



Interaction of Script Knowledge and Temporal Discourse Cues in a Visual World Study

Alessandra Zarcone & Vera Demberg

To cite this article: Alessandra Zarcone & Vera Demberg (2021) Interaction of Script Knowledge and Temporal Discourse Cues in a Visual World Study, *Discourse Processes*, 58:9, 804-819, DOI: [10.1080/0163853X.2021.1930807](https://doi.org/10.1080/0163853X.2021.1930807)

To link to this article: <https://doi.org/10.1080/0163853X.2021.1930807>



© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 26 Jul 2021.



Submit your article to this journal [↗](#)



Article views: 834



View related articles [↗](#)



View Crossmark data [↗](#)

Interaction of Script Knowledge and Temporal Discourse Cues in a Visual World Study

Alessandra Zarcone and Vera Demberg

HumAln-IIS - Audio und Medientechnologien Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany;
Department of Language Science and Technology, Universität des Saarlandes, Saarbrücken, Germany

ABSTRACT

There is now a well-established literature showing that people anticipate upcoming concepts and words during language processing. Commonsense knowledge about typical event sequences and verbal selectional preferences can contribute to anticipating what will be mentioned next. We here investigate how temporal discourse connectives (*before*, *after*), which signal event ordering along a temporal dimension, modulate predictions for upcoming discourse referents. Our study analyses anticipatory gaze in the visual world and supports the idea that script knowledge, temporal connectives (*before eating* → *menu*, *appetizer*), and the verb's selectional preferences (*order* → *appetizer*) jointly contribute to shaping rapid prediction of event participants.

Introduction

During text understanding, comprehenders need to process the linguistic signal and integrate it with prior world knowledge to construct a mental model of the situation. The situation model then allows listeners to represent the content that is being conveyed and even to anticipate upcoming linguistic material (Altmann & Kamide, 1999; McRae & Matsuki, 2009). Discourse connectives can thereby provide an informative signal for how to combine different discourse segments. Previous work has shown that discourse connectives (causals vs. concessives, e.g., *therefore* vs. *nevertheless*) are processed incrementally and are quickly integrated with contextual information (Köhne & Demberg, 2013; Kuperberg et al., 2011; Münte et al., 1998; Traxler et al., 1997), but connectives may differ in terms of the processing difficulty they elicit for the comprehender. Concessive connectives have been shown to be harder to process than causals (Drenhaus et al., 2014; Xiang & Kuperberg, 2015), as they require the revision of the default event structure. In this article, we focus on the processing of *temporal* discourse connectives like *before* versus *after*, which signal in what order events unfold in time and help comprehenders map events along a temporal dimension (Clark & Clark, 1968; Mandler, 1986; Münte et al., 1998).

Besides temporal connectives, knowledge of everyday activities may also contribute to mapping events along a temporal dimension and to building expectations that are coherent with the discourse representation. Imagine hearing a story such as the one in (1):

- (1) Peter went for dinner to his favorite restaurant. He sat down at a free table right by the window. After his meal, he ordered . . .

Before hearing the verb (*ordered*), comprehenders may already be in a position to construct some hypotheses about what event happened next, for instance, that Peter *asked for the bill* or *ordered*

CONTACT Alessandra Zarcone  a.zarcone@gmail.com  HumAln-IIS - Audio und Medientechnologien, Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, Erlangen 91058, Germany

 Supplemental data for this article can be accessed on the [publisher's website](#).

© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

something else, as cued by the combination of the temporal connective *after* and of our knowledge of typical events in the restaurant script. After the verb, the search space of possible expectations could be further reduced to entities that can be *ordered* (*desserts, coffees, etc.*) and the predicting comprehender could expect that Peter ordered a *dessert*, or a *coffee*, or maybe an after-dinner *digestif*, but probably not an *appetizer* or an *aperitif*.

The goal of our study is to investigate how our knowledge of typical events in common routines (*script knowledge*, e.g., *getting a haircut, going to a restaurant*; Nieuwland, 2015; Van der Meer et al., 2002, 2005) is cued by the interplay between temporal discourse cues and lexical cues (the verb's selectional preferences) and how early such cues are integrated during predictive processing, driving expectations for upcoming discourse referents.

Explicit and implicit temporal relations in discourse

In comprehending two sentences presented one after the other, people may infer an implicit relation between them (e.g., a temporal relation, e.g., *Max stood up. John greeted him*, or a causal relation, e.g., *John pushed Max. He fell*, Lascarides & Asher, 1993; see also Murray, 1997; Sanders, 2005). According to the Iconicity Assumption, comprehenders assume that events are usually reported following their chronological order (e.g., *Max stood up. John greeted him*; Dowty, 1986; Fleischman, 1990; Zwaan et al., 2001), unless signaled otherwise.

This signaling role is played by explicit temporal connectives, which establish relationships between events and thus contribute to discourse coherence (Halliday & Hasan, 1976; Van den Broek & Helder, 2017; Zwaan & Radvansky, 1998). Explicit discourse connectives (e.g., *before, after, because, however*) have been shown to influence processing (Haberlandt, 1982; Sanders & Noordman, 2000), increase coherence (Britton et al., 1982; Halliday & Hasan, 1976), and affect the activation of concepts and their retrieval (Caron et al., 1988; Millis & Just, 1994). Connectives are rapidly integrated during online comprehension (Drenhaus et al., 2014; Köhne & Demberg, 2013; Traxler et al., 1997) and act as cues to indicate what information is relevant and what is not while comprehenders integrate upcoming information, shift their attention focus to different parts of the situation, and build a coherent mental representation of the discourse (Gernsbacher & Givón, 1995; Zwaan & Rapp, 2006). Temporal connectives in particular (e.g., *before / after*) provide specific cues regarding the temporal relations between clauses and sentences, contributing to the temporal structure of the discourse representation. When the order of presentation does not correspond to the chronological order of the events, temporal connectives may signal reordering (e.g., *Before she left the house, she grabbed her umbrella*).

Furthermore, several studies (Clark & Clark, 1968; Mandler, 1986; Münte et al., 1998; Ye, Kutas, et al., 2012; Politzer-Ahles et al., 2017) have shown evidence in support of the idea that an iconic order of mention of events (i.e., an order of mention matching the order in which they occurred in the world) facilitates processing. The Iconicity Assumption suggests that “deviations from chronological order are relatively difficult—but not impossible—to process because a default assumption has to be overridden” (Zwaan et al., 2001, p. 79). The semantic markedness of the inverse chronological order is supported by experimental results that ascribe higher processing difficulties to constructions with inverse chronological order (compared to the iconic order) in healthy adults (Baker, 1978; Clark & Clark, 1968), in aphasics (Sasanuma & Kamio, 1976), in Parkinson patients (Natsopoulos et al., 1991), and in children (Amidon & Carey, 1972; Mandler, 1986; Trosborg, 1982). Additionally, Münte et al. (1998) found that event-related potential responses to *before*-initial sentences (which present events out of chronological order) elicit a larger prolonged left frontal negativity 300 ms after the onset of the connective compared to *after*-initial sentences. They ascribed this to a higher working memory load, as the difference is more pronounced for individuals with a higher working memory span, and argue that *before*-initial sentences require additional discourse-level computations to rearrange the events to match chronological order (see also Ye, Kutas, et al., 2012; Ye, Milenkova, et al., 2012; Politzer-Ahles

et al., 2017), compared to sentences when the order of mention matches the chronological order of the events.

Everyday events, their participants, and their order

Besides discourse cues, people can also use knowledge of everyday activities (e.g., *going to a restaurant*) to build a discourse representation and map events along a temporal dimension. If we hear a story about a restaurant visit, we expect that *ordering food* happens after *reading the menu*, and if the speaker omits a typical event such as *sitting down at the table*, we can easily infer that it did take place anyway. Such events and their participants are commonly referred to as *script knowledge* (Schank & Abelson, 1977).

The typical order of everyday events is arguably part of our representation of such events. Human cognition is typically future-oriented in a broader sense: Observers simulate and predict an object's future trajectory (Freyd, 1987) as well as future states in complex movements (Güldenpenning et al., 2012), and when recalling the end phase of a motion event, people have a tendency to recall a phase that is actually beyond its actual ending point (Freyd & Finke, 1984; see also Hubbard, 2005). Work on event sequences (Chwilla & Kolk, 2005; Khalkhali et al., 2012; Van der Meer et al., 2002, 2005) has shown that the mention of earlier script events in chronological order cues later events (e.g., *cook, sit* → *dine*). These results point toward a temporally oriented, order-sensitive organization of script events and their participants in long-term memory, which influences expectation-based processing, arguably interacting with linguistic cues.

Interplay of discourse cues and temporally ordered script knowledge

An interesting open question is how order-sensitive knowledge such as script knowledge interacts with contextual cues. Temporal order information, cued by discourse connectives, has also been shown to influence the comprehension of script event sequences (e.g., *after bite off – chew* vs. *before digest – swallow*): In a relatedness judgment task, pairs presented in chronological order (the *after* pairs) were recognized as related (i.e. as part of the same script) faster than pairs presented in the inverse order (the *before* pairs; Van der Meer et al., 2002, 2005).

However, the temporal orientation of script knowledge may also overcome the assumption that the order of mention corresponds to event order (the Iconicity Assumption). Mandler (1986) suggested that the Iconicity Assumption may primarily hold for arbitrarily connected events and not for causally related events, because temporal order in causally related events already exists and does not have to be constructed. In Mandler's (1986) self-paced reading experiments, the processing of noniconic order was found to be facilitated in cases where there was a relationship between the events, compared to arbitrarily connected events. The same may be true for script knowledge, which already encodes the order between events (e.g., in the restaurant script, *reading the menu* typically comes before *eating a meal*), and hence less cognitive effort may be required for constructing and reordering events, as script knowledge would potentially neutralize any effect of iconicity.

Considering previous results on future-oriented script knowledge and on the nonmarkedness of iconic order, we can expect that people will build expectations as to how a story continues, anticipating upcoming events and their participants, *in particular when* the order of mention matches their chronological order of the events (i.e., the upcoming event is both the next event to be mentioned *and* the next event in chronological order). Alternatively, the strong temporal relations between script events may facilitate the role played by temporal connectives when reordering is signaled, thus making it easier to build early expectations for upcoming events and their participants *even when* the order of mention does not match the chronological order (i.e., the upcoming event is the next event to be mentioned *but not* the next event in chronological order). If, on the other hand, comprehenders are strongly influenced by script knowledge only, then we can expect their predictions to be geared toward the events and participants that *come next* in the temporal sequence that is encoded in the script

representation, *regardless of* discourse cues that may signal reordering or lexical cues indicating the event which is more likely to be *mentioned next*.

Current study

The present study addresses the interplay between script knowledge and temporal connectives in naturalistic language processing. We manipulate the temporal connective in the context of order-sensitive script knowledge to investigate the modulation of comprehender expectations with respect to upcoming events. We measure such expectations using the visual world paradigm, a well-established paradigm for studying anticipatory processing (Altmann & Kamide, 1999; Huettig et al., 2011), which has also been used in previous studies investigating anticipatory gaze in the context of the processing of causal or concessive connectives (Köhne & Demberg, 2013). In our experiment, we depict participants of different script events and observe comprehender's gaze to these event participants to infer which script event a comprehender anticipated based on the combination of contextual cues (script, connective, and local context). Consider the example in (2):

- (2) Peter went to his favorite restaurant for dinner. He sat down at a free table right by the window.
- (a) After his meal, he ordered . . .
 - (b) Before his meal, he ordered . . .

The first sentence in (2) has the function of triggering the script knowledge. The connective then signals an upcoming reference to a script-related event, which is placed in a specific temporal relation (*after* or *before*) with a reference event (the *meal*). Finally, the verb (*ordered*) constrains the types of referents that can be expected to be mentioned next.

Our hypothesis is that scripts are organized in long-term memory as temporally ordered structures and that people make an early use not only of broad world knowledge information about (temporally oriented) scripts but also of discourse cues as soon as they become available, seamlessly integrating cues to build expectations about what will be mentioned next. In (2a), we expect the discourse connective *after*, in conjunction with the script knowledge and the central event (*his meal*), to elicit predictions for events that would happen next in the script (e.g., *reading the bill*, *ordering dessert*). At the region following the verb (which constitutes the critical region in our experiment), these predictions may then be quickly integrated with the verb (*ordered*), such that the comprehender may be able to anticipate the discourse participant that matches both the cued order and the verb's selectional preference (*after meal* + *ordering* → *dessert*) before this referent is mentioned. A similar integration of discourse, verb, and script cues should also hold for sentences such as (2b), where the connective *before* signals that a reordering is required (with comprehenders anticipating events such as *reading the menu* or *ordering drinks*, *ordering an appetizer*). The combination of the verb's selectional preference with these event anticipations could then elicit anticipatory looks to objects like an *appetizer*, a *drink*, or an *aperitif*. We expect this to happen for both connectives, *even when* the order of mention does not match the chronological order. Alternatively, iconicity may also play a more prominent role on expectation-based processing, leading to stronger predictions for the target for *after*-sentences such as (2a), where no reordering is required, compared to *before*-sentences such as (2b), which require reordering. Finally, if comprehenders are strongly influenced by script knowledge only, that is by the encoded temporal relations between script events (independent of the way they are presented), this may lead to more anticipatory gazes to the next event *encoded* in the script (and the corresponding referent), no matter what target referent the contextual cues (discourse connectives or a verb's selectional preference) point to as the most likely to be mentioned next.



Figure 1. Visual scene for sentences in (3).

Methods

Materials

Experimental items consisted of a static scene and a short context followed by a target sentence (see Example 3, below, and Figure 1). To study the interplay between temporal discourse cues, script knowledge and local linguistic cues such as the fit of a referent with a verb, each sentence could appear in one of four different variations, which differed with regard to the temporal discourse connective (*before / after*) and the verb used in combination with the discourse referent (two verbs with different selectional preferences). An exemplary item with the four variations is shown in (3a-d).

- (3) Florian machte am Wochenende einen Großeinkauf. Im Supermarkt arbeitete er seine lange Einkaufsliste nach und nach ab.

Florian went for a big grocery shopping trip during the weekend. In the supermarket he worked through the list, item after item.

- (a) Vor dem [Einpacken der Einkäufe]_{pre-target} [fuhr er routiniert]_{target} [den Einkaufswagen]_{target+1} an die Kasse
Before [bagging the groceries]_{pre-target} [drove he experiencedly]_{target} [the shopping cart]_{target+1} to the cashier.
- (b) Nach dem [Einpacken der Einkäufe]_{pre-target} [fuhr er routiniert]_{target} [das Auto]_{target+1} vom Parkplatz
After [bagging the groceries]_{pre-target} [drove he experiencedly]_{target} [the car]_{target+1} from the parking lot.
- (c) Vor dem [Einpacken der Einkäufe]_{pre-target} [steckte er routiniert]_{target} [die EC-Karte]_{target+1} in das Kartenlesegerät
Before [bagging the groceries]_{pre-target} [inserted he experiencedly]_{target} [the debit card]_{target+1} in the card-reader

(d) Nach dem [Einpacken der Einkäufe]_{pre-target} [steckte er routiniert]_{target}
[die Schlüssel]_{target+1} ins Auto.

After [bagging the groceries]_{pre-target} [inserted he experiencedly]_{target} [the key]_{target+1} in the car.

The short context introduces a scenario (e.g., *grocery shopping*), the target sentence begins either with an *after* or a *before* temporal discourse connective, followed by a main script event (*bagging the groceries*), and by a region including a subject, a verb, and an adverb (*he drove experiencedly*). After this region, the discourse referent is mentioned, which is always congruent with the preceding discourse (a: before + drive: *the shopping cart*; b: after + drive: *the car*; c: before + insert: *debit card*; d: after + insert: *key*).

The visual worlds in the static scenes included a total of six objects, corresponding to the Areas of Interest (AOIs) (see Figure 1). Two of them were script-congruent fillers, which did not match the verb or were participants in an obvious *before*- or *after*-event (here a *clock* and a *cashier divider*). The other four were a target discourse referent, a temporal-order distractor, one selectional-preference distractor, and one unrelated distractor. For example, for the sentence variation 3a (*before + drive*), the *shopping cart* is the target object, matching both the expectations from the temporal connective (*connective match*) and from the verb's selectional preference (*verb match*); the temporal-order distractor is the *debit card*, which holds the same temporal order relation (*before* in this case) with the main event as the target (*connective match*, *verb mismatch*); the car is the selectional-preference distractor, matching the same main verb as the target but holding a different temporal order relation with the main event (*connective mismatch*, *verb match*); and the *key* is the unrelated distractor (*connective mismatch*, *verb mismatch*).

As we used four different sentence variations for each sentence, the experiment was fully counter-balanced. Each scene could be accompanied by one of four sentence variations, determining which one (out of the four objects, excluding the fillers) would have been the target. Each of the four possible target objects was in turn the target, the temporal-order distractor, the selectional-preference distractor, or the unrelated distractor, depending on the combination of *before* or *after* discourse cues and main verbs used in the target sentences. Before the main verb was mentioned, there were always two referents matching the temporal connective (the target and the temporal-order distractor) and two that did not match the temporal connective (the selectional-preference distractor and the unrelated distractor). After the verb, there were always two referents matching the verb (the target and the selectional-preference distractor) and only one that matched both the connective and the verb (the target).

Norming studies

We initially constructed 37 experimental items with four variations (148 sentences), which we tested in a sentence completion study (Norming Study 1) on the crowdsourcing platform Prolific (<https://prolific.ac/>) to evaluate the predictability of the discourse referent in context. The sentences were presented up to the verb region, excluding the discourse referent. On average, 11 participants saw each sentence (range: 9–13). For 10 items, where the target referent was not the most elicited one in the cloze completion study, the discourse referent was changed to match the most frequently elicited one. Six items for which no preference for a specific target referent could be found in one of the four variations were discarded.

A total of 31 items (with 4 objects each, that is 124 objects) were selected after Norming Study 1 for Norming Study 2. For each target object, we selected three pictures depicting it from freely available sets, namely the Bank of Standardized Stimuli (Brodeur et al., 2010), the HatField image test (Adlington et al., 2009), the datasets by Moreno-Martínez & Montor (2012) and Van der Linden et al. (2015), and Google Image searches for pictures labeled for noncommercial reuse. We collected picture-naming data for the 372 pictures depicting the target objects. The pictures were divided into three lists, so that the same participant would only see one of three pictures depicting the same target object, and each was presented together with the script trigger sentences to provide a context comparable to the experiment context (e.g., *Florian machte am Wochenende einen Großeinkauf. Im*

Table 1. Results from the Norming Studies for the Final Set of 28 Experimental Items

	<i>After-Objects</i>	<i>Before-Objects</i>	Average
Cloze completion match (Norming Study 1)	36%	34%	35%
Picture naming match (Norming Study 2)	82%	80%	81%

Supermarkt arbeitete er seine lange Einkaufsliste nach und nach ab. “Florian went for a big grocery shopping trip during the weekend. In the supermarket he worked through the list, item after item.”). Each picture with corresponding sentence was named by 11 participants on average (range: 9–14). After Norming Study 2, we excluded 3 more items because the target items could not be named with sufficient reliability and we selected for each object the picture with the best match between elicited nouns and nouns in the experimental sentences.

Table 1 shows the results of the two norming studies for the final set of 28 items selected for the visual world experiment (28×4 sentences + 28×4 object pictures). *Before-* and *after-*sentences were matched both in terms of average cloze completions (36% for *after-*sentences and 34% for *before-*sentences, n.s. difference; $t = 0.524$, $p > .05$) and in terms of picture naming (82% of elicited nouns for the pictures matched the ones in the sentences for *after-*objects and 80% for *before-*objects, n.s. difference; $t = 0.82$, $p > .05$). The final set of items can be found in the Supplementary Materials.

Participants

Thirty-two Universität des Saarlandes students initially participated in the experiment. We then replaced seven of them (five due to eye-tracking problems and two more for their low accuracy scores in the comprehension questions). All participants (the original 32 plus the 7 replacement participants) were native German speakers, had normal or corrected-to-normal visual acuity, and were paid to participate in the experiments. Data from the final set of 32 participants (22 women; age: $M_{\text{age}} = 25$ years; range: 19–34) were analyzed.

Equipment

Visual world scenes were presented on a Dell E2209W LCD monitor (size of the lit-up portion of the screen 47.5×29.5 cm). Eye movement data were acquired and recorded via a SR-Research Eye-Link 1000 Plus system with desktop mount, head support, and a 16-mm lens and with a sampling rate of 250 Hz and a nine-point calibration. Participants sat at a distance of about 83 cm from the screen with their eyes at level with the bottom 75% of the monitor height.

Procedure

Four lists of 84 sentences (each list contained 14 *after-*sentences, 14 *before-*sentences, and 56 fillers) were created using our final 28 items in their four sentence variants. Each list contained each item only in one variant. Each participant was assigned to one list and saw each scenario at most once, as different versions of the sentences sharing a scenario were assigned to different lists. The order of the items in the list was randomized and subsequently manually rearranged to make sure experimental items were separated by one to three fillers. For each list, half of the participants assigned to it were presented the list in reverse order. The fillers followed the same general pattern but used different discourse markers (e.g., *however*, *therefore*, *additionally*). Data from the fillers were not analyzed or reported elsewhere. Participants were instructed to look at the scenes and listen to the sentences to be able to respond to the comprehension questions. Before the presentation of the items in the lists, 3 additional items were presented as a practice session.

First, a fixation cross appeared in the middle of the screen for 500 ms. Then, the stories were played, and a visual world scene was presented simultaneously with the onset of each story. The stories were played via speakers and were read by a female voice; the speaker was trained as a professional speaker in standard German. The context sentences lasted on average 8 seconds and the target sentences 7 seconds on average for each condition.

After each story, the scene disappeared and was immediately followed by a yes-no comprehension question, which the participants answered by button press. Half of the questions' correct answer were "yes" the other half "no." Participants took breaks after each third of the items. The experiment lasted about 30 minutes.

Analysis

We analyzed fixations in three time regions: the pretarget region (i.e., the event noun following the connective *before / after*), the target region (including the main verb, the subject, and an adverb, where we expect the target referent to be predicted), and the post-target region, where the actual referent that is also displayed on the screen is mentioned:

- (4) *Nach dem Einpacken der Einkäufe_{pre-target} fuhr er routiniert_{target} das Auto_{post-target} vom Parkplatz*
After bagging the groceries_{pre-target} drove he experiencedly_{target} the car_{post-target} from the parking lot
 (transliteration)

For each of these regions, we chose an analysis window starting 200 ms after the onset of the first word in the region, as this is known as the amount of time needed to program a saccade. Each region ended when the next one began.

The dependent variable in our analyses was the number of fixations to the four AOIs for each trial and time region: the target (e.g., *car*), the temporal-order distractor (e.g., *key*), the selectional-preference distractor (e.g., *shopping cart*), and the unrelated distractor (e.g., *debit card*). For each time region, we analyzed four data points per trial, corresponding to the number of fixations to each of the four AOIs in the corresponding scene for the trial. As our dependent variable consists of number of fixations, we used generalized linear models for analyzing the three time regions. Our data follow a quasi-Poisson distribution (i.e., it is count data with a much heavier tail than the standard Poisson distribution). We found that the data can best be fit using a gamma distribution (with log-link); to the best of our knowledge, there are currently no reliable mixed effects model implementations for the quasi-Poisson distribution.

As predictors, we used the match with the discourse connective and the match with the verb's selectional preference. The match-predictors are used to test if a match with contextual cues (first the discourse connective, at the pretarget region, and then the connective and the verb, at the target region) influence the number of inspections that an AOI receives. Table 2 shows an example of all AOIs for two target sentences and how they map in terms of the match predictors.

Additionally, we coded the type of connective in the sentence as the iconicity predictor to distinguish between sentences where the discourse connective indicated that the events were

Table 2. Value of the Match Predictors for Each AOI

		Verb Match	
		Match	Mismatch
<i>After bagging the groceries, he drove...</i>			
Connective match	Match	Target (<i>the car</i>)	Temporal-order distractor (<i>the key</i>)
	Mismatch	Selectional-preference distractor (<i>the shopping cart</i>)	Unrelated distractor (<i>the debit card</i>)

mentioned in an iconic order (*after*-sentences, iconic order) and sentences where the connective indicated a reordering (*before*-sentences, noniconic order). We used the iconicity of the order of mention as a predictor to test if order iconicity had an effect in facilitating predictions to the target AOI (that is the AOI with connective match and verb match).

In our first analysis the AOIs were coded depending on their match with the connective and the verb's selectional preference. Additionally, we also recoded the AOIs in a different way to test the effect of future-oriented script knowledge: AOIs matching future events were recoded as *future-matching* and AOIs nonmatching future events as *non-future-matching*. This recoding is independent of what sentence variation was presented as context (specifically, independent of the iconic order cued by the connective and independent of the AOI match with the verb and with the connective) and only dependent on the temporal relation between the anchor event and the AOI (e.g., given the anchor event *bagging the groceries*, *car* and *key* were coded as past-oriented referents/AOIs and *shopping cart* and *debit card* were coded as future-oriented referents/AOIs). In our analyses using future-matching as a binary predictor, we did not use any of the match predictors described above.

Subject and item were entered as random factors. Main effects were tested based on model comparison using a χ^2 test (Baayen et al., 2008). We started with the maximal random effects supported by the design, as suggested by Barr et al. (2013), and followed their "best path" procedures whenever the maximal model did not converge (first by removing the correlation between random intercepts and random slopes, and then removing random intercepts or random slopes). For each region we report the results of the maximal converging model.

Additionally, we also performed a Bayes factor analysis using full Bayesian mixed linear models fitted using the R package 'rstanarm' (Goodrich et al., 2020). Bayesian inference was done using Markov Chain Monte Carlo sampling with 4 chains, each with iter = 2000; warmup = 1000; thin = 1; post-warmup = 1000. The model's priors were set as follows: \sim normal (location = (0, 0, 0, 0), scale = (2.50, 2.50, 2.50, 5)). For each model and each parameter, we report the Bayes factor expressed as BF₁₀, indicating the intensity of the evidence that the data provide for H₁ versus H₀, the median of the posterior distribution (a point estimate comparable to the beta from frequentist linear models); its median absolute deviation (MAD; a robust equivalent of standard deviation); the 90% credible interval (CI), representing a range of possible parameter values; the maximum probability of effect (MPE), that is the probability that the effect is positive or negative (depending on the median's direction); and the overlap, that is the percentage of overlap between the posterior distribution and a normal distribution of mean 0 and same standard deviation than the posterior.

We also ran a post-hoc power analysis on our data to estimate the power that was achieved given the actual observed effect sizes. We performed our power analyses using the R package 'simR' (Green, MacLeod, Alday et al., 2016; Green & MacLeod, 2016) for the glmer models reported below. We used 800 simulation runs for each power estimate, with an alpha level of 0.05.

Predictions

We now lay out our predictions for the different time regions in this study. Our hypothesis predicts that all cues are combined and allow people to anticipate upcoming input in accordance with all contextual cues (knowledge-based and linguistic) used jointly. At the target region, we expect the verb's selectional preferences to interact with the connective and cue the relevant script knowledge, restricting the expectations to the target object (verb match and connective match). We should hence observe more gazes to the target objects, both in sentences cued with *before* and *after*, and thus an interaction between connective match and verb match. There could potentially also be an effect of the connective match as early as the pretarget region, resulting in more gazes to the objects related to events matching the connective (target and temporal order distractor) compared to the other objects, as they match people's script knowledge of what comes *before* or *after* the main event (*before bagging the groceries* → *debit card*, *shopping cart* / *after bagging the groceries* → *key*, *car*).

Alternatively, as the Iconicity Assumption holds that event mentions following chronological order of events in the world should be easier to process than event mentions in the inverse order, this could also influence predictions for the target object. We do not expect a main effect of iconicity at any region, as the number of fixations is a window into the comprehender's expectations but not a measure of processing difficulty (which we may expect from reordering events presented in a noniconic order). However, processing difficulty could nevertheless interfere with the comprehender's predictions. We thus additionally tested whether there was a difference between *before*-sentences, which are noniconic, and *after*-sentences, which have iconic event order. At the pretarget region, this could yield an interaction between iconic order (*before* or *after* connective) and connective match, with matching AOIs being fixated more quickly in the iconic condition than in the noniconic condition. At the target region, we would expect a stronger effect of anticipating the target in iconic sentences than in noniconic sentences, which in our experimental design corresponds to a three-way interaction between iconic order, connective match, and verb match.

A shallower processing mechanism, which is strongly influenced by temporally oriented and order-sensitive script knowledge, may result in expectations based on future-oriented world knowledge only, without directly taking into account discourse cues indicating reordering or the verb's selectional preferences. We thus also conducted an analysis to test if the order-sensitivity of scripts may steer expectations toward *future* events and participants in the script representation (e.g., *eating a meal* → *reading the bill*, *ordering dessert*) regardless of contextual cues, leading to an advantage for AOIs that refer to events following the anchor event (e.g., *dessert* – future-matching) over AOIs that refer to events preceding the anchor event (e.g., *appetizer* – non-future-matching)

We would of course expect most gazes to the target object in the post-target region, when the referent is named. This region of interest is not analyzed to test our hypothesis but only to confirm that participants preferentially look at the target object when it is mentioned.

Results

All participants scored better than 77% correct on the comprehension questions ($M = 92\%$, $SD = 0.06$). Trials for which participants incorrectly answered the comprehension question were excluded from analyses (8% data points). An overview of the time course of fixation proportions to the different AOIs is provided in Figure 2 and Table 3.

Pretarget region: noun following *before* / *after*

In the region before the verb, all AOIs received a similar number of inspections: Objects matching the connective (target + temporal order distractor) were not inspected significantly more frequently than the objects not matching it (selectional preference distractor + unrelated distractor; connective match: $t = -0.974$, $p > .05$; $BF_{10} = 2.04$, median = -0.059 , MAD = 0.048, 90% CI $[-0.14, 0.020]$, overlap = 54.56%, MPE = 88.48%), and there was no effect of iconic order (iconic order: $t = 0.523$, $p > .05$; $BF_{10} = 1.02$,

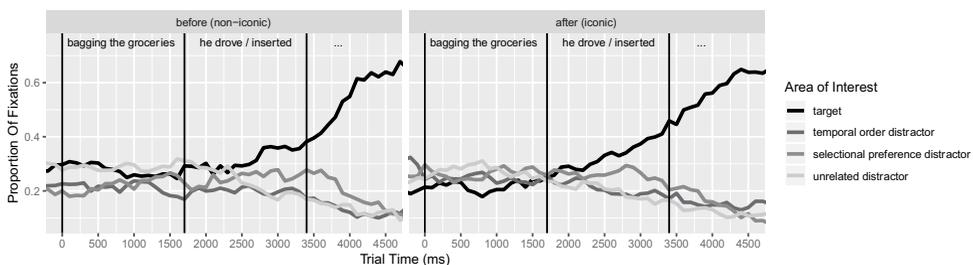


Figure 2. Proportions of fixations per AOI, region, and discourse cue over trial time.

Table 3. Fixation Proportions for Each AOI and Region

after	pre-target	target	post-target
<i>matching temporal order</i>	0.35		
target		0.23	0.47
temporal order distractor		0.17	0.14
<i>not matching temporal order</i>	0.41		
selectional preference distractor		0.20	0.12
unrelated distractor		0.19	0.12
<i>fillers</i>	0.24	0.21	0.15
before	pre-target	target	post-target
<i>matching temporal order</i>	0.38		
target		0.23	0.45
temporal order distractor		0.17	0.14
<i>not matching temporal order</i>	0.39		
selectional preference distractor		0.19	0.15
unrelated distractor		0.20	0.11
<i>fillers</i>	0.22	0.21	0.15

median = 0.018, MAD = 0.029, 90% CI [-0.030, 0.064], overlap = 76.15%, MPE = 73.60%). An interaction between connective match and iconic order was not significant ($t = 0.956$, $p > .05$; $BF_{10} = 1.89$, median = 0.067, MAD = 0.064, 90% CI [-0.032, 0.17], overlap = 59.85%, MPE = 86.33%) and did not improve model fit ($\chi^2 = 0.892$, $p > .05$). We did not test for an effect of verb match because it is not relevant at this region, as the verb has not been presented yet.

We found no significant effect of future match ($t = 0.516$, $p > .05$; $BF_{10} = 1.11$, median = 0.039, MAD = 0.054, 90% CI [-0.052, 0.13], overlap = 72.80%, MPE = 76.78%; iconicity: $t = 0.526$, $p > .05$; $BF_{10} = 0.91$, median = 0.018, MAD = 0.030, 90% CI [-0.031, 0.065], overlap = 77.17%, MPE = 71.50%). In addition, an interaction between future match and iconic order was not significant ($t = -1.290$, $p > .05$; $BF_{10} = 2.99$, median = -0.11, MAD = 0.068, 90% CI [-0.23, 0.0026], overlap = 42.63%, MPE = 94.35%) and did not improve model fit ($\chi^2 = 1.8819$, $p > .05$).

Target region

Table 4 shows the overview of the results for the target region (verb region). Objects matching the verb's selectional preference received significantly more gazes (verb match: coef = 0.09, $t = 2.258$, $p < .05$; $BF_{10} = 5.81$, median = 0.095, MAD = 0.032, 90% CI [0.041, 0.15], overlap = 14.56%, MPE = 99.70%). Additionally, there was a significant interaction of connective match and verb match, showing that the target object received most fixations (coef = 0.13, $t = 2.094$, $p < .05$; $BF_{10} = 5.57$, median = 0.14, MAD = 0.052, 90% CI [0.050, 0.22], overlap = 19.34%, MPE = 99.62%; the interaction was found to improve model fit significantly: $\chi^2 = 4.020$, $p < .05$). There was no significant main effect of connective match ($t = 0.898$, $p > .05$; $BF_{10} = 1.55$, median = 0.023, MAD = 0.024, 90% CI [-0.018, 0.062], overlap = 64.90%, MPE = 82.50%) or iconic order ($t = 0.393$, $p > .05$; $BF_{10} = 0.50$, median = 0.0069, MAD = 0.022, 90% CI [-0.027, 0.043], overlap = 87.41%, MPE = 62.25%), indicating that the amount of fixations to the target did not significantly differ between connectives. These findings are consistent with our hypothesis that script knowledge, discourse, and verb cues are seamlessly integrated.

We however note that effect size is a bit smaller than what we expected based on previous work. We performed a post-hoc power analysis using simr, which included the same random effect structure as in the data analysis. The power estimate for verb match was 62.7% for verb match and 45.2% for the interaction between verb match and connective match.

We also tested a model including a three-way interaction (connective match \times verb match \times iconic order) against a model including all two-way interactions only. This would allow us to test not only if comprehenders focus on the target object that matches both the temporal connective and the verb's

Table 4. Fixed and random effects and linear model for the verb region

Fixed Effects				
Predictor	Estimate	Standard Error	<i>t</i>	<i>p</i>
(Intercept)	0.87291	0.044452	19.637	< 0.001
connective_match	0.02593	0.028871	0.898	> 0.05
verb_match	0.09394	0.041607	2.258	< 0.05
iconicity	0.00853	0.021715	0.393	> 0.05
connective_match * verb_match	0.12951	0.061839	2.094	< 0.05
Random Effects				
Groups	Name	Variance	Standard Deviation	
Participant	(Intercept)	0.022840	0.15113	
	connective_match	0.007889	0.08882	
	verb_match	0.012940	0.11376	
	connective_match * verb_match	0.038729	0.19680	
Item	(Intercept)	0.006962	0.08344	
	verb_match	0.010199	0.10099	

Number of obs: 3272, groups: participant, 32; item, 28

fixations + 1 ~ connective_match * verb_match + iconicity + (1 + connective_match * verb_match | subject) + (1 + verb_match | item)

selectional preferences but also if this effect differs in strength between sentences that require reordering (noniconic order) and sentences that do not (iconic order). We found that the three-way interaction was not significant ($t = 1.088$, $p > .05$; $BF_{10} = 2.29$, median = 0.15, MAD = 0.12, 90% CI [0.030, 0.35], overlap = 50.80%, MPE = 90.85%) and did not improve model fit ($\chi^2 = 1.1923$, $p > .05$). We note that our study did possibly not have sufficient power for detecting the relevant three-way interaction.

Post-target region

Table 5 shows the overview of the results for the Post-target region (object region). As expected, the target is inspected more than the other AOIs at the object region, after it was mentioned, as shown by a significant effect of verb match (coef = 0.41, $t = 8.759$, $p < .001$; $BF_{10} > 150$, median = 0.41,

Table 5. Fixed and random effects and linear models for the object region

Fixed Effects				
Predictor	Estimate	Standard Error	<i>t</i>	<i>p</i>
(Intercept)	0.84132	0.03461	24.306	< 0.001
connective_match	0.42423	0.04895	8.666	< 0.001
verb_match	0.40987	0.04679	8.759	< 0.001
iconicity	0.01625	0.01946	0.835	> 0.05
connective_match * verb_match	0.65655	0.06778	9.687	< 0.001
Random Effects				
Groups	Name	Variance	Standard Deviation	
Participant	(Intercept)	0.006520	0.08075	
	connective_match	0.035952	0.18961	
	verb_match	0.032730	0.18091	
	connective_match * verb_match	0.062592	0.25018	
Item	(Intercept)	0.009738	0.09868	
	verb_match	0.397647	0.63059	

Number of obs: 3272, groups: participant, 32; item, 28

fixations + 1 ~ connective_match * verb_match + iconicity + (1 + connective_match * verb_match || subject) + (1 | item)

MAD = 0.041, 90% CI [0.34, 0.48], overlap = 0%, MPE = 100%) and of connective match (coef = 0.42, $t = 8.666$, $p < .001$; $BF_{10} > 150$, median = 0.43, MAD = 0.044, 90% CI [0.36, 0.50], overlap = 0%, MPE = 100%) and a verb match by connective match interaction (coef = 0.66, $t = 9.687$, $p < .001$; $BF_{10} > 150$, median = 0.66, MAD = 0.056, 90% CI [0.57, 0.75], overlap = 0%, MPE = 100%). There was no statistically significant effect of iconic order ($t = 0.835$, $p > .05$; $BF_{10} = 1.29$, median = 0.015, MAD = 0.019, 90% CI [-0.016, 0.048], overlap = 69.92%).

Discussion

Comprehenders rely on a variety of cues to build predictions about upcoming referents. Predictions may be directed toward participants of well-known scripts, which have been activated during language comprehension. Script knowledge, however, does not only include knowledge of typical everyday events and their participants but also the temporal order of such events. Combined with temporal connectives, script knowledge can direct expectations toward typical participants of events happening *before* or *after* a reference event.

We addressed the role of order-sensitive script knowledge, cued by temporal connectives and the verb's selectional preferences, in influencing anticipatory gaze to script event participants in a visual world paradigm. In particular, our goal was to investigate whether temporal connectives and verb selectional preferences interact early to cue script knowledge, helping comprehenders anticipate an upcoming event participant. In the pretarget region, when comprehenders had heard the first cue (*before* / *after* + script event, e.g., *After bagging the groceries*), the temporal discourse connective in combination with the script event did not preferentially steer expectations toward the event participant matching the connective. At this time region, objects matching the connective were not inspected more than objects not matching it. This was the case both for sentences with iconic order and sentences that required reordering: There was no interaction between iconicity and connective match. We did not find any inspections for AOIs depicting future participants (with reference to the main event) either. This could be because script events are highly associated between each other, and the temporal connective does not seem to influence anticipatory gaze strongly enough to overcome these associations: on hearing *after bagging the groceries* or *before bagging the groceries*, people do not only look at participants from the event following the bagging event but also at participants preceding it, regardless of the discourse connective. It is possible that contextual constraints are too weak at the point before the verb, meaning that anticipation of an object of an as-yet unknown verb is not sufficiently beneficial to comprehension.

When comprehenders heard the second cue (the main verb, e.g., *drove*), they were able to identify the target object as the only one matching the script knowledge, discourse connective, and verb cues and inspected it significantly more than the other distractors before it was mentioned. Our results are in line with other studies supporting immediate interpretation of discourse markers (Köhne & Demberg, 2013; Kuperberg et al., 2011; Traxler et al., 1997), selectional preferences (Altmann & Kamide, 1999; Hare et al., 2009), and script knowledge (Chwilla & Kolk, 2005; Khalkhali et al., 2012; Raisig et al., 2010; Van der Meer et al., 2002, 2005) as they show the early integration from all these types of cues. The different cues were rapidly integrated at the verb region regardless of what connective was used (both for *after*- and *before*-sentences), showing no advantage for iconic order.

The Iconicity Assumption predicts increased difficulties for constructions that require event reordering (noniconic constructions). However, in our experiment, this did not affect the comprehender's ability to anticipate the correct target when events were not mentioned in iconic order (sentences marked by *before*). The nature of script knowledge may have neutralized the cost of reordering: If comprehension involves the creation of a mental model of the events (including their temporal order), this process can be costly if the order has to be encoded (and possibly reordered) in the case of arbitrarily connected events (Mandler, 1986) but not if the temporal order is already encoded in the representation of events that are strongly connected by causal or inference links, such as script events.

Previous work by Van der Meer et al. (2002) and Van der Meer et al. (2005) showed that in a relatedness judgment task script event sequences following the iconic order (*after bite off – chew*) were recognized as related faster than pairs presented in the inverse order (*before digest – swallow*), and their results appear to be in conflict with ours, as do results from priming of script event sequences by Chwilla and Kolk (2005), Khalkhali et al. (2012), Van der Meer et al. (2002), and Van der Meer et al. (2005). These studies, however, present a few key differences with ours: They investigate decision latencies on relatedness or lexical decision tasks, they do not present whole sentences read by a speaker, and they do not look at anticipatory gaze. It is possible that our more naturalistic setting gave our subjects more time to cue events in both temporal directions compared to the nonsyntactic presentation in priming studies. An alternative possibility is that there is indeed a difference between *before*- and *after*-sentences in terms of effort, but it is simply not detected by anticipatory eye movements in a visual world paradigm. Pupillometry (see Demberg & Sayeed, 2016; Nuthmann & Van der Meer, 2005; Raisig et al., 2010) or a study using event-related potentials may offer a way to evaluate differences in terms of cognitive load between *before*- and *after*-sentences that anticipatory gaze simply does not detect.

Conclusion

We presented a novel visual world experiment demonstrating the interplay of script knowledge, temporal discourse markers, and linguistic knowledge. A script scenario was activated and then a sentence introduced a reference event, *after which* or *before which* a main event took place. We tested inspections to objects in a visual scene, which were relevant for different parts of the script, at different regions during sentence comprehension. After the reference event (e.g., *After bagging the groceries*), none of the objects was inspected more than the others. After the main verb (e.g., *he drove*), the object matching both the connector cue (*before / after*) and the verb cue (*drove*) was inspected significantly more than the competitors. Our results hence show that *before* and *after* connectives are integrated early with information about typical events in a script and about selectional preferences, giving rise to early predictions for a target referent. We did not find any difference in the degree to which *after* (indicating iconic order) and *before* sentences (requiring reordering) gave rise to expectations of the upcoming referent. We also did not find evidence for shallower processing strategies such as increased looks to future-oriented objects.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the German Research Foundation (DFG) as part of SFB 1102 (Project A3, Project-ID 232722074) and the Cluster of Excellence “Multimodal Computing and Interaction” (EXC 284). The work was carried out while the first author was affiliated with the Universität des Saarlandes.

References

- Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of Clinical and Experimental Neuropsychology*, 31(6), 731–775. <https://doi.org/10.1080/13803390802488103>
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Verbal Learning and Verbal Behavior*, 11(4), 417–423. [https://doi.org/10.1016/S0022-5371\(72\)80022-7](https://doi.org/10.1016/S0022-5371(72)80022-7)
- Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>

- Baker, L. (1978). Processing temporal relationships in simple stories: Effects of input sequence. *Journal of Verbal Learning and Verbal Behavior*, 17(5), 559–572. [https://doi.org/10.1016/S0022-5371\(78\)90337-7](https://doi.org/10.1016/S0022-5371(78)90337-7)
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Britton, B. K., Glynn, S. M., Meyer, B. J., & Penland, M. J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, 74(1), 51. <https://doi.org/10.1037/0022-0663.74.1.51>
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One*, 5(5), e10773. <https://doi.org/10.1371/journal.pone.0010773>
- Caron, J., Micko, H. C., & Thüring, M. (1988). Conjunctions and the recall of composite sentences. *Journal of Memory and Language*, 27(3), 309–323. [https://doi.org/10.1016/0749-596X\(88\)90057-5](https://doi.org/10.1016/0749-596X(88)90057-5)
- Chwillia, D. J., & Kolk, H. H. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*, 25(3), 589–606. <https://doi.org/10.1016/j.cogbrainres.2005.08.011>
- Clark, H. H., & Clark, E. V. (1968). Semantic distinctions and memory for complex sentences. *The Quarterly Journal of Experimental Psychology*, 20(2), 129–138. <https://doi.org/10.1080/14640746808400141>
- Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS One*, 11(1), 1. <https://doi.org/10.1371/journal.pone.0146194>
- 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. <https://doi.org/10.1007/BF00627434>
- Drenhaus, H., Demberg, V., Köhne, J., & Delogu, F. (2014). Incremental and predictive discourse processing based on causal and concessive discourse markers: ERP studies on German and English. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.) *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 403–408). Québec City, Canada: Cognitive Science Society.
- Fleischman, S. (1990). *Tense and narrativity*. UT Press.
- Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94(4), 427. <https://doi.org/10.1037/0033-295X.94.4.427>
- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 126. <https://doi.org/10.1037/0278-7393.10.1.126>
- Gernsbacher, M. A., & Givón, T. (1995). *Coherence in spontaneous text*. John Benjamins.
- Goodrich B., Gabry J., Ali I., & Brilleman S. (2020). *rstanarm: Bayesian applied regression modeling via Stan*. [Computer software manual]. <https://mc-stan.org/rstanarm> .
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493. <https://doi.org/10.1111/2041-210X.12504>
- Green, P., MacLeod, C. J., & Alday, P. (2016). *Package 'simr'* [Computer software manual]. <https://cran.r-project.org/web/packages/simr/simr.pdf>
- Güldenpenning, I., Kunde, W., Weigelt, M., & Schack, T. (2012). Priming of future states in complex motor skills. *Experimental Psychology*, 59(5), 286–294. <https://doi.org/10.1027/1618-3169/a000156>
- Haberlandt, K. (1982). Reader expectations in text comprehension. In J.-F. Le Ny & W. Kintsch (Eds.), *Language and comprehension* (pp. 239–249). Amsterdam: North-Holland. [https://doi.org/10.1016/S0166-4115\(09\)60055-8](https://doi.org/10.1016/S0166-4115(09)60055-8)
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- Hare, M., Elman, J. L., Tabaczynski, T., & McRae, K. (2009). The wind chilled the spectators, but the wine just chilled: Sense, structure, and sentence comprehension. *Cognitive Science*, 33(4), 610–628. <https://doi.org/10.1111/j.1551-6709.2009.01027.x>
- Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, 12(5), 822–851. <https://doi.org/10.3758/BF03196775>
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171. <https://doi.org/10.1016/j.actpsy.2010.11.003>
- Khalkhali, S., Wammes, J., & McRae, K. (2012). Integrating words that refer to typical sequences of events. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, 66(2), 106. <https://doi.org/10.1037/a0027369>
- Köhne, J., & Demberg, V. (2013). The time-course of processing discourse connectives. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, p. 2760–2765. Berlin, Germany.
- Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23(5), 1230–1246. <https://doi.org/10.1162/jocn.2010.21452>
- Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16, 437–493. <https://doi.org/10.1007/BF00986208>
- Mandler, J. M. (1986). On the comprehension of temporal order. *Language and Cognitive Processes*, 1(4), 309–320. <https://doi.org/10.1080/01690968608404680>
- McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, 3(6), 1417–1429. <https://doi.org/10.1111/j.1749-818X.2009.00174.x>

- Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33(1), 128–147. <https://doi.org/10.1006/jmla.1994.1007>
- Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS One*, 7(5), e37528. <https://doi.org/10.1371/journal.pone.0037527>
- Münste, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71–73. <https://doi.org/10.1038/25731>
- Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*, 25(2), 227–236. <https://doi.org/10.3758/BF03201114>
- Natsopoulos, D., Mentenopoulos, G., Bostantzopoulou, S., Katsarou, Z., Grouios, G., & Logothetis, J. (1991). Understanding of relational time terms before and after in Parkinsonian patients. *Brain and Language*, 40(4), 444–458. [https://doi.org/10.1016/0093-934X\(91\)90142-N](https://doi.org/10.1016/0093-934X(91)90142-N)
- Nieuwland, M. S. (2015). The truth before and after: Brain potentials reveal automatic activation of event knowledge during sentence comprehension. *Journal of Cognitive Neuroscience*, 27(11), 2215–2228. https://doi.org/10.1162/jocn_a_00856
- Nuthmann, A., & Van der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology*, 42(3), 306–317. <https://doi.org/10.1111/j.1469-8986.2005.00291.x>
- Politzer-Ahles, S., Xiang, M., & Almeida, D. (2017). “Before” and “after”: Investigating the relationship between temporal connectives and chronological ordering using event-related potentials. *PLoS One*, 12(4), e0175199. <https://doi.org/10.1371/journal.pone.0175199>
- Raisig, S., Welke, T., Hagendorf, H., & Van der Meer, E. (2010). I spy with my little eye: Detection of temporal violations in event sequences and the pupillary response. *International Journal of Psychophysiology*, 76(1), 1–8. <https://doi.org/10.1016/j.ijpsycho.2010.01.006>
- Sanders, T. (2005, November). Coherence, causality and cognitive complexity in discourse. In M. Aurnague, M. Bras, A. Le Draoulec, & L. Vieu (Eds.), *Proceedings of the first international symposium on the exploration and modelling of meaning* (pp. 105–114). Université de Toulouse.
- Sanders, T. J., & Noordman, L. G. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), 37–60. https://doi.org/10.1207/S15326950dp2901_3
- Sasanuma, S., & Kamio, A. (1976). Aphasic's comprehension of sentences expressing temporal order of events. *Brain and Language*, 3(4), 495–506. [https://doi.org/10.1016/0093-934X\(76\)90044-4](https://doi.org/10.1016/0093-934X(76)90044-4)
- Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates.
- Traxler, M. J., Sanford, A. J., Aked, J. P., & Moxey, L. M. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 88–101. <https://doi.org/10.1037/0278-7393.23.1.88>
- Trosborg, A. (1982). Children's comprehension of 'before' and 'after' reinvestigated. *Journal of Child Language*, 9(2), 381–402. <https://doi.org/10.1017/S0305000900004773>
- Van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54(5–6), 360–372. <https://doi.org/10.1080/0163853X.2017.1306677>
- Van der Linden, L., Mathôt, S., & Vitu, F. (2015). The role of object affordances and center of gravity in eye movements toward isolated daily-life objects. *Journal of Vision*, 15(5), 1–18. <https://doi.org/10.1167/15.5.8>
- Van der Meer, E., Beyer, R., Heinze, B., & Badel, I. (2002). Temporal order relations in language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 770. <https://doi.org/10.1037/0278-7393.28.4.770>
- Van der Meer, E., Krüger, F., & Nuthmann, A. (2005). The influence of temporal order information in general event knowledge on language comprehension. *Zeitschrift Für Psychologie / Journal of Psychology*, 213(3), 142–151. <https://doi.org/10.1026/0044-3409.213.3.142>
- Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6), 648–672. <https://doi.org/10.1080/23273798.2014.995679>
- Ye, Z., Kutas, M., George, M. S., Sereno, M. I., Ling, F., & Münste, T. F. (2012). Rearranging the world: Neural network supporting the processing of temporal connectives. *NeuroImage*, 59(4), 3662–3667. <https://doi.org/10.1016/j.neuroimage.2011.11.039>
- Ye, Z., Milenkova, M., Mohammadi, B., Kollwe, K., Schrader, C., Dengler, R., Samii, A., & Münste, T. F. (2012). Impaired comprehension of temporal connectives in Parkinson's disease—A neuroimaging study. *Neuropsychologia*, 50(8), 1794–1800. <https://doi.org/10.1016/j.neuropsychologia.2012.04.004>
- Zwaan, R. A., Madden, C. J., & Stanfield, R. A. (2001). Time in narrative comprehension. In D. H. Schram & G. Steen (Eds.), *Psychology and sociology of literature* (pp. 71–86). John Benjamins Publishing.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162. <https://doi.org/10.1037/0033-2909.123.2.162>
- Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 725–764). Elsevier.