

Eventuality type predicts temporal order inferences in discourse comprehension

Elena Marx, Department of Cognitive Science, Central European University, Vienna, Austria, marx_elena@phd.ceu.edu

Eva Wittenberg, Department of Cognitive Science, Central European University, Vienna, Austria, wittenberge@ceu.edu

One kind of temporal inference in discourse operates over iconicity, such that inferred temporal order follows reported order. In two preregistered experiments (combined N = 930), we asked whether this temporal inference is predictably modulated by linguistic eventuality. Based on event-structural theories of temporal interpretation, stative descriptions, corresponding to cognitively less salient states in the world, should serve as backgrounds for eventive descriptions, locating states earlier in time. Participants read descriptions like *Mary got/was married to John. She got/was pregnant* and indicated which happened first. Eventuality type of both sentences and reported order were crossed. We find that states tend to be ordered before events, and longer states before shorter states. Our results support a model of discourse comprehension in which eventuality framing is crucial for (temporal) inferences.



1. Introduction

Understanding temporal order is a crucial human capacity (Downes et al., 2002; Kausler et al., 1988; Sacks, 1985). When we read a story, some of the most important information to keep track of is what happens when, and how readers do this in discourse comprehension has been a vivid question in research for decades (e.g., Graesser, Singer, & Trabasso, 1994; Jakobson, 1965; Kehler, 1994, 2002; Oversteegen, 2005; Tai, 1983). Here, we focus on the temporal inferences people draw to mentally order situations expressed in successive sentences.¹ Our starting point is a classic example:

- (1) Mary got pregnant. She got married to John.
- (2) Mary got married to John. She got pregnant.

Examples (1) and (2) can be found in various versions throughout the literature on discourse comprehension over the past 70 years or so (e.g., Fleischman, 1990; Horn, 2022; Kamp & Rohrer, 1983; Lascarides & Asher, 1991, 1993; Levelt, 1989; Schmerling, 1975; Strawson, 1952; Webber, 1988). They have been used to illustrate that the order of successive sentences in a narrative is interpreted as iconically mapping to the temporal order of the expressed situations: The inferred order in (1) is that Mary was pregnant before she married John; example (1) may also trigger a causal inference, namely, that perhaps Mary and John got married because Mary was already pregnant, presumably – another inference – by John. Conversely, the inference triggered by (2) is that Mary and John first got married, and then, Mary became pregnant. In any case, the example pair in (1) and (2) has been presented to generations of linguistics students as evidence that inferred orders are triggered by reported order in linguistic discourse. However, this claim is incomplete on two accounts: First, there is a lack of empirical or experimental data to confirm these intuitions (Gibson & Fedorenko, 2010); and second, while the literature clearly recognizes effects of eventuality type (Hinrichs, 1986; Hopper, 1979), i.e., whether a situation is relayed using an eventive or a stative description, it is lacking a truly methodical investigation of how linguistic eventuality type systematically affects these inferences.

Most work on discourse inference has been concerned with causal inference (Briner et al., 2011; Oversteegen, 2005; Wolfe et al., 2005). Causal inference is, of course, tightly linked to temporal inference, since causes precede effects (cf. the Causal Law, Lascarides & Asher, 1993).

¹ There is considerable variation in the notions employed around the linguistic expression of event structure. We will use the following terminology, to minimize confusion between linguistic notions and cognitive counterparts: We refer to the order of events imagined as they happened, as *inferred order*, but to the order of successive sentences as *reported order*. When talking about non-linguistic events or states, we call them *situations*, and their linguistic encodings, following Bach (1986), are called *eventualities*. Situations comprise both non-dynamic *states* and change-of state *events*; which are expressed linguistically as *stative descriptions* or *eventive descriptions*. For the purpose of this paper, we do not distinguish between achievements and accomplishments.

However, not all temporal sequences are also causally related, or form contingency relations. For instance, in (2), Mary's marriage may or may not have causally contributed to the subsequent pregnancy; and even any potential causal inference in (1) is, at least in our intuition, weaker than the temporal order inference. For the moment, we therefore leave aside the question of causality and focus only on the temporal inference.

Temporal inference in discourse is made easier when two successive sentences are marked through temporal adverbs or conjunctions, or with different tense morphology. In (1') and (2'), for instance, the temporal order inferences are countercorrelated with the reported orders, because in (1'), a temporal conjunction leads to the temporal reordering of the described situations, whereas in (2'), the first sentence is marked with present tense and the second with past tense:

(1') Mary got pregnant **after** she got married to John.

(2') Mary **gets**_{PRESENT} married to John. She **got**_{PAST} pregnant.

There are two broad classes of theoretical accounts explaining how people infer temporal order of events in discourse through tense: one in which tense serves as an anaphora that guides the relative temporal ordering of situations (Carroll & von Stutterheim, 2010; Hinrichs, 1986; Klein, 1994, 2009; Oversteegen, 2005; Partee, 1973; von Stutterheim et al., 2003; Webber, 1988), and another in which temporal order inferences are byproducts of establishing overall coherence in the discourse, using tense as a cue (Kehler, 1994, 2002, 2006; Lascarides & Asher, 1993). Based on tense alone, both families of accounts would predict that in (1) and (2), the inferred order follows the reported order, since there is no tense differential.

While both of these accounts discuss examples of tense differentials affecting discourse interpretation, a factor that is not manipulated in (1) and (2), only one of them additionally makes clear predictions according to eventuality type, i.e. whether a situation is encoded by a stative description, or by an eventive description. Specifically, Hinrichs (1986) elaborates that stative descriptions following eventive descriptions (examples (5) and (6) below) can be interpreted as temporally overlapping with events (see also Dowty, 1986; Partee, 1984). But why should this be? Hinrichs (1986) simply refers to Vendler's (1967) Aktionsarten as the descriptive categories determining whether successive descriptions trigger sequential or overlapping inferences.

That said, a possible explanation is given by accounts that closely link non-linguistic event cognition to the temporal interpretation of linguistic descriptions, by proposing a *temporal anchor* relative to which other descriptions are interpreted (Carroll & von Stutterheim, 2010; Klein, 1994, 2000, 2009; von Stutterheim et al., 2003; but see also Kameyama et al., 1993; Kamp & Reyle, 1993). Specifically, it has been argued that temporal expressions, such as tenses, denote complex temporal structures that can be formalized along several interrelated temporal intervals: With each temporal expression, speakers make an assertion about a specific time – a topic time

– which for the simple past, selects an interval of the time when a situation took place, including its rightmost boundaries (i.e., a culmination point for achievements and accomplishments, or a post-state for states and activities). This topic time is then matched to another temporal interval: a temporal anchor. Such an anchor can be, among other possibilities, determined by knowledge about event structure; the crucial property is that the temporal anchor should be the most salient element in a discourse. Research on event perception has shown that events are cognitively more salient than stable states (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007); in linguistic discourse, this should mean that eventive predicates tend to function as anchoring situations, and states tend to function as anchored situations. We find a related idea in the distinction between foreground and background, with implications for narrative progression (cf. Hopper, 1979; Wårvik, 2004).

Crucially, given that for simple past, a topic time always includes the rightmost boundary of the situation time (i.e., culmination point/post-state), this also imposes a constraint on its leftmost boundary: That is, the anchored situation cannot start after the time of the temporal anchor. Therefore, we should find that (a), if a situation is linguistically described as eventive, it should serve as a temporal anchor for a stative description. Furthermore, in two sentences in simple past, we should find that (b), in the absence of a simultaneous choice, people’s temporal intuitions should order stative descriptions before eventive descriptions.

We conducted two preregistered experiments in order to systematically address these predictions, using a forced-choice task that probes inferences of temporal order. Experiment 1 manipulated linguistic eventuality in the classical examples (1) and (2) in a single-trial design. In Experiment 2, we extended Experiment 1 in two ways: First, we used a broader range of situation descriptions to assess to what extent our findings would generalize across verbs and events. Second, using multiple items allowed us to test people’s temporal inferences in a within-subject design.

For both experiments, we argue that the widely reported intuitions for sentences like (1) and (2), according to which the inferred temporal order matches the reported linear order of the two sentences, only arise because both Mary’s marriage and her pregnancy are linguistically conveyed by eventive descriptions (1” and 2”):

(1”) Mary **got pregnant**_{EVENTIVE}. She **got married to**_{EVENTIVE} John.

(2”) Mary **got married to**_{EVENTIVE} John. She **got pregnant**_{EVENTIVE}.

The same correlation between reported and inferred order should also hold when the first sentence contains a stative description, such as in (3) and (4). This prediction follows from assuming an iconic mapping between sentence order and temporal order, but more importantly for our purposes, it arises independently from a general cognitive tendency to order states first in time.

(3) Mary **was pregnant**_{STATIVE}. She **got married to**_{EVENTIVE} John.

(4) Mary **was married**_{STATIVE} to John. She **got pregnant**_{EVENTIVE}.

In contrast, only from an event-structural perspective, when a stative description follows an eventive description, such as in (5) and (6), should the inferred order be opposite to the reported order:

(5) Mary **got pregnant**_{EVENTIVE}. She **was married**_{STATIVE} to John.

(6) Mary **got married to**_{EVENTIVE} John. She **was pregnant**_{STATIVE}.

In (5), the state of being married should be inferred to have occurred before the event of becoming pregnant; and in (6), the state of being pregnant should be inferred to have occurred before the wedding event. That is, (5) and (6) are predicted to reverse the temporal inference classically reported in the literature (e.g., Fleischman, 1990; Horn, 2022; Kamp & Rohrer, 1983; Lascarides & Asher, 1991, 1993; Levelt, 1989; Schmerling, 1975; Strawson, 1952; Webber, 1988).

Finally, one open question is how two stative descriptions following each other are ordered relative to each other, as in (7) and (8):

(7) Mary **was pregnant**_{STATIVE}. She **was married**_{STATIVE} to John.

(8) Mary **was married**_{STATIVE} to John. She **was pregnant**_{STATIVE}.

While none of the theoretical approaches mentioned above clearly predicts a temporal order between two states, the assumption that stative descriptions elicit overlapping inferences (Dowty, 1986; Hinrichs, 1986; Partee, 1984) might prevent people from sequentially ordering the two situations in (7) and (8) in time (see also a parallel coherence relation in Kehler, 1994, 2006, without reference to eventuality). In such a case, the two possible temporal orders should be at chance level when participants are asked to choose one over the other.

However, there is broad consensus in the literature that a crucial dimension of states is their duration (see Vendler, 1967; or Carlson, 1977, and Kratzer, 1995, for durational differences between stage/individual-level predicates), apart from their lack of change or dynamicity. Thus, if duration influences the way in which people conceptualize states, we would predict that differences in the duration typically associated with two stative predicates should lead to different temporal construals: That is, given that being pregnant typically lasts a shorter time than being married, pregnancies should be less stative, and thus more salient, than marriages. For (7), we would therefore predict a reversal of reported order, as in (5) and (6), whereas the double-stative sequence in (8) should follow the reported order of the two clauses as in (3) and (4). Predictions for both experiments are outlined in **Table 1**.

Table 1: Outline of predicted inferred temporal order based on eventuality type order in Experiment 1 and 2. Diagonal lines indicate similar predictions following from reported order and event structure.

Eventuality order	Predicted inferred order	Exp.	Example
eventive - eventive	reported order	1	<i>got pregnant – got married</i>
		2	<i>boarded a train – went on holiday</i>
eventive - eventive	reported order	1	<i>got married – got pregnant</i>
		2	<i>went on holiday – boarded a train</i>
stative - eventive	reported order = state before event	1	<i>was pregnant – got married</i>
		2	<i>was on a train – went on holiday</i>
stative - eventive	reported order = state before event	1	<i>was married – got pregnant</i>
		2	<i>was on holiday – boarded a train</i>
eventive - stative	opposite order → state before event	1	<i>got pregnant – was married</i>
		2	<i>boarded a train – was on holiday</i>
eventive - stative	opposite order → state before event	1	<i>got married – was pregnant</i>
		2	<i>went on holiday – was on a train</i>
stative - stative	opposite order → long before short state	1	<i>was pregnant – was married</i>
		2	<i>was on a train – was on holiday</i>
stative – stative	reported order = long before short state	1	<i>was married – was pregnant</i>
		2	<i>was on holiday – was on a train</i>

2. Experiment 1

2.1 Methods

2.1.1 Participants

We recruited 800 native English speakers via Prolific.co, to collect 100 observations per condition.² The experiment was programmed in PsychoPy and conducted via [Pavlov.org](https://pavlov.org/); it lasted approximately 2 minutes. 42 additional participants were excluded, because they terminated the experiment prematurely, or due to data recording problems.

² Sample size was based on the work of Morgan et al. (2020), who used a single-trial design to collect 50 data points per condition (cf. Exp. 4), but we doubled the number of observations per cell, due to unknown effect sizes.

2.1.2 Materials and procedure

Critical stimuli consisted of two main clauses describing two situations (i.e., pregnancy and marriage) in simple past. Situations were crossed in reported order (i.e., pregnancy first: (1), (3), (5), (7) vs. marriage first: (2), (4), (6), (8)); we also crossed eventuality type in the first clause (i.e., eventive: (1–2), (5–6) vs. stative: (3–4), (7–8)) and the second clause (i.e., eventive: (1–4) vs. stative: (5–8)).

We used a single-trial, between-subjects design (see e.g., Morgan et al., 2020 Experiments 1 and 4; von der Malsburg et al., 2020 particularly the belief-estimation tasks). This choice was motivated by task adaptation effects which have been shown to emerge in a variety of experimental designs, both due to the repetition of critical trials (e.g., Balota et al., 2018; Demberg & Sayeed, 2016; Fine et al., 2013; Hammerly et al., 2019; Ness & Meltzer-Asscher, 2021; Pregla et al., 2021), and the inclusion of filler trials (e.g., Cowart, 1997; Keller, 2000; Schütze & Sprouse, 2013), commonly employed in within-subject designs (Arehalli & Wittenberg, 2021; Laurinavichyute & von der Malsburg, 2023, for discussion).

Participants were randomly assigned to one of the eight conditions, and then asked to indicate which of the two situations had occurred first by choosing between two buttons (i.e., pregnancy vs. marriage; **Figure 1**).



Figure 1: Example of the experimental screen participants saw in the forced-choice task, here in the eventive-eventive combination with marriage first. The side of the buttons was randomized.

2.1.3 Statistical analysis

We conducted two types of statistical analyses: First, a logistic regression model was fitted in the R statistical environment (i.e., *glm* from the stats package, R Core Team, 2014) with reported order (i.e., marriage-first vs. pregnancy-first), the linguistic eventuality type in the first sentences (i.e., stative vs. eventive), and the combination of eventuality types across sentences (i.e., same: eventive-eventive/stative-stative vs. mixed: eventive-stative/stative-eventive) as sum-contrast coded predictor variables. The dependent variable was whether people chose the situation that was reported first as having happened first. To assess the significance of the full model, we performed likelihood-ratio tests between the full model and the reduced models, excluding one of the predictors or interactions individually. Model comparison was performed with each of the three predictor variables and the four possible interactions. Second, we ran a series of planned pairwise comparisons to assess the interactions of the regression models. Preregistrations for all analyses are available at <https://osf.io/fnmq4>.

2.2 Results

Participants' mean response choices are shown in **Figure 2** (planned pairwise comparisons between conditions are outlined in **Figure 3**): When the sentences described two events (dark green bars), participants' choices matched the reported order, such that the first reported situation was chosen to occur first. This dovetails with the temporal inferences for (1) and (2) as described in the literature, in which the reported order of situations iconically determines how they are ordered in time – at least when both situations are linguistically encoded as events. Accordingly, the statistical analysis revealed a significant difference in the choices participants made depending on the eventuality type of both clauses (see **Figure 3.I**, $\beta = -0.52$, $t = -13.77$, $p < 0.001$): That is, first reported events were chosen significantly more often when both clauses were eventive descriptions ($\text{mean}_{\text{event-event}} = 0.95$, $\text{SD} = 0.22$) than when they were stative descriptions ($\text{mean}_{\text{state-state}} = 0.43$, $\text{SD} = 0.50$).

When the two reported situations did not match in eventuality type (teal and olive bars), participants chose the first reported situation only when it was a stative description. Thus, for (3) and (4), where the first clause described a state, and the second an event, participants followed the reported order of the two clauses, ordering the situation in the first clause first in time, but for (5) and (6), where the first clause was eventive and the second stative, the reported order was reversed: Here, participants were more likely to choose the second reported situation to have happened first.

Accordingly, the statistical analysis revealed a significant difference between the two mixed-eventuality conditions where events were reported first and the two conditions where states were reported first (see **Figure 3.II**, $\beta = 0.71$, $t = 20.55$, $p < 0.001$). As predicted, participants were

significantly more likely to choose first described situations when they were stative (mean_{stative_first} = 0.93, SD = 0.25) than when they were eventive (mean_{eventive_first} = 0.22, SD = 0.42).

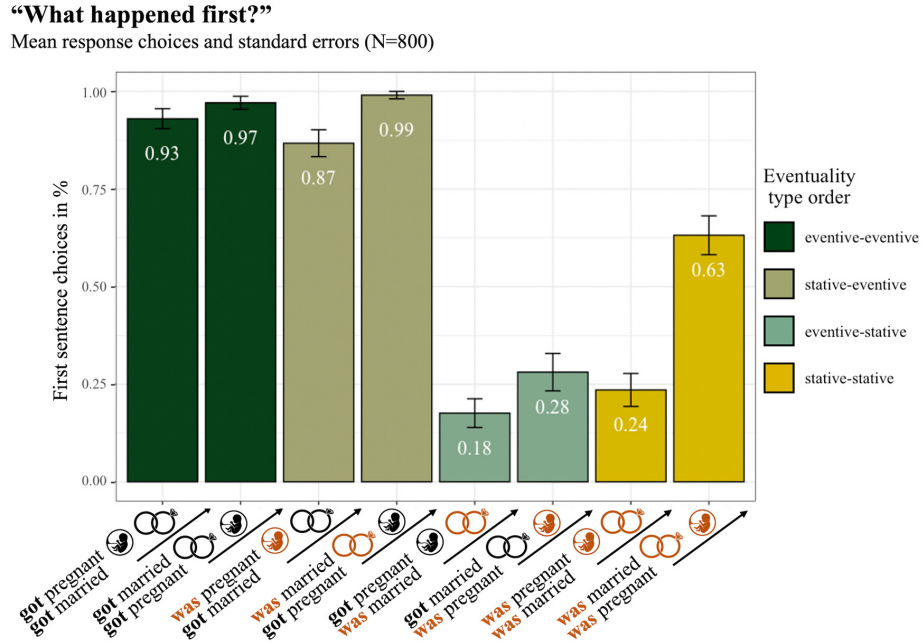


Figure 2: Mean proportions of first sentence choices across eight conditions, error bars represent standard errors. The x axis shows the reported order of the two situations, orange icons indicate stative eventuality type.

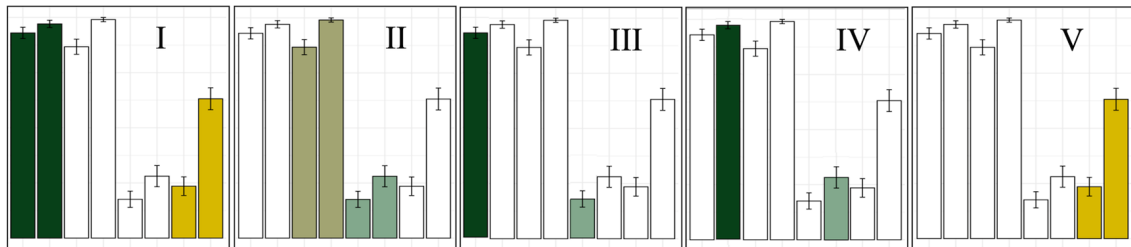


Figure 3: Outline of planned pairwise comparisons, colors of bars reflect compared conditions.

Similarly, when comparing between conditions where an eventive pregnancy was mentioned first (see **Figure 3.III**, $\beta = -0.75$, $t = -16.56$, $p < 0.001$) as well as between conditions where an eventive marriage was mentioned first (see **Figure 3.IV**, $\beta = -0.69$, $t = -14.38$, $p < 0.001$), the situation described first was more likely to be chosen when followed by another eventive description (eventive pregnancy first, ex. (1): mean_{eventive_pregnancy} = 0.93, SD = 0.26; eventive

marriage first, ex. (2): $\text{mean}_{\text{eventive_marriage}} = 0.97$, $\text{SD} = 0.17$) than by a stative description (eventive pregnancy first, ex. (5): $\text{mean}_{\text{eventive_pregnancy}} = 0.18$, $\text{SD} = 0.38$; eventive marriage first, ex. (6): $\text{mean}_{\text{eventive_marriage}} = 0.28$, $\text{SD} = 0.45$).

Neither the main effects of eventuality in the first clause ($Df = 1$, $\chi^2 = 3.22$, $p = 0.07$) and eventuality combination ($Df = 1$, $\chi^2 = 0.95$, $p = 0.33$) reached significance, nor did the two-way interaction between reported order and eventuality combination ($Df = 1$, $\chi^2 = 0.31$, $p = 0.58$) or the three-way interaction that also included eventuality in the first clause ($Df = 1$, $\chi^2 = 1.17$, $p = 0.28$). However, we found a significant interaction between eventuality combination and eventuality in the first clause ($Df = 1$, $\chi^2 = 363.47$, $p < 0.001$): That is, the first reported eventive situation was chosen to happen first, but only if the second reported situation was also eventive. When the first clause described an event and the second a state, participants chose the state to happen first. Conversely, when the first clause described a state, it was only more likely to happen first when the second clause encoded an event.

For double-stative descriptions (ochre bars), i.e., (7) and (8), the pattern was split: When marriage was reported first, participants followed the clauses' reported order ($\text{mean}_{\text{stative_marriage}} = 0.63$, $\text{SD} = 0.48$), but when pregnancy was reported first, they tended to choose the second mentioned situation ($\text{mean}_{\text{stative_pregnancy}} = 0.24$, $\text{SD} = 0.43$). This was also reflected in the statistical analysis: There was (a) a significant difference of reported order between the double-stative conditions (see **Figure 3.V**, $\beta = -0.40$, $t = -6.10$, $p < 0.001$), (b) a main effect of reported order ($Df = 1$, $\chi^2 = 29.30$, $p < 0.001$) such that people were generally more likely to order marriage first when it was reported first ($\text{mean}_{\text{marriage_first}} = 0.74$, $\text{SD} = 0.44$) than when it was reported second ($\text{mean}_{\text{pregnancy_first}} = 0.54$, $\text{SD} = 0.50$), and (c) a significant interaction between reported order and eventuality in the first clause ($Df = 1$, $\chi^2 = 5.57$, $p = 0.018$): Overall, participants chose the first reported situation more often when it was stative and encoded marriage, rather than pregnancy; however, for first reported eventive situations, the difference between reported orders was less pronounced.

2.3 Discussion of Experiment 1

In Experiment 1, we asked about the mechanisms of temporal order inference in discourse, taking the classic example of Mary's pregnancy and her marriage to John as a starting point. By systematically manipulating the reported order as well as the eventuality type of both sentences, we found that linguistic eventuality was a strong predictor of people's temporal interpretations: As predicted, participants were more likely to order states before events, regardless of whether they read about the state in the first or the second sentence. In contrast, reported order, which has been argued to be a strong cue to temporal order in discourse comprehension, only guided people's temporal inferences in double-eventive sentences.

Regardless of the reported order, in double-stative descriptions participants were more likely to order marriage before pregnancy. Although this pattern of results is consistent with our assumption that people conceptualize a described state as more or less stative on the basis of its duration, and thus choose the longer state to have happened first, this pattern of results could also be due to a causal asymmetry between the two situations according to people's world knowledge: That is, people may expect marriages to happen first because of cultural conventions.

In Experiment 2, these two possibilities were examined by testing whether differences in the expected duration of states would also affect people's temporal inferences about other situation descriptions.

3. Experiment 2

Experiment 2 broadened our empirical base from just a single classic item (i.e., Mary's pregnancy vs. marriage) to a range of items tested in a within-participants design. In addition, we followed up on the question whether other dimensions of event structure influence the way in which people conceptualize states: that is, duration. If longer states are conceived as more stative, and thus, more backgrounded, than shorter states, we should find a systematic asymmetry between double-stative sentences that encode different durations across items.

3.1 Methods

3.1.1 Participants

Due to a technical error, we overshot our goal of 100 participants and ended up with 130 full datasets, recruited via Prolific.co. We excluded 23 additional participants who did not meet the sanity check criterion (i.e., at least 50% of the filler trial responses had to be answered correctly). The experiment took approximately 10 minutes. Only complete data sets were submitted to statistical analysis.

3.1.2 Materials and procedure

As in Experiment 1, the critical stimuli consisted of two sentences that were crossed with respect to their reported order (i.e., predicate 1–2 vs. predicate 2–1), the eventuality type in the first sentence (i.e., eventive vs. stative), and the eventuality type in the second sentence (i.e., eventive vs. stative). All sentences consistently used the simple past tense (for a full list of stimuli, see <https://osf.io/5tz6r>).

The two sentences in each item pair were created such that they described two situations that could have happened in either order. For instance, in examples (9–16), Laura's boarding the train and her going on holiday do not necessarily entail a causal order or one-directed contingency:

Laura could board a train to go to her holiday destination, or she could make a train trip after starting her holiday.

- (9) Laura **boarded**_{EVENTIVE} a train. She **went**_{EVENTIVE} on holiday.
- (10) Laura **went**_{EVENTIVE} on holiday. She **boarded**_{EVENTIVE} a train.
- (11) Laura **boarded**_{EVENTIVE} a train. She **was**_{STATIVE} on holiday.
- (12) Laura **went**_{EVENTIVE} on holiday. She **was**_{STATIVE} on a train.
- (13) Laura **was**_{STATIVE} on a train. She **went**_{EVENTIVE} on holiday.
- (14) Laura **was**_{STATIVE} on holiday. She **boarded**_{EVENTIVE} a train.
- (15) Laura **was**_{STATIVE} on holiday. She **was**_{STATIVE} on a train.
- (16) Laura **was**_{STATIVE} on a train. She **was**_{STATIVE} on holiday.

While in Experiment 1, eventiveness and stativeness was based on a minimal difference of the finite verb (eventive: *to get* + adjective vs. stative: *to be* + adjective), in Experiment 2 we based the alternations on the stative and eventive meaning of the entire predicate. Therefore, there was less control over the surface features of the predicate, such as length, but we deliberately were able to broaden the range of predicates (in the above examples, copula constructions with a prepositional phrase, *to board*, and *to go*).

As in Experiment 1, participants were asked to select one of two buttons to indicate which of the two sentence situations had occurred first. However, in this experiment, participants encountered 29 trials in a within-subject design: After a single practice trial, participants completed eight critical trials, one in each condition. In addition, participants also made temporal judgments on 20 fillers, which likewise consisted of two sentences, but used either grammatical (i.e., tense/aspect) or lexical markers (i.e., adverbs and connectives) to establish an unambiguous temporal order between the two reported situations (e.g., *Julius stared out of his window. Then, he went to play with his little brother*). In half of the filler trials, the temporal order did not match the sentences' reported order, in order to induce a reversal of the reported order in some of the trials. Filler trials functioned as a sanity check.

To counterbalance between subjects, participants were randomly assigned to one of eight lists using a Latin square design (reported order, eventuality type combination, situation item). All experiments were programmed in PsychoPy and conducted online at Pavlovia.org.

3.1.3 Statistical analysis

Similar to Experiment 1, we performed two types of preregistered statistical analyses (available at <https://osf.io/5tz6r>): First, we built a logistic regression model with the reported order of situations (i.e., predicate 1–2 vs. predicate 2–1), the eventuality type in the first sentence (i.e., stative vs. eventive), and the combination of eventuality types across sentences (i.e., same: eventive-eventive/stative-stative vs. mixed: eventive-stative/stative-eventive) as sum-contrast coded predictor variables in the R statistical environment (i.e., *glmer* from the stats package, R Core Team, 2014). The within-subjects design allowed us, in contrast to Experiment 1, to include participants and items as random intercepts.

As before, we used as the dependent variable whether participants chose the first sentence to have happened first in time. We performed likelihood ratio tests to compare the full model with reduced models, which excluded one of the predictors or one of the interactions at a time. Additionally, we performed a series of planned pairwise comparisons to further assess the nature of the effects and interactions of the regression models.

For analysis, predicate 1 was coded as the predicate we roughly estimated to be shorter in duration in the stative form (*being on a train* in the examples above) compared to predicate 2 (*being on holiday* in the examples above).

3.2 Results

Participants' mean response choices are shown in **Figure 4**: Overall, we replicated the pattern of results from Experiment 1, extending it to a broader range of situation descriptions as well as to a within-subjects design. All of the effects we found previously were present again, with some additional significant interactions and main effects.

As in Experiment 1, and as predicted, participants were highly likely to follow an iconic interpretation between reported order and temporal order, that is, to choose the first described event to have happened first, but only if both sentences contained eventive descriptions (dark green bars).

However, this tendency was strongly influenced by eventuality type: First reported situations were chosen significantly more often for double-eventive descriptions ($\text{mean}_{\text{event-event}} = 0.83$, $\text{SD} = 0.38$) than for double-stative descriptions ($\text{mean}_{\text{state-state}} = 0.45$, $\text{SD} = 0.50$), which appeared as a significant difference between people's response choices depending on the eventuality type of both clauses (see **Figure 5.I**, $\beta = 0.37$, $t = 9.60$, $p < 0.001$).

Relatedly, for sentences in which eventuality types differed between clauses (olive and teal bars), participants chose the first reported situation first, but only when it was stative (olive bars, $\text{mean}_{\text{state-first}} = 0.90$, $\text{SD} = 0.31$).

“What happened first?”

Mean response choices and standard errors (N=130)

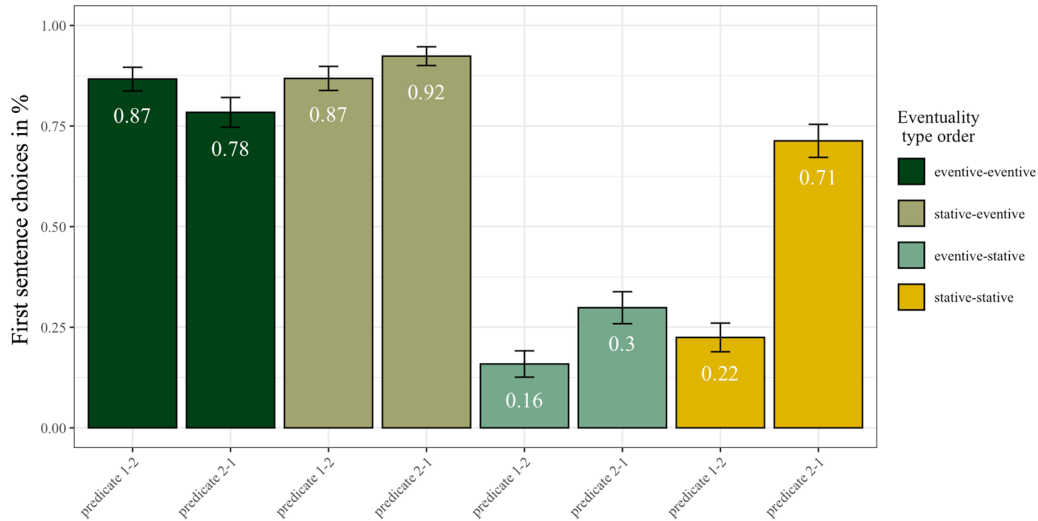


Figure 4: Mean proportions of first sentence choices across eight conditions, error bars represent standard errors. The x axis shows the reported order of the two situations.

For sentences in which the first situation was eventive (teal bars), participants chose the second situation to have happened first (mean_{event_first} = 0.23, SD = 0.42). This observation was statistically supported by a main effect of the eventuality in the first clause ($Df = 1, \chi^2 = 28.33, p < 0.001$), a significant interaction between eventuality type combination and the eventuality in the first clause ($Df = 1, \chi^2 = 296.98, p < 0.001$), as well as a significant difference between the stative-eventive and eventive-stative combinations, collapsed across reported orders (i.e., teal vs. olive bars, see **Figure 5.II**, $\beta = 0.67, t = 20.59, p < 0.001$).

Within reported orders, when the first reported situation was eventive, participants were less likely to order it first in time when it was followed by a stative description (predicate order 1-2: mean_{event-state} = 0.16, SD = 0.37; predicate order 2-1: mean_{event-state} = 0.30, SD = 0.46), than by an eventive description (predicate order 1-2: mean_{event-event} = 0.87, SD = 0.34, predicate order 2-1: mean_{event-event} = 0.78, SD = 0.41), which was also confirmed by pairwise comparisons (predicate order 1-2, see **Figure 5.III**: $\beta = -0.71, t = -16.15, p < 0.001$, predicate order 2-1, see **Figure 5.IV**: $\beta = 0.49, t = -8.92, p < 0.001$).

Unlike in Experiment 1, the interaction between reported order and eventuality type in the first clause reached significance ($Df = 1, \chi^2 = 13.02, p = 0.0003$), such that the reported order of the two sentences had a bigger effect on people’s response choices when stative situations were reported first (predicate order 1-2: mean_{stative_first} = 0.54, SD = 0.50 vs. predicate order 2-1: mean_{stative_first} = 0.82, SD = 0.38) than when eventive situations were reported first (predicate order 1-2: mean_{eventive_first} = 0.52, SD = 0.50 vs. predicate order 2-1: mean_{eventive_first} = 0.53, SD

= 0.50). That is, when states preceded events in reported order, participants were more likely to simply rely on this order of situations. Furthermore, in contrast to Experiment 1, we found a significant three-way interaction between the three predictors ($Df = 1$, $\chi^2 = 19.80$, $p < 0.001$).

As in Experiment 1, double-state descriptions differed in terms of their reported order: In Experiment 2, participants followed the reported order more frequently when the first reported situation was the relatively longer state (in the example above, *being on holiday*; $\text{mean}_{\text{predicate order } 2-1} = 0.71$, $SD = 0.45$) than when the first reported situation was the relatively shorter state (in the example above, *being on a train*; $\text{mean}_{\text{predicate order } 1-2} = 0.22$, $SD = 0.42$). This was statistically reflected in a main effect of reported order ($Df = 1$, $\chi^2 = 21.84$, $p < 0.001$) as well as in a significant difference between the double-stative conditions (see **Figure 5.V**, $\beta = -0.49$, $t = -9.02$, $p < 0.001$).

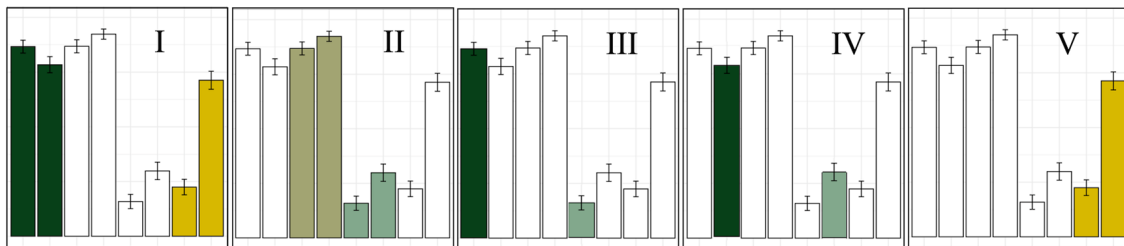


Figure 5: Outline of planned pairwise comparisons, colors of bars reflect compared conditions.

We had expected that differences in the duration of a state might play a role when inferring temporal order. Therefore, we preregistered selected additional analyses for the double-stative conditions to probe selected items, and performed pairwise comparisons only for these items, where we expected durational differences to be more pronounced: First, we replicated the findings from Experiment 1, such that participants followed the reported order when marriage was described first ($\text{mean}_{\text{marriage_first}} = 0.59$, $SD = 0.50$), but not when pregnancy was described first ($\text{mean}_{\text{pregnancy_first}} = 0.08$, $SD = 0.28$); this was reflected in a significant difference for the double-stative conditions in marriage-pregnancy sentences ($\beta = -0.51$, $t = -3.27$, $p = 0.003$): Similarly, a person being on holiday and sitting on a train, reported in this order, led to significantly more first sentence first choices ($\text{mean}_{\text{holiday_first}} = 0.94$, $SD = 0.24$) than when sentences reported the train sitting first ($\text{mean}_{\text{train_first}} = 0.08$, $SD = 0.28$, $\beta = 0.86$, $t = 9.09$, $p < 0.001$). Visual inspection of the data, broken down by item (**Figure 6**), further revealed that this duration-driven asymmetry between double-stative descriptions was stable across items: For sentence pairs in which the longer state was reported first (i.e., predicate order 2-1), the first reported situation was numerically selected more often to have happened first than for sentence pairs in which the shorter state was reported first.

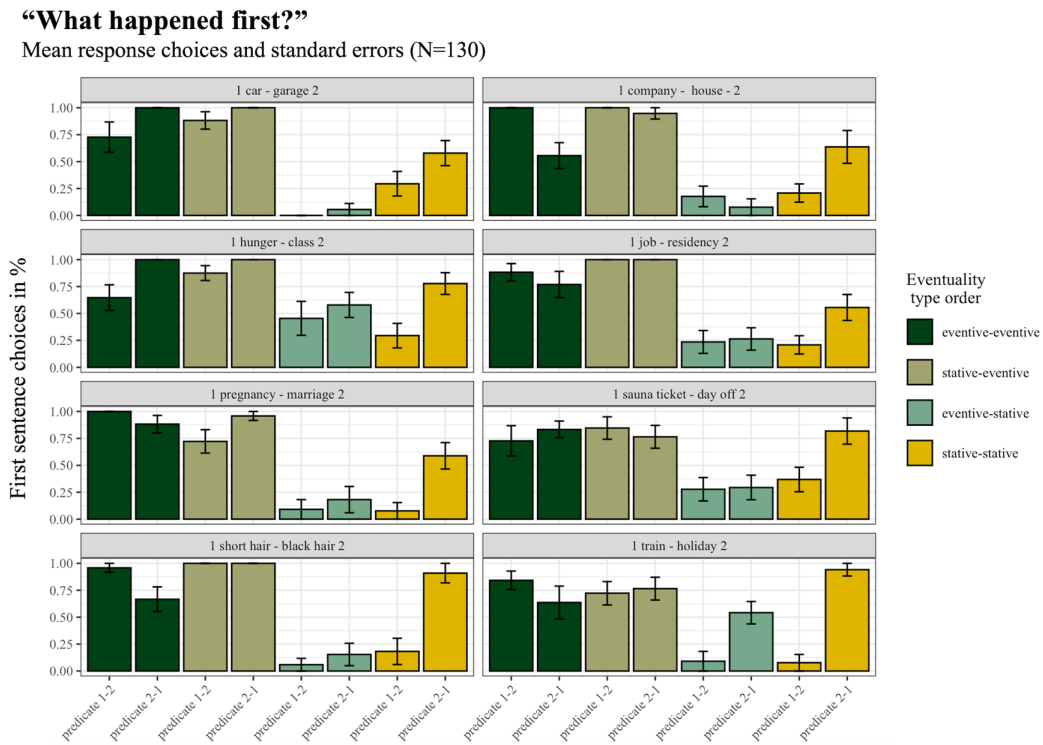


Figure 6: Mean proportions of first sentence choices across eight conditions item-by-item, error bars represent standard errors. The x axis shows the reported order of the two situations.

As in Experiment 1, the only interaction that did not reach significance was between reported order and eventuality type combination ($Df = 1, \chi^2 = 0.50, p = 0.48$). The main effect of event type combination did also not reach significance ($Df = 1, \chi^2 = 2.49, p = 0.11$).

3.3 Discussion of Experiment 2

Experiment 2 served as a conceptual replication of the first study; in addition, it broadened our empirical base by using a within-participants design. We replicated the pattern obtained in Experiment 1. That is, while participants iconically mapped reported order to temporal order in double-eventive sequences, this tendency was completely suppressed as soon as one of the events was encoded as a state: then, states were more likely to be ordered before events. In other words, eventuality type reliably predicted temporal order inferences in discourse comprehension.

Furthermore, the within-subjects design allowed us to account for noise through including items and participants as random intercepts, which increased the power of our statistical analysis. For instance, we found that the predicted main effect of event type in the first clause as well as the three-way interaction between reported order, eventuality in the first clause, and eventuality combination were significant in Experiment 2, but not in Experiment 1. As we have shown in

Figure 6, there was limited variability in items that could have contributed to these effects. Therefore, Experiment 2 supports the generalizability of the conclusion that non-linguistic event structure is an important contributor to temporal inferences in discourse.

Experiment 2 also allowed us to touch upon a question that is related to the question of what makes a state a state, and an event an event: the question of temporal duration. We return to this point in Section 4.

4. Discussion

In this paper, we asked whether linguistic eventuality type (i.e., stative vs. eventive descriptions) modulates the way in which people understand the temporal order between two situations reported in two independent sentences. More precisely, does a preference for ordering stative descriptions first override the widely reported observation that, in discourse comprehension, people take the reported order of a narrative to map iconically to temporal inferences?

In two experiments, we found that temporal inferences are reliably predicted by linguistic eventuality: When participants read two sentences containing eventive descriptions, they followed the reported order of the two sentences and inferred that the first reported situation also happened first in time. In this respect, our data confirms a classic observation in the literature on temporal inference in discourse, according to which the reported order of successive sentences is iconically mapped onto the temporal order of the described situations (Fleischman, 1990; Hinrichs, 1986; Oversteegen, 2005; Webber, 1988). However, when situation descriptions differed in their eventuality type, the reported order did not serve as a cue to temporal order. Instead, participants were more likely to locate stative descriptions before eventive descriptions, regardless of which of the two situations they read first (3–6).

These results suggest a systematic correspondence between linguistic eventuality type and the inferred temporal order between situations, confirming previous findings from other linguistic contexts (see Marx & Wittenberg, 2022, for relative clause constructions). Therefore, our results lend strength to event-structural accounts that link temporal inference in language comprehension to non-linguistic event cognition.

According to such accounts, deriving the temporal structure of a narrative requires that each described situation is temporally anchored by another time or situation. While multiple factors can in principle influence temporal anchoring relations (i.e., causality or explicit temporal marking), event structure – that is, the fact that states are stable and thus cognitively less salient than events (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007) – seems to be a robust determinant for constructing temporal relations in language.

When both situations are encoded with stative descriptions, we found in Experiment 1, that people tended to order marriage before pregnancy, especially when marriage was linguistically

reported second. While this general preference for ordering marriage first may originate from a contingency asymmetry between the two situations in world knowledge, with people perhaps viewing pregnancy as more contingent on marriage than vice versa, the fact that this preference is more pronounced when it is opposite to the reported order may be explained in two ways: First, participants may have paid more attention to the temporal order of the two situations because of the markedness of the reported order. That is, people's expectation that marriage should occur first may have conflicted with their expectation that the reported order should reflect the order in which the situations occurred in the world (Behaghel, 1932; Tai, 1983).

However, there is also another possible explanation for the temporal inferences found in stative-stative conditions, which is independent of people's expectations about the contingency relations between pregnancy and marriage. Experiment 2 allowed us to further explore the possibility that this effect is driven by other dimensions of event structure: the estimated duration of states. We tentatively predicted that longer states should precede shorter states: If duration is a relevant distinction in event structure, shorter states should tend to pattern with events. This was the case in our data but has also been shown in linguistic tests.

Linguistic tests of stativity (Dowty, 1979; Katz, 2003; Lakoff, 1966) tend to vary in reliability. In general, they are quite robust for long-lived individual-level predicates (e.g., *to be French*), but short-lived stage-level predicates (e.g., *to be cooperative*) often resist. For instance, states are claimed to be incompatible with progressive tense, which is true for individual-level predicates (e.g., *?I am being French*), but not for stage-level predicates (e.g., *I am being cooperative*), which pattern with events (e.g., *I am singing*). Second, imperative morphology is marked with stative meaning (e.g., *??Be French!*, but see *Be cooperative!* or *Sing!*). Also, embedding states under *must* should elicit an epistemic interpretation, which we find in individual-level predicates (e.g. *You must be French*); however, stage-level predicates receive a deontic, or at least ambiguous, interpretation (e.g., *You must be cooperative*) similar to events (e.g., *You must sing*).

Now, while there are manifold differences between individual-level and stage-level predicates (Carlson, 1977; Kratzer, 1995), for instance, that one has limited agency over being French, but full agency over being cooperative or singing, one very salient recurring difference between these different stative predicates is relevant to our project: time. Being French in most cases lasts from crib to grave, whereas being cooperative rarely does.

Therefore, we hypothesized that the estimated duration of a state could influence temporal order inferences in double-stative conditions; and we have argued that this explained the inferred order in these conditions in Experiment 1. Indeed, in Experiment 2, we found more generally that longer states tended to be ordered before shorter states. While a more nuanced exploration of this effect is beyond the scope of this paper and it should be confirmed by future work, we would argue that these findings further support a theory of language processing in which

non-linguistic event knowledge reliably and predictably drives temporal inferences in discourse comprehension. In this sense, our study also touches on the more fundamental question of what makes a state a state, and an event an event in language and cognition.

Overall, these experiments are a first systematic attempt to empirically investigate the importance of linguistic eventuality in how people understand the temporal order of unrelated discourse entities. Our results suggest that whether we talk – and therefore likely think – about situations as static or dynamic is more relevant for our readers' inferences about what happened first than the linear order in which the situations are reported.

Data accessibility statement

The data and analysis scripts are publicly available on OSF: <https://osf.io/fnmq4> and <https://osf.io/5tz6r>.

Ethics and consent

The experimental research presented in this article was approved by the Psychological Research Ethics Board (PREBO) at the Central European University. The reference number of the approval is 2023/02. Participants gave informed consent.

Acknowledgements

We thank Attila Balla for his support in data collection, Edward Matthew Husband and the three reviewers for their critical comments, and the audiences at AMLaP 2023 and the University of Stuttgart for their helpful feedback.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Conceptualization, methodology, visualization, writing – original draft, writing – review & editing: EM, EW; Statistical analysis: EM (lead).

ORCID IDs

EM [0000-0002-2199-3625](https://orcid.org/0000-0002-2199-3625)

EW [0000-0002-3188-6145](https://orcid.org/0000-0002-3188-6145)

References

- Arehalli, S., & Wittenberg, E. (2021). Experimental filler design influences error correction rates in a word restoration paradigm. *Linguistics Vanguard*, 7(1), 1–15. DOI: <https://doi.org/10.1515/lingvan-2020-0052>
- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, 9(1), 5–16. DOI: <https://doi.org/10.1007/BF00627432>
- Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2018). Dynamic adjustment of lexical processing in the lexical decision task: Cross-trial sequence effects. *Quarterly Journal of Experimental Psychology*, 71(1), 37–45. DOI: <https://doi.org/10.1080/17470218.2016.1240814>
- Behaghel, O. (1932). *Deutsche Syntax. Eine geschichtliche Darstellung*. Heidelberg: Carl Winters.

- Briner, S. W., Virtue, S., & Kurby, C. A. (2011). Processing causality in narrative events: Temporal order matters. *Discourse Processes*, *49*(1), 61–77. DOI: <https://doi.org/10.1080/0163853X.2011.607952>
- Carlson, G. (1977). A unified analysis of the English bare plural. *Linguistics and Philosophy*, *1*(3), 413–457. DOI: <https://doi.org/10.1007/BF00353456>
- Carroll, M., & von Stutterheim, C. (2010). Event representation, time event relations, and clause structure: A crosslinguistic study of English and German. In J. Bohnemeyer & E. Pederson (Eds.), *Event representation in language and cognition* (pp. 68–83). Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511782039.004>
- Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, *11*(1), 1–14. DOI: <https://doi.org/10.1038/s41467-020-17851-9>
- Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. SAGE Publications.
- Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS ONE*, *11*(1), 1–29. DOI: <https://doi.org/10.1371/journal.pone.0146194>
- Downes, J. J., Mayes, A. R., MacDonald, C., & Hunkin, N. M. (2002). Temporal order memory in patients with Korsakoff's syndrome and medial temporal amnesia. *Neuropsychologia*, *40*(7), 853–861. DOI: [https://doi.org/10.1016/S0028-3932\(01\)00172-5](https://doi.org/10.1016/S0028-3932(01)00172-5)
- Dowty, D. R. (1979). *Word meaning and Montague Grammar: The semantics of verbs and times in Generative Semantics and in Montague's PTQ*. Springer Dordrecht. DOI: <https://doi.org/10.1007/978-94-009-9473-7>
- Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? *Linguistics and Philosophy*, *9*(1), 37–61. DOI: <https://doi.org/10.1007/BF00627434>
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10). DOI: <https://doi.org/10.1371/journal.pone.0077661>
- Fleischman, S. (1990). *Tense and narrativity: From medieval performance to modern fiction*. University of Texas Press. DOI: <https://doi.org/10.7560/780903>
- Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, *14*(6), 233–234. DOI: <https://doi.org/10.1016/j.tics.2010.03.005>
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. DOI: <https://doi.org/10.1037/0033-295X.101.3.371>
- Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, *110*, 70–104. DOI: <https://doi.org/10.1016/j.cogpsych.2019.01.001>
- Hinrichs, E. (1986). Temporal anaphora in discourses of English. *Linguistics and Philosophy*, *9*(1), 63–82. DOI: <https://doi.org/10.1007/BF00627435>

- Hopper, P. J. (1979). Aspect and foregrounding in discourse. In T. Givón (Ed.), *Discourse and syntax* (pp. 211–241). Brill. DOI: https://doi.org/10.1163/9789004368897_010
- Horn, L. (2022). Contrast and clausal order: Beyond Behaghel. *Language*, 98(4), 812–843. DOI: <https://doi.org/10.1353/lan.2022.0022>
- Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, 13(51), 21–37. DOI: <https://doi.org/10.1177/039219216501305103>
- Kameyama, M., Passonneau, R., & Poesio, M. (1993). Temporal centering. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 70–77. DOI: <https://doi.org/10.3115/981574.981584>
- Kamp, H., & Reyle, U. (1993). Tense and aspect. In H. Kamp & U. Reyle (Eds.), *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory* (pp. 483–689). Kluwer Academic Publishers. DOI: https://doi.org/10.1007/978-94-011-2066-1_6
- Kamp, H., & Rohrer, C. (1983). Tense in texts. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 250–269). De Gruyter. DOI: <https://doi.org/10.1515/9783110852820.250>
- Katz, G. (2003). On the stativity of the English perfect. In A. Alexiadou, M. Rathert, & A. von Stechow (Eds.), *Perfect explorations* (pp. 205–234). Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110902358.205>
- Kausler, D. H., Salthouse, T. A., & Sauls, J. S. (1988). Temporal memory over the adult lifespan. *The American Journal of Psychology*, 101(2), 207–215. DOI: <https://doi.org/10.2307/1422835>
- Kehler, A. (1994). Temporal relations: Reference or discourse coherence? *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 319–321. DOI: <https://doi.org/10.3115/981732.981779>
- Kehler, A. (2002). *Coherence, reference and the theory of grammar*. CSLI Publications.
- Kehler, A. (2006). Discourse coherence. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 241–265). Wiley-Blackwell. DOI: <https://doi.org/10.1002/9780470756959.ch11>
- Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. (Doctoral dissertation). DOI: <https://doi.org/10.7282/T3GQ6WMS>
- Klein, W. (1994). *Time in language*. Routledge.
- Klein, W. (2000). An analysis of the German perfekt. *Language*, 76(2), 358–382. DOI: <https://doi.org/10.1353/lan.2000.0140>
- Klein, W. (2009). How time is encoded. In W. Klein & P. Li (Eds.), *The expression of time* (pp. 39–82). De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110199031.39>
- Kratzer, A. (1995). Stage-level and Individual-level predicates. In G. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 125–175). University of Chicago Press.
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79. DOI: <https://doi.org/10.1016/j.tics.2007.11.004>

- Lakoff, G. (1966). Stative adjectives and verbs in English. *Harvard Computational Laboratory Report NSF-17*.
- Lascarides, A., & Asher, N. (1991). Discourse relations and defeasible knowledge. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, (1), 55–62. DOI: <https://doi.org/10.3115/981344.981352>
- Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16, 437–493. DOI: <https://doi.org/10.1007/BF00986208>
- Laurinavichyute, A., & von der Malsburg, T. (2023). *Agreement attraction in grammatical sentences and the role of the task*. 1–23. DOI: <https://doi.org/10.31234/osf.io/n75vc>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.
- Marx, E., & Wittenberg, E. (2022). Event structure predicts temporal interpretation of English and German past-under-past relative clauses. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 439–445.
- Morgan, A. M., von der Malsburg, T., Ferreira, F., & Wittenberg, E. (2020). Shared syntax between comprehension and production: Multi-paradigm evidence that resumptive pronouns hinder comprehension. *Cognition*, 205, 104417. DOI: <https://doi.org/10.1016/j.cognition.2020.104417>
- Ness, T., & Meltzer-Asscher, A. (2021). Rational adaptation in lexical prediction: The influence of prediction strength. *Frontiers in Psychology*, 12, 622873. DOI: <https://doi.org/10.3389/fpsyg.2021.622873>
- Oversteegen, L. (2005). Causality and tense – Two temporal structure builders. *Journal of Semantics*, 22(3), 307–337. DOI: <https://doi.org/10.1093/jos/ffh021>
- Partee, B. H. (1973). Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy*, 70(18), 601–609. DOI: <https://doi.org/10.2307/2025024>
- Partee, B. H. (1984). Nominal and temporal anaphora. *Linguistics and Philosophy*, 7(3), 243–286. DOI: <https://doi.org/10.1007/BF00627707>
- Pregla, D., Lissón, P., Vasishth, S., Burchert, F., & Stadie, N. (2021). Variability in sentence comprehension in aphasia in German. *Brain and Language*, 222, 105008. DOI: <https://doi.org/10.1016/j.bandl.2021.105008>
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Sacks, O. (1985). The lost mariner. In O. Sacks (Ed.), *The man who mistook his wife for a hat and other clinical tales*. Summit Books.
- Schmerling, S. F. (1975). Asymmetric conjunction and rules of conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 211–231). DOI: https://doi.org/10.1163/9789004368811_009
- Schütze, C., & Sprouse, J. (2013). Judgment data. In D. Podesva & R. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). DOI: <https://doi.org/10.1017/CBO9781139013734.004>
- Strawson, P. F. (1952). *Introduction to logical theory*. Methuen.

- Tai, J. H. Y. (1983). Temporal sequence and Chinese word order. *Iconicity in syntax: Proceedings of a Symposium on Iconicity in Syntax*. John Benjamins Publishing Company.
- Vendler, Z. (1967). Verbs and times. In Z. Vendler (Ed.), *Linguistics and Philosophy* (pp. 97–121). Cornell University Press. DOI: <https://doi.org/10.7591/9781501743726-005>
- von der Malsburg, T., Poppels, T., & Levy, R. P. (2020). Implicit gender bias in linguistic descriptions for expected events: The cases of the 2016 United States and 2017 United Kingdom elections. *Psychological Science*, 31(2), 115–128. DOI: <https://doi.org/10.1177/0956797619890619>
- von Stutterheim, C., Carroll, M., & Klein, W. (2003). Two ways of construing complex temporal structures. In F. Lenz (Ed.), *Deictic conceptualisation of time, space and person* (pp. 97–133). John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/pbns.112.07stu>
- Wårvik, B. (2004). What is foregrounded in narratives? Hypotheses for the cognitive basis of foregrounding. In T. Virtanen (Ed.), *Approaches to cognition through text and discourse* (pp. 99–148). De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110892895.99>
- Webber, B. L. (1988). Tense as discourse anaphor. *Technical Reports (CIS)*, 441.
- Wolfe, M. B. W., Magliano, J. P., & Larsen, B. (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*, 39(2–3), 165–187. DOI: <https://doi.org/10.1080/0163853X.2005.9651678>
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, 133(2), 273–293. DOI: <https://doi.org/10.1037/0033-2909.133.2.273>

