conditions will be that they impose additional restrictions (e.g., the duality restriction in our case). This opens up the possibility that a reinterpretation leading towards their loss is also one that leads to a generalization. While in actual change both generalization and specialization can occur, this is experimentally a path that we plan to control in future experiments, also independently of purely presuppositional items. The fact that we only studied one item equally invites further research.

From a language change perspective, notice that we did not start from the expectation of replicating a particular change, but rather from the idea of distilling relevant cognitive behavior

patterns that can provide some of the fundamental building blocks in the construction of change.<sup>4</sup> The complexities of actualized changes can require experimental checking with respect to multiple dimensions. Places in which the potential weakening of the duality restriction we have studied here may show are, as discussed in Gergel (2022), developments of the quantified *both* in early Middle English times in which it was reinforced by the numeral *two* as well as similar developments in the histories of some Romance languages in which versions of the original Latin quantifier *ambo* have equally been reinforced by numerals as e.g., in developments of Romanian *amândoi* or Spanish *ambos dos* (while the latter is prescriptively often argued against, the former is unmarked).

More generally, however, notice that with respect to theories of actual change our experimental results came closest to Eckardt's assumptions about the possibility of losing presuppositions. But the interesting suggestions from Eckardt's work, too, can and must be qualified from the present perspective, as we did not introduce any particularly critical and burdensome situations in which speakers would be overwhelmed with multiple inferences. Therefore, as it stands, our current result detaches potential presuppositional loss from a principle such as Avoid Pragmatic Overload.

## 5. References.

- Bates, Douglas, Martin Mächler, Ben Bolker & Steven Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67,1. 1–48. <https://>
- Beck, Sigrid & Remus Gergel. 2015. The diachronic semantics of English *again*. *Natural Language Semantics* 23,3. 157–203.
- Beck, Sigrid, Polina Berezovskaya, & Katja Pflugfelder. 2009. The use of *again* in 19th-century English versus Present-Day English. *Syntax*, 12,3. 193-214.
- Carlier, Anne & Béatrice Lamiroy. 2018. The emergence of the grammatical paradigm of nominal determiners in French and in Romance: Comparative and diachronic perspectives. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 63. 141-166.
- Christensen, Rune Haubo Bojesen. 2019. ordinal: Regression Models for Ordinal Data. <https://cran.r-project.org/web/packages/ordinal/>.
- Eckardt, Regine. 2006. *Meaning Change in Grammaticalization*. Oxford: OUP.
- Eckardt, Regine. 2009. APO: Avoid pragmatic overload. In Jacqueline Visconti and Maj-Britt Mosegaard Hansen (eds.), *Current Trends in Diachronic Semantics and Pragmatics*, 21–42. London: Emerald.
- Fedzechkina, Masha & Gareth Roberts. 2020. Learners sacrifice robust communication as a result of a social bias. *Proceedings of the 42<sup>nd</sup> Annual Meeting of the Cognitive Science Society*, 2281–2287.
- Fox, John, Brad Price, & Sanford Weisberg. 2019. car: Companion to Applied Regression. <https://cran.r-project.org/web/packages/car/>.
- Fuchs, Martín, Ashwini Deo & María Mercedes Piñango. 2020. The Progressive-to-Imperfective shift. Contextually determined variation in Rioplatense, Iberian, and Mexican Altiplano Spanish. In Alfonso Morales-Front, Michael J. Ferreira, Ronald P. Leow and Cristina Sanz (eds.), *Hispanic Linguistics: Current issues and new directions*, 119–136. Amsterdam: Benjamins.

---

<sup>4</sup> For so-called atoms from the perspective of semantic theory, see e.g., Grosz (2022) for a recent take and the literature cited there. Our focus here is of course on cognitively verified building blocks.

- Gergel, Remus. 2020. Sich ausgehen: Actuality entailments and further notes from the perspective of an Austrian German motion verb construction. *Proceedings of the Linguistic Society of America*, 5(2). 5–15. <https://doi.org/10.3765/plsa.v5i2.4790>.
- Gergel, Remus. 2022. Cyclicity effects in the development of presuppositions. Ms. Saarland University.
- Gergel, Remus, Andreas Blümel & Martin Kopf. 2016. Another heavy road of decompositionality: Notes from a dying adverb. *University of Pennsylvania Working Papers in Linguistics* 22, 1.
- Gergel, Remus & Martin Kopf-Giammanco. 2021. ‘Sich ausgehen’: On modalizing go constructions in Austrian German. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 66,2. 141-190.
- Gergel, Remus, Martin Kopf-Giammanco & Maïke Puhl. 2021. Simulating semantic change: a methodological note. In *Proceedings of Experiments in Linguistic Meaning (ELM) 1*, 184–196. University of Pennsylvania: LSA Publications. <https://doi.org/10.3765/elm.1.4869>.
- Grosz, Patrick. 2022. Scalarity as a meaning atom in *wohl*-type particles. In Remus Gergel, Ingo Reich, Augustin Speyer (eds.), *Particles in German, English, and Beyond*, 243–268. Amsterdam: Benjamins.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.
- Los, Bettelou & Erwin Komen 2012. Clefts as resolution strategies after the loss of a multifunctional first position. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, 884–898. Oxford: OUP.
- Lüdecke, Daniel. 2021. *sjPlot: Data Visualization for Statistics in Social Science*. <https://CRAN.R-project.org/package=sjPlot>.
- Puhl, Maïke & Remus Gergel. 2022. *Final though*. In Remus Gergel, Ingo Reich, Augustin Speyer (eds.), *Particles in German, English, and Beyond*, 177–208. Amsterdam: Benjamins.
- R Core Team. 2021. R: A language and environment for statistical computing. Vienna, Austria. <https://www.R-project.org/>.
- Schwenter, Scott & Richard Waltireit. 2010. Presupposition accommodation and language change. In Kristin Davidse, Lieven Vandelanotte, and Hubert Cuyckens (eds.), *Subjectification, Intersubjectification and Grammaticalization*, 75–102. Berlin: Mouton de Gruyter.
- Traugott, Elizabeth Closs & Richard B. Dasher. 2012. *Regularity in Semantic Change* Cambridge: Cambridge University Press.
- Zhang, Muye, Maria Mercedes Piñango & Ashwini Deo. 2018. Real-time roots of meaning change: Electrophysiology reveals the contextual-modulation processing basis of synchronic variation in the location possession domain. *40th Annual Conference of the Cognitive Science Society*, 2783–2788.