

Abductive inferences in causal discourse: Evidence from eyetracking during reading

When we interpret causal statements in discourse, we not only integrate causes and effects incrementally, but also immediately take relevant world knowledge into consideration in doing so (Köhne-Fuetterer et al. 2021, Kuperberg et al. 2011, Xiang/Kuperberg 2015, Xu et al. 2017). Accordingly, Kuperberg et al. (2011) showed that two-sentence discourses violating domain knowledge immediately give rise to an N400 effect even in the absence of explicit discourse marking. The present study contributes to this line of research by providing first online evidence that even more fine-grained subtypes of inferential processes occur in online processing. Consider (1):

- (1) *Weil [Alex sich an die Aufbauanleitung hielt]_{cause2}, [ging die Spülmaschine kaputt]_{effect}.*
Because [Alex followed the assembly instructions]_{cause2}, [the dishwasher broke down]_{effect}.

In isolation, (1) seems *anomalous*. World knowledge predicts an effect to the contrary (cause2 \Rightarrow \neg effect; cf. Aliseda 2006): Without other evidence, following the instructions prevents a machine from breaking down. (1) should therefore either be rejected, or taken to constitute a partial explanation, leading to the introduction of an additional cause via abductive inferencing (Aliseda 2006).

The anomaly may disappear once (1) is embedded in a larger context. One can think of a number of situations in which (1) could make sense. Consider, for instance, the complex cause in (2), in which a cause1 has been added to cause2 in (1):

- (2) *Weil [die Aufbauanleitung einen Fehler enthielt]_{cause1} und [Alex sich an die Anleitung hielt]_{cause2}, [ging die Spülmaschine kaputt]_{effect}.*
'Because [the assembly instructions contained an error]_{cause1} and [Alex followed the instructions]_{cause2}, [the dishwasher broke down]_{effect}.'

Taken together, the erroneous assembly instructions (cause 1) and Alex following these (cause 2) may be taken to *fully explain* the effect.

The present study investigated how partial explanations like (1) are processed in discourse. Moreover, we compared two types of causal relations differing in their involvement of domain knowledge. In addition to anomalous sequences as in (1), where the opposite effect is expected, we included situations where world knowledge doesn't make a particular prediction, introducing what we characterize as *novel* causal relations (cause2 \nRightarrow effect and cause2 \nRightarrow \neg effect):

- (3) *Weil [Maria sich auf die Bank setzte]_{cause2}, [bekam sie einen schlimmen Ausschlag]_{effect}.*
Because [Mary sat on the bench]_{cause2}, [she got a bad skin rash]_{effect}.

(3) doesn't contradict world knowledge: Rather, cause and effect seem unrelated: Whatever the possible effects of sitting on a bench are, getting a skin rash is usually not among them.

Materials: 15 anomalous and 15 novel discourses were constructed in three discourse order variants according to a 3x2 design (discourse order x causal relation). Discourses with the two causes conjoined within a *because* clause (*because cause1 and cause2, effect*, cf. (2)) served as controls. *Left dislocation* conditions, where *cause1* preceded the *because* clause (*cause1. because cause2, effect*), tested how easily causes can be integrated when not embedded under a causal connective. In the *right dislocation* condition of most interest here, *cause1* followed the *because* clause (*because cause2, effect. cause1, you see*). Importantly, all three orderings contained exactly the same 'nucleus' (*because cause2, effect*). All discourses were preceded by a two-sentence sequence introducing all referents and ended with a sentence concluding the story.

Pretests: Materials were pretested with respect to several aspects. Most importantly, the causal connectedness of anomalous and novel causal relations was rated (N=24) on a scale from -3 (*highly contradictory*) to +3 (*highly natural*) with 0 explicitly requested to indicate *no causal*

connection. We tested three conditions: (i) similar to (1)/(3): *because cause2, effect*; (ii) negating *cause2*: *because negated(cause2), effect* ("because Alex did **not** follow the instructions, the dishwasher broke down"), and (iii) *because cause1 and cause2, effect*, as in (2). As expected, anomalous relations were rated oppositely in the positive (i) and negative (ii) cases (mean ratings: -2.0 vs. 2.1), whereas novel and negated novel cases both had no causal connection (0.0 vs. 0.1). Crucially, both types were rated as natural when they were part of a complex cause (*anomaly* 2.0; *novelty* 2.1). Another pretest (N=30) established that all three conditions for both *anomaly* and *novelty* items were rated equally plausible as a whole.

Predictions concerning **right dislocation** (*because cause2, effect*) as in (1) and (3) were captured in the framework of Halpern/Pearl (2005). Both anomalous and novel causal relations invoke a causal network consisting of a cause and an effect variable. However, the networks differ in one important respect: Anomalous relations violate established world knowledge, predicting a contrary distribution of cause and effect. Consequently, integrating *because cause2, effect* leads to a contradictory causal model calling for revision. Readers are therefore predicted to regress from the effect region to earlier parts of the discourse to check whether they had parsed *cause2* incorrectly. In novelty cases, on the other hand, the simple model invoked by *because cause2, effect* isn't contradictory, but insufficient. This is predicted to lead to abductive reasoning as to how the model could be plausibly *extended*. We thus expected integration difficulty right at the effect clause, that is, enhanced first-pass times on the effect clause, but less regressive eye-movements than for *anomaly*. In the **left dislocation conditions** (*cause1. because cause2, effect*), we assumed incremental discourse interpretation with immediate access to the global discourse representation (Hagoort/van Berkum, 2007). Thus, integrating the effect clause shouldn't be more difficult than in the control condition. Similarly, in Halpern/Pearl's theory, left dislocation (and control) provide full explanations, for which no abductive modelling effort is required.

Eye-tracking experiment: Participants (N=27) read the discourses plus 30 filler texts while their eye-movements were monitored using an EyeLink 1000 system. In line with our predictions, *left dislocation* didn't differ from control at any segment. By contrast, *right dislocation* led to longer first-pass times and more regressions from the *effect* ROI. Furthermore, the effects differed for *anomaly* and *novelty*. **Inferential statistics** analyzing residual first-pass times of the *effect* ROI revealed a reliable interaction: Whereas *novelty* led to significantly longer first-pass times than control (mean difference: 169.7ms; $p < .01$), *anomalous right dislocations* didn't differ reliably from control (mean difference: -6.7ms). A logit mixed effects model analysis of first-pass regression ratios revealed an opposite pattern with significantly more regressions out of anomalous effect clauses (16.3%, control: 7.4%; $p < .05$) than for novel ones (12.6%, control: 8.9%; $p = .34$). Analyses of the second-pass times of *cause2* revealed the same interaction. *Right dislocation* led to longer second-pass times (SPT) than control, but this effect was more pronounced for *anomaly* (mean SPT: 679.1ms, control: 250.0ms) than *novelty* (mean SPT: 467.1ms, control: 292.6ms), as shown by a significant interaction (estimate = -262.05, $t = 2.42$, $p < .05$).

In **conclusion**, the eye-tracking record of *anomalous* vs. *novel right dislocation* shows that subtle world knowledge distinctions and their associated inferential profiles are reflected in different temporal profiles when inferring from partial to full explanations during text comprehension.

REFERENCES

Aliseda (2006): *Abductive Reasoning*. Springer. **Hagoort/van Berkum** (2007). Beyond the sentence given. *PTRS*, B: 362. 801–811. **Halpern/Pearl** (2005). Causes and Explanations: A Structural-model Approach. *BJPS*, 56. 843–911. **Hobbs et al.** (1993). Interpretation as Abduction, *AI*, 63 (1-2): 69–142. **Köhne-Fuetterer et al.** (2021). The online processing of causal and concessive discourse connectives. *Linguistics* 59. 417–448. **Kuperberg et al.** (2011). Establishing Causal Coherence across Sentences. *JCN*, 23(5): 1230–1246. **Xiang/Kuperberg** (2015). Reversing expectations during discourse comprehension. *LaCoN* 30. 648–672. **Xu et al.** (2017). Influence of Concessive and Causal Conjunctions on Pragmatic Processing. *Disc. Proc.* 55. 387–409.