Title
Yesterday Is History, Tomorrow Is a Mystery: An Eye-Tracking Investigation of the Processing of Past and Future Time Reference During Sentence Reading

Journal
Journal of Experimental Psychology Learning Memory and Cognition, 48(7)

ISSN
0278-7393

Authors
Biondo, Nicoletta
Soilemezidi, Marielena
Mancini, Simona

Publication Date
2022-07-01

DOI
10.1037/xlm0001053

Yesterday is history, tomorrow is a mystery: an eye-tracking investigation of the processing of past and future time reference during sentence reading

Nicoletta Biondo[a,b], Marielena Soilemezidi[c,d,a] & Simona Mancini[a]

[a]Basque Center on Cognition, Brain and Language, Spain; [b]University of Siena, Italy; [c]National and Kapodistrian University of Athens, Greece; [d]University of Oslo, Norway

Corresponding author:

Nicoletta Biondo (nicoletta.biondo@unisi.it)

Department of Social, Political and Cognitive Sciences – University of Siena

Room 303, Palazzo San Niccolò

Via Roma, 56

53100 Siena, Italy

Abstract

The ability to think about non-present time is a crucial aspect of human cognition. Both the past and future imply a temporal displacement of an event outside the "now". They also intrinsically differ: the past refers to inalterable events; the future to alterable events, to possible worlds. Are the past and future processed similarly or differently? In this study, we addressed this question by investigating how Spanish speakers process past/future time reference violations during sentence processing, while recording eye movements. We also investigated the role of verbs (in isolation; within sentences) and adverbs (deictic; non-deictic) during time processing. Existing accounts propose that past processing, which requires a link to discourse, is more complex than future processing, which—like the present—is locally bound. Our findings show that past and future processing differs, especially at early stages of verb processing, but this difference is not limited to the presence/absence of discourse linking. We found earlier mismatch effects for past compared to future time reference in incongruous sentences, in line with previous studies. Interestingly, it took longer to categorize the past than the future tense when verbs were presented in isolation; however, it took longer to categorize the future than the past when verbs were presented in congruous sentences, arguably because the future implies alterable worlds. Finally, temporal adverbs were found to play an important role in reinspection and reanalysis triggered by the presence of undefined time frames (non-deictic adverbs) or incongruences (mismatching verbs).

Keywords: time reference, tense, adverbs, sentence comprehension, eye movements

Introduction

Mentally locating events in the past or future plays a key role in many fundamental aspects of human cognition and behavior, such as planning, decision-making, and self-regulation (e.g. Schacter, 2012; Szpunar, 2010; Boyer, 2008; Suddendorf & Corballis, 2007; Atance & O'Neill, 2001). Recent developmental models suggest that the ability to distinguish between completed/not-alterable and yet-to-happen/alterable events is present at very early stages of human development; a full distinction between the past/future is internalized from 4-5 years of age (McCormack & Hoerl, 2017; see also McCormack, 2014; Zhang & Hudson, 2018). However, it remains unclear whether locating events in the past or future relies on the same cognitive mechanisms.

In this study we address this issue from a psycholinguistic perspective, focusing on whether native speakers of Spanish process past and future temporal information similarly or differently during online sentence processing and during the categorization of verbs presented in isolation.

Time in sentences

Languages such as Spanish or English encode temporal information through the *tense* morphology of finite verbs[1], which is indeed considered the "grammaticalization of time" (Comrie, 1985) and, optionally, through temporal adverbs such as "yesterday, at noon, afterwards" (Smith, 1981). For example, in sentences (1) and (2) temporal information is encoded through the deictic adverbs yesterday/tomorrow and through verb inflection.

(1) Ayer (ellos) bailaron swing en el Boulevard.

'Yesterday they danced swing on the Boulevard'

(2) Mañana (ellos) bailarán swing en el Boulevard.

'Tomorrow they will dance swing on the Boulevard'

[1] Because of the nature of the language under investigation in this study (Spanish), our review of the literature focuses on tensed languages. For temporal processing on tenseless languages like Chinese, see e.g., Qiu & Zhou (2012); Collart & Chan (2021) and references therein.

Tense features allow for the categorization of events as past or non-past, with respect to an evaluation time, which often corresponds to the speech time "now"[2]. For example, "danc-ed" describes an event that happened in the past relative to the current moment. Moreover, the point in time in which the event takes place can either be deduced from previous context or can be explicitly provided through a temporal adverb (Aronson, 1977; Partee, 1973), such as "yesterday" in (1), which defines the 24-hour time interval preceding today (the time of utterance). The same reasoning can be applied to the future tense and the temporal adverb in (2), which define an event that will happen in the future, specifically the day after today. Some researchers (e.g., Weger & Pratt, 2008; Ulrich & Maienborn, 2010; Bonato, Zorzi & Umiltà, 2012; Anelli et al., 2018) have proposed that time is conceptualized as running from left to right (at least in Western-European countries) along a mental timeline. A graphic representation of the events encoded in (1) and (2) is provided in Figure 1.

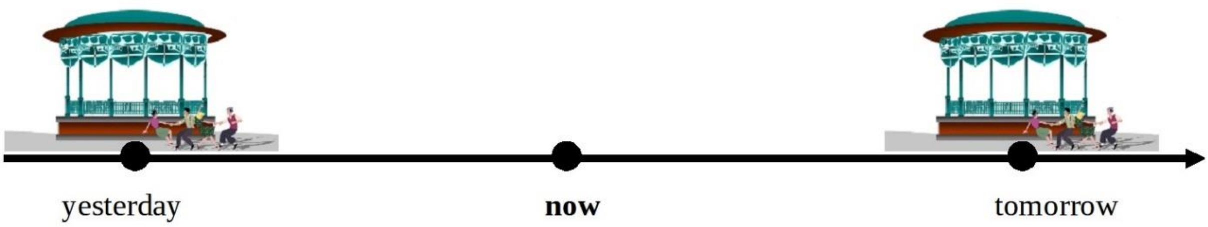

Figure 1. Graphic representation of the events encoded in sentences (1) and (2)

Past - future dissociation in previous studies

Previous literature has mainly focused on the dissociation between past and present, while the difference between the past and future has been discussed only inferentially, by generalizing evidence found for one non-past time (i.e. the present) to the other (i.e. the future).

[2] In main clauses and embedded relative clauses, the evaluation time is the time of utterance. In embedded sentences (i.e. complement clauses such as "John said that…") and sentences preceded by narrative context, evaluation times may be provided by previous tenses (Enç, 1987).

Some linguistic theories (Zagona, 2003; see also Avrutin, 2000) propose that when we refer to the past, the time of the event and the evaluation time (i.e., the utterance time) do not coincide: a link must thus be established between them in order to interpret the past tense. In this sense, the past is said to be *discourse-linked,* that is, it references a time specified by the discourse context, which lies outside the 'now'. Conversely, when referring to the non-past, that is, the present time, there is simultaneity between the time of the event and the utterance time. The present tense is thus considered to be *locally bound* to the time of utterance. Figure 2 provides a graphical representation of the locally bound and simultaneous nature of the present and the discourse-linked and externally anchored nature of past time reference.

The analysis formulated for the present tense has been extended to the future tense, considered as a form derived from the present tense via modal and aspectual features (Zagona, 2013), that is, via properties which specify how likely/real the event is and how it will unfold over time.

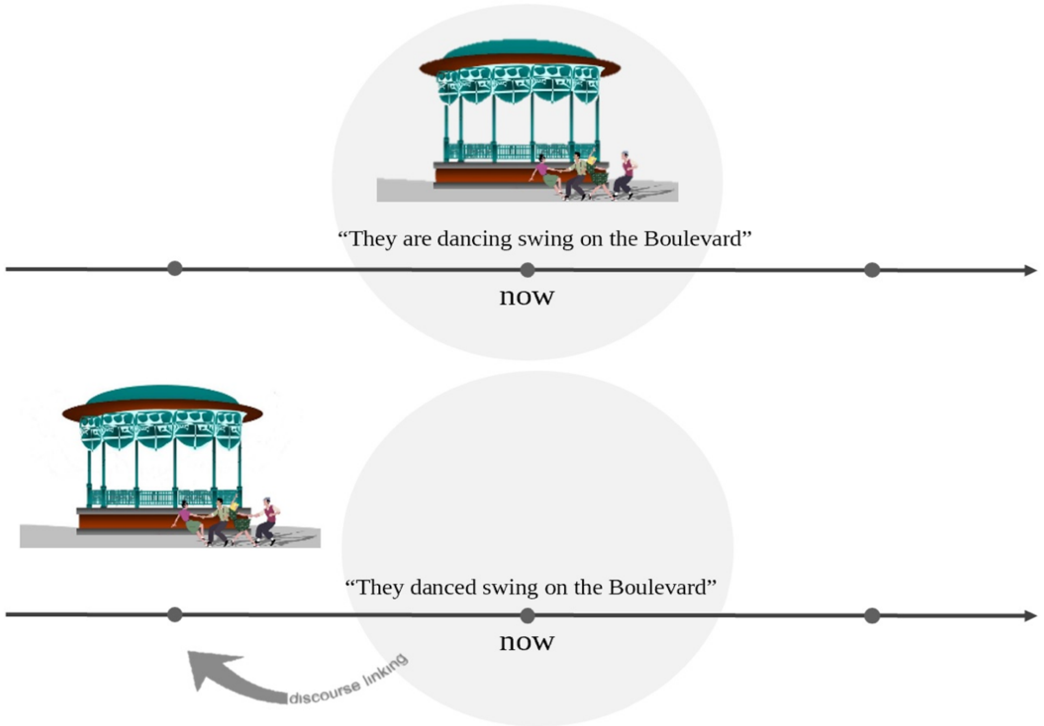

Figure 2. Graphic representation of the locally bound nature of present and discourse-linked nature of past time reference.

The hypothesis of a past-present dissociation has been corroborated by numerous experimental studies with language impaired and non-impaired speakers. Yet, the studies comparing the past and the future report less straightforward results.

On the one hand, data from studies with aphasic patients have shown impaired production and comprehension of past compared to present time reference (Bastiaanse et al., 2011; Dragoy et al., 2012; Bos et al., 2013). Interestingly, this neurocognitive dissociation between past and present time reference processing also seems to occur in speakers with unimpaired linguistic abilities (a summary is presented in Table 1). In an error detection task, Dragoy et al. (2012) found that a present tense verb violation of past time context (i.e., set by a temporal adverb, such as “just before”) was detected earlier than the violation of present time context by a past tense verb (Table 1, B). This pattern is in line with some ERP studies (Baggio, 2008; Bos et al., 2013) showing that violation of a past time frame by a present tense verb triggered either a Left Anterior Negativity (LAN) followed by a P600 or just a P600 compared to the control condition (Table 1, A). Conversely, detection of a present time frame violation by a past tense verb takes longer and does not trigger any reliably different pattern of neural activity from its correct counterpart (Table 1, C), apart from Sentence Final Negativity (SFN), as shown by Dragoy et al.’s study (2012). Overall, these findings are in line with the so-called *PAst DIscourse LInking Hypothesis* (PADILIH, Bastiaanse et al., 2011; Bastiaanse, 2013): processing past time reference requires establishing a link to discourse and is therefore more difficult than processing non-past time reference, where the event time is locally bound and thus easier to interpret.

	Compared conditions	Results
A	Adverb$_{PAST}$ Verb$_{PAST}$	**ERPs:** LAN+P600 (Baggio, 2008);
	Adverb$_{PAST}$ Verb$_{PRESENT}$	P600 (Bos et al., 2013)
B	Adverb$_{PAST}$ Verb$_{PRESENT}$	**Error detection time:**
	Adverb$_{PRESENT}$ Verb$_{PAST}$	$Adv_{PAST}V_{PRES} < Adv_{PRES}V_{PAST}$ (Dragoy et al., 2012)
C	Adverb$_{PRESENT}$ Verb$_{PAST}$	**ERPs:** SFN (Dragoy et al., 2012)
	Adverb$_{PRESENT}$ Verb$_{PRESENT}$	
D	Adverb$_{PAST}$ Verb$_{PRESENT}$	**ERPs:** P600+SFN (Dragoy et al., 2012)
	Adverb$_{PRESENT}$ Verb$_{PRESENT}$	

Table 1. Summary of previous ERP and behavioral studies on present/past verb dissociation. The first column describes the experimental conditions that were compared; the second column describes the related ERP or behavioral effects.

On the other hand, in a grammaticality judgment task, Faroqi-Shah and Dickey (2009) showed that both the aphasic and control groups detected present temporal violations faster than past and future tense violations, and found no difference between past and future violations. This finding is clearly at odds with the assumption that the present and future belong to the same non-past tense category and should therefore be processed similarly, as assumed by PADILIH. A potential explanation is that past and future violations led to similarly longer reaction times compared to present tense violations because in both cases the time of the event was located outside the "now", either before or after the time of utterance (i.e. in the past or in the future, respectively), as shown in Figure 1. Fyndanis et al. (2018) also found no reliable difference between past and future time reference in a sentence completion task, where Greek and Italian speakers with agrammatic aphasia were asked to produce past and future time verbs triggered by sentences containing past/future deictic temporal adverbs. The authors failed to find a time reference effect both in the separate and aggregated analysis of Greek and Italian agrammatic patients. Interestingly, however, they found that a greater number of errors unrelated to time reference were made in the future time condition. This finding was clearly at odds with PADILIH, which predicts that past time reference will impose a greater cognitive load on patients (a population with limited processing resources). The authors suggested that future seems to tax the processing system of aphasic patients more than past time reference, since it involves more abstract representations (reference to possible worlds). In a sentence-picture matching task where eye movements were recorded, Bos et al. (2014) showed that both the aphasic and control groups were slower and less accurate in selecting the right picture for sentences including verb forms with future than past time reference. Nevertheless, this unexpected difficulty for future time reference was explained by a preference for pictures depicting past events, which were visually more salient than those representing future events, as confirmed by the greater number of fixations that past-related pictures received. Finally, Martinez-Ferreiro & Bastiaanse (2013) tested Spanish- and Catalan-speaking aphasic patients and a control group on temporal reference production and comprehension, in a sentence-picture matching task. The results showed ceiling performance in the control group and a gradient pattern in the aphasic group: past time reference triggered the most errors, followed by the future, and present time reference triggered the least errors.

In sum, most studies confirm that it is more difficult to process past than present time reference, while potential dissociations between the past and future remain unclear. Conceptually speaking, the past and future could be considered two sides of the same coin. They both imply a temporal displacement of the event expressed by the verb, respectively to the left and the right side of the time of utterance on the mental time line. One can thus hypothesize that past and future are similarly processed through a mechanism that allows for the interpretation of temporally displaced events. However, there are several reasons to believe that future and past time reference differ. First, the past is immutable and beyond our control, since it deals with events that have already taken place, while the future is necessarily more speculative, as it describes events that are yet to happen (Comrie, 1985). The study of past and future (as opposed to present) thus allows us to better investigate the crucial cognitive dissociation between completed/not alterable events and yet-to-happen/alterable events, a fundamental dissociation that is already made in the very first years of human development. Second, in addition to involving temporal displacement of an event, the future also involves modal displacement, that is, reference to possible worlds (Bochnack, 2019 and references therein). The relation between modality and future time reference has been mentioned in previous accounts (e.g. Zagona, 2013). However, to our knowledge, no study has investigated if/how the modal properties of future time reference affect sentence processing, and whether this difference between the past and future is reflected in the deployment of different cognitive mechanisms during online sentence processing.

A final issue that remains unresolved by previous studies is whether processing past (compared to non-past) tense is more difficult because it violates the past time frame provided by the adverb, or simply because it is more difficult to process past verbs *per se*, independently of the presence of a temporal adverb. For example, Dragoy and colleagues (2012) kept the inflected verb constant and manipulated the temporal information provided by the adverb in non-past conditions (Table 1, D), while they kept the adverb constant and changed the verb tense in past conditions (Table 1, C). Given that this discrepancy in the selection of the baseline condition could have affected the interpretation of the data, in subsequent work (i.e., Bos et al. 2013) the authors kept the adverb constant and manipulated the tense of the following verb (Table 1, A), permitting cleaner observation of the effects that violating a temporal frame had on processing. However, this manipulation still does not allow us to exclude an alternative explanation, namely that processing past time reference on the verb may be more difficult than processing the non-past, in itself, independently of the presence of a (mismatching) temporal adverb. Indeed, previous findings showing that a violation on a non-past verb is detected earlier than a violation on a past verb (Table 1, B) are compatible with both the adverb- and the verb-related interpretation of the processing cost.

The current study

The main aim of this study was to provide a systematic investigation of the processing of past and future time reference during sentence comprehension through the use of the eye-tracking technique, in order to clarify unresolved issues in the current literature on time processing. A schema of the questions addressed, of the conditions tested and of the main hypotheses is provided in Table 2.

Compared to other behavioral methods (e.g. error detection tasks, self-paced reading) or to electrophysiological paradigms used in previous studies, eye-tracking allows for a more ecological presentation of the linguistic material, without losing temporal resolution. A variety of dependent variables can be analyzed. Latency measures include *first-pass reading time*, which measures the sum of all fixations on a region of interest (e.g., the verb) before leaving it either to the left or to the right; *go-past time* (also called *regression path duration*) that is the sum of the first-pass plus the time spent rereading previous regions before moving past the region of interest; *total time* that is the sum of all fixations made on a region of interest, including refixations made after the region has been exited to the right. Regression measures include the probability or number of regressive saccades made inside or outside a given region. First-pass is generally considered an “early” measure while total time is considered a “late” measure, arguably reflecting different stages (early and late, respectively) of processing (e.g. Clifton, Staub & Rayner, 2007; but see also Vasishth, von der Malsburg & Engelmann, 2012). Other measures such as go-past time or regression measures have been considered both early and late measures (*ibidem*). For this reason, they have also been defined as hybrid or intermediate measures (e.g. Conklin, Pellicer-Sánchez & Carrol, 2018).

Are past and future time reference violations differently processed? (Q1)

The first goal of our study was to expand our understanding of time reference processing during sentence comprehension, by further investigating future time processing and comparing it with past time reference processing in Spanish.

Critically, by contrasting past and future, rather than past and present time forms, we aimed to create a sharp contraposition between immutable past events and alterable future events not only at the conceptual level, but also linguistically, thanks to the use of synthetic future-inflected verb forms. Spanish is particularly suitable, as it encodes both past and future through a morphological suffix attached to the verb (e.g. past: *compraron*; future: *comprarán*). Other languages such as English often have a periphrastic realization of future time reference, which differs from the way simple past is encoded (e.g. past: *(they) buy;* future: *(they) will buy*). The verb tense (future/past) of the experimental sentences was manipulated so as to match or mismatch the time frame set by a deictic adverbial, which was kept constant in line with previous research (Baggio, 2008; Bos et al., 2013). A sample of the experimental manipulation is shown in (3) and (4) below.

(3) A la salida del trabajo, **ayer** las chicas **compraron/***comprarán** pan en la tienda.
 'While leaving from work, yesterday the girls bought/*will buy bread at the shop'

(4) A la salida del trabajo **mañana** las chicas **comprarán/*compraron** pan en la tienda.
 'While leaving from work, tomorrow the girls will buy/*bought bread at the shop'

When the adverb reference and the time expressed by the verb do not match it is impossible to temporally locate the event. Nevertheless, the parser may spend some time attempting to link the verb with the preceding adverb. It is thus expected that the mismatch between the adverb and the verb may increase reading times at the verb position. In particular, both early and late measures are expected to reflect longer reading times in the verb region when sentences contain a temporal mismatch between adverbs and non-adjacent verbs. This prediction would be in line with previous eye-tracking studies investigating the processing of temporal violations. In particular, Biondo et al. (2018) tested adverb-verb violations by using an eye-tracking-while-reading paradigm, in Spanish. Their results showed that when the adverb and verb were separated by a subject NP (as in the current study), temporal violations gave rise to longer reading times compared to the control condition, both in first-pass and total time measures.

Two different hypotheses can be formulated concerning the timing and strength of the processing disruptions generated by past and future time violations. On the one hand, PADILIH (Bastiaanse et al., 2011; Dragoy et al., 2012; Bos et al., 2013; Bos et al., 2014) predicts that past and future processing should give rise to different effects, due to differences related to discourse anchoring. Past time reference is discourse-linked, i.e., the time of the event and the utterance time do not coincide. This extra-syntactic link may be cognitively more complex to process compared to the locally bound future time reference. For this reason, if a future time frame is followed by a verb referring to the past, a discourse-link needs to be established at the verb, and this processing load can lead to later detection of the violation. Conversely, when a past time frame is followed by a future verb, there is no-discourse linking launched at the verb, so temporal violations give rise to immediate effects. This prediction would be in line with previous research (e.g. Baggio, 2008; Bos et al., 2013). On the other hand, future and past time reference both imply non-coincidence between the event time and the utterance time, i.e., they are similarly anchored to a point in time that either precedes or follows the time of utterance, so they may be processed similarly. Within this scenario, violations of past and future reference should give rise to similar processing penalties, both in early and late reading measures.

What is the role of temporal adverbs and verbs during time processing? (Q2)

The second objective of this study was to address another question that was left unanswered by previous studies, namely the role that adverbs and verbs play during time processing.

To address the question about the role of temporal adverbs, we included an additional set of correct sentences in which the adverb did not contain deictic information. Differing from the deictic temporal adverbs (such as "yesterday"), which provided an unambiguous ±past/future interpretation, these were non-deictic adverbs (e.g. *el lunes*, on Monday), which did not explicitly mark temporal reference (Alexiadou, 1997), such that their temporal interpretation depended on the time expressed by the verb. By comparing sentences containing deictic temporal adverbs with sentences containing non-deictic temporal adverbs (see 5 and 6 below), we intended to ascertain whether and how past and future time processing on the verb changes as a function of the temporal information provided by different temporal adverbs.

(5) A la salida del trabajo **ayer/el lunes** las chicas **compraron** pan en la tienda.
'After leaving work yesterday/on Monday the girls bought bread at the shop'

(6) A la salida del trabajo **mañana/el lunes** las chicas **comprarán** pan en la tienda.
'After leaving work tomorrow/on Monday the girls will buy bread at the shop'

If the temporal adverb plays a major role in the processing of time reference, the presence of a deictic adverb is expected to facilitate the reading of a tensed verb, because it provides contextual information that drives expectations concerning the tense of the upcoming verb. This could generate overall faster processing of deictic compared to non-deictic sentences. In this scenario, it is plausible that a potential past-future dissociation would only emerge in the presence of deictic adverbs, since these are marked for temporal reference. For example, some experimental studies have shown that the parser is more likely to engage in predictive processing when the subject carries marked features (1st/2nd person features) compared to unmarked (3rd person) features (Nevins et al., 2007; Alemán Bañón & Rothman, 2019). It could be that, like marked subjects, (deictic) temporal adverbs with a defined temporal value play a different role than (non-deictic) temporal adverbs whose temporal value is undefined or ambiguous. Alternatively, if there is greater reliance on verbal rather than adverbial information during sentence comprehension, verb processing should not be affected by the presence of different types of temporal adverbs. Within this scenario, a potential past-future dissociation should arise in terms of higher processing costs for past compared to future time processing regardless of the type of adverb involved. This finding would also imply that the early/late latencies of any mismatch effect are likely to be related to the ease/difficulty of processing the tensed verb, independently of the time frame set by the adverb. Recent accounts suggest that the cognitive mechanisms at play during the processing of core phrases such as verbs may differ from those underlying the analysis of optional phrases such as adjuncts (e.g., Boland & Blodgett, 2006; Tutunjian & Boland, 2008; see also Martorell, 2018). As a consequence, it is possible that verb processing is not affected by different types of temporal adverbs, because adjuncts play a secondary role during processing.

In order to test how time reference on verbs is processed, *per se*, the eye-tracking experiment was followed by a verb temporal decision task. In this task, participants were asked to read tensed past and future verbs and to categorize them as past or future. We had two distinct predictions in this respect. If past and future time reference also differ during the processing of verbs in isolation, shorter reaction times should be found for the categorization of future verbs compared to past verbs, since past time is more difficult to process (as proposed by PADILIH). If past and future time reference do not differ when verbs are presented in isolation, no difference in reaction time should be found during the categorization of past and future verbs (alternative hypothesis). Based on previous studies reporting ceiling performance in the accuracy data of healthy speakers, and given the simplicity of the task, we expected similarly high accuracy in the offline categorization of past and future verbs.

Question 1: Are past and future time reference violations differently processed?		
Eye-tracking while reading (Q1)		***Main hypotheses***
Past mis/match	e.g., '... ayer ... compraron/*comprarán...' ('…yesterday … bought/*will buy …')	mismatch > match (early)
Future mis/match	e.g., '... mañana ... comprarán/*compraron...' ('…tomorrow … will buy/*bought …')	mismatch > match (late)
Question 2: What is the role of temporal adverbs and verbs during time processing?		
Eye-tracking while reading (Q2)		***Main hypotheses***
Past non-/deictic	e.g., '... el lunes/ayer ... compraron...' ('…on Monday/yesterday … bought …')	past > future; non-deictic > deictic
Future non-/deictic	e.g., '... el lunes/mañana ... comprarán...' ('…on Monday/tomorrow … will buy …')	
Temporal decision task		
Past	e.g., 'compraron' ('bought')	past > future
Future	e.g., 'comprarán' ('will buy')	

Table 2. Summary of the questions, conditions and main hypotheses tested.

Method

Participants

Sixty-six Spanish native speakers (48 women; age range 18-35 years; mean = 23.1; SD = 4.2) were invited to participate in this study for a small payment. Because participants were recruited from the local Spanish-Basque bilingual community, care was taken to select participants whose first language was Spanish (Spanish AoA: mean = .4 years, SD = .84; Basque AoA: mean = 3.07 years, SD = 2.74); who were predominantly exposed to Spanish (Spanish exposure: mean = 71.07%, SD = 13.88; Basque exposure: mean = 21.43%, SD = 1.35); and who were more proficient in Spanish than Basque, based on self-report and standardized BEST (De Bruin et al., 2017) and LexTale (Lemhöfer & Broersma, 2012) scores (Spanish BEST: 64.73 (SD: .63); Spanish LexTale: 52.1 (SD: 6.64); Basque BEST: 30 (SD: 20.79); Basque LexTale: 22.33 (SD: 16.54). Sample size was determined a-priori using the *simR* package (Green & MacLeod, 2016) to run simulations (Brysbaert & Stevens, 2018). The output showed that more than 50 participants were necessary to achieve the 80% power threshold required to reliably detect a violation effect from early measures onwards (see Supplementary Material). All the participants had normal or corrected to normal vision and no reading disorders. The experiment was approved by the BCBL Ethics Review Board. All participants signed a consent form in line with the Helsinki Declaration.

Material and design

Eye-tracking experiment

For the creation of the stimuli, a list of indicative mood verbs in 3rd person plural were generated via the EsPal subtitles database (Duchon et al., 2013) with log count constraints set to a minimum value of 1.26, i.e., the average frequency value according to the database, and a maximum value of 7.16, i.e., the highest frequency value available in the database. The log frequency range of the extracted verbs was .017 – 2.01. Then, we selected verbs whose log frequency in the future and the past tense (*pasado simple,* simple past) was similar (past: mean = .06, SD = .13; future: mean = .06, SD = .13). We chose 3rd person plural verbs to match past and future verbs in length, i.e., number of characters (Past: mean = 10.05, SD = 1.87; Future: mean = 10.03, SD = 1.87).

In the experimental sentences we used a variety of deictic adverbials in order to avoid participants becoming habituated to the critical constituents (e.g. *ayer/mañana por la noche,* ‘yesterday/tomorrow night’; *el año pasado,* ‘last year’; *la próxima semana*, ‘next week’). Past and future deictic adverbs were balanced in length (Past: mean = 15.82, SD = .71; Future: mean = 16.13, SD = 3.92; $t_{(203)} = .58$, $p = .56$). A variety of non-deictic temporal adverbs was also used to avoid excessive repetition (e.g. *el sábado a la una,* ‘Saturday at one o’clock’; *en diciembre,* ‘in December’; *en el mes de mayo*, ‘in the month of May’). Non-deictic and deictic temporal adverbs were also matched in length (Deictic: mean = 15.98, SD = 3.73; Non-Deictic:

mean = 16.28, SD = 3.59; $t_{(304)}$ = .69, p = .49). In order to verify that the adverbs used for the experimental stimuli were unmistakably perceived by native Spanish speakers as deictic (i.e., adverbs that refer exclusively either to the past or future) or non-deictic (i.e., adverbs that can potentially refer to both the past and the future), an adverb decision task was conducted (24 participants not recruited for the main experiment, 17 women, age range 19-32 years, mean age = 23, SD = 3.8). In this task, all the adverbs used in the experimental material were presented one by one to the participants, who had to decide whether each adverb could be followed by a verb in the past, in the future, or both. The adverbs were correctly categorized as referring to past, future, or both tenses with an accuracy above 90% (deictic past: mean = .99, SD = .02; deictic future: mean = .93, SD = .04; non-deictic: mean = .92, SD = .05). Past and future adverbials were combined with past and future verbs to create correct and incorrect sentences, in a 2 x 2 design, with *time reference* (past, future) and *verb match* (match, mismatch) as within-subject factors, as shown in Table 3. An additional set of correct stimuli was included to investigate the role of adverbs and verbs during time processing. Two factors were manipulated: adverb type (deictic and non-deictic) and time reference (past, future) in a 2x2 within-subject factor design, as shown in Table 4. A total of one-hundred and two sentences were created.

Past	
Match	***Mismatch***
Gracias a la beca, el año pasado los investigadores progresaron en sus estudios sobre la polución. 'Thanks to the scholarship, last year the researchers made progress on their studies on pollution'	Gracias a la beca, el año pasado los investigadores progresarán en sus estudios sobre la polución. 'Thanks to the scholarship, last year the researchers will make progress on their studies on pollution'
Future	
Match	***Mismatch***
Gracias a la beca, el próximo año los investigadores progresarán en sus estudios sobre la polución. 'Thanks to the scholarship, next year the researchers will make progress on their studies on pollution'	Gracias a la beca, el próximo año los investigadores progresaron en sus estudios sobre la polución. 'Thanks to the scholarship, next year the researchers made progress on their studies on pollution'

Table 3. Experimental material addressing question 1

Past	
Deictic	***Non-deictic***
Gracias a la beca, el año pasado los investigadores progresaron en sus estudios sobre la polución. 'Thanks to the scholarship, last year the researchers made progress on their studies on pollution'	Gracias a la beca, a lo largo del año los investigadores progresaron en sus estudios sobre la polución. 'Thanks to the scholarship, throughout the year the researchers made progress on their studies on pollution'
Future	
Deictic	***Non-deictic***
Gracias a la beca, el próximo año los investigadores progresarán en sus estudios sobre la polución. 'Thanks to the scholarship, next year the researchers will make progress on their studies on pollution'	Gracias a la beca, a lo largo del año los investigadores progresarán en sus estudios sobre la polución. 'Thanks to the scholarship, throughout the year the researchers will make progress on their studies on pollution'

Table 4. Experimental material addressing question 2

All the experimental sentences contained 5 regions of interest: an adjunct phrase in sentence-initial position; followed by a deictic time adverbial (encoding either past or future information) or a non-deictic time adverbial; a third person plural animate subject; the critical verb (encoding either past or future tense features); and, finally, an adverbial adjunct in sentence final position.

In order to balance the design and achieve an equal number of grammatical and ungrammatical sentences, as well as to vary the position of the anomalous word, 122 filler sentences were added to the list in which gender agreement was manipulated either between the noun and a predicative adjective (32 sentences; e.g., *el museo es bella*, the museum$_{masc}$ is beautiful$_{fem}$) or between the determiner and the noun (44 correct, 44 incorrect; e.g., *siempre nos visitaba el/*la capitán del equipo*, the captain of the team always visited us). Importantly, filler sentences contained inflected verbs whose tense was not limited to the past and future indicative. Indicative present, other forms of the indicative past (e.g., *comía*, ate (imperfective aspect), *ha comido*, has eaten), as well as verbs in the conditional and subjunctive mood were included, with 3rd as well as 1st, and 2nd person singular and plural inflections.

The naturalness of the experimental material was evaluated by the same participants recruited for the adverb decision task (the rating was administered before the adverb decision task). In the naturalness judgment task, the participants read both the experimental and filler sentences, and had to evaluate their naturalness on a 5-point Likert scale (1=completely unnatural, 5=completely natural). After the ratings, one item that had received an overall low naturalness score in its correct version (mean rating score < 3) was excluded from the original Latin Square design for the eye-tracking experiment. The remaining 96 experimental items (16 per condition) were randomly assigned to six different lists according to a Latin Square design, so that each participant was presented with only one version of each sentence.

The mean rating scores related to the experimental material are reported in Table 5. Data were analyzed through two linear mixed-effect model analyses (Gibson, Piantadosi & Fedorenko, 2011; Kizach, 2014). All the factors were sum-coded (match, past, deictic = -0.5; mismatch, future, non-deictic = 0.5) and the random effect structure was selected parsimoniously (Bates et al., 2015). In the first analysis we checked whether the mismatch sentences were in effect rated as less natural than the match sentences, and whether there was any effect or interaction with time reference in the deictic conditions. Results show that deictic sentences were rated as less natural in presence of a verb mismatch, as expected (Intercept: 3.14; Estimate: -2.25, SE: .17, t = -13.41, p < .05). The rating for past and future sentences did not differ (Intercept: 3.14; Estimate: .05, SE: .05, t = .98, p > .05) and there was no interaction between time reference and verb mis/match (Intercept: 3.14; Estimate: .05, SE: .1, t = .45, p > .05). In the second analysis we compared the deictic and non-deictic match conditions, and whether there was any difference related to past/future time reference. Non-deictic sentences were rated as less natural compared to the deictic sentences (Intercept: 4.17; Estimate: -.21, SE: .05, t = -3.96, p < .05), although both conditions showed similar numerical ratings. There was no difference between past and future conditions (Intercept: 4.17; Estimate: -.06, SE: .05, t = -1.28, p > .05) and the time reference did not interact with the deicticness of the adverb (Intercept: 4.17; Estimate: .18, SE: .11, t = -1.69, p > .05).

Adverb	Verb	Rating (SD)
Deictic (Past)	***Match***	4.3 (.6)
	Mismatch	2 (.7)
Deictic (Future)	***Match***	4.3 (.5)
	Mismatch	2 (.7)
Non-deictic	***Match (past)***	4.1 (.6)
	Match (future)	4 (.6)

Table 5. Mean rating scores (standard deviation in brackets) of the naturalness judgement task.

Temporal decision task

A "temporal decision task" was administered after the eye-tracking experiment. The same set of past and future verbs presented in the eye-tracking experiment was used, that is, 96 past verbs and 96 future verbs. Taking into account that in Spanish the future form of the experimental verbs is always stressed on the last syllable (e.g., *trabajarán,* will work) whereas the past form is never stressed (e.g. *trabajaron,* worked), it was imperative to enrich the list of verbs in this task so as to achieve greater variability and ensure participants did not classify verbs solely by considering the presence or absence of stress on the final syllable. For this reason, in addition to the 48 past and 48 future experimental verbs in each of the two lists, 48 future filler verbs without an accent on the last syllable (e.g., *recogeremos,* will collect - 1st person plural) and 48 past fillers with an accent on the last syllable were added (e.g., *ilustró,* illustrated - 3rd person singular*; aceleré,* speeded up - 1st person singular). These had a similar number of characters as the experimental verbs (filler verbs: mean = 10.21, SD = 2.48; experimental verbs: mean = 10.04, SD = 1.86) and similar log frequency (filler verbs: mean = .09, SD = .16; experimental verbs: mean = .06, SD = .13), although only experimental verbs were considered in the analysis. A full list of experimental material can be found at https://osf.io/hkxw8/.

Procedure

Eye-tracking

Eye movements were recorded using an SR Eye-Link 1000 machine interfaced with a 19″ CRT Viewsonic monitor (60 cm from participants' eyes). The stimuli were displayed via Experiment Builder Software (SR Research, Ontario, Canada) and were distributed in four experimental blocks. Participants had binocular vision while their eye movements were recorded, but only the right eye was tracked. The experiment was conducted in a dimly lit cabin to provide a favorable viewing environment. In order to minimize head movements, a chin rest bar and a forehead restraint were used. Before the experiment, and when necessary during the experiment, a calibration procedure was conducted in which participants had to fixate 13 positions indicated by a white dot, linearly distributed along the bottom, central and top lines of the screen. Each trial was initiated by fixating on a white dot on the left side of the screen, where the first word of the sentence would appear. Once fixation on the white dot reached a stable value, the sentence was displayed. All sentences were presented in 19-point font (Times New Roman). Subjects had to press the space bar to end the presentation of a sentence. Participants were instructed to read each sentence at their own pace, to avoid blinking (except before the sentences appeared on the screen), and to answer yes/no comprehension questions by pressing one of the two corresponding buttons on the keyboard (C, M). Comprehension questions followed 33% of the sentences (e.g., the sentence "During the festival last night the actors presented a play by Molière" was followed by the question "Was a play by Molière presented?"). To familiarize participants with the experimental procedure, the session started with a short practice block of 5 trials (featuring sentences that were similar to but not included among those in the experimental list). The eye-tracking experiment lasted approximately 60 minutes, including practice, calibration, and breaks after the end of each of the four experimental blocks.

Temporal decision task

This task was administered to all participants after the eye-tracking experiment. Participants were instructed to read the verbs appearing on the screen (one at a time) and indicate as fast as possible whether they were in the future or past tense, by pressing either the up or down arrow keys on the keyboard, respectively. We used the up and down keys to avoid biased left-right associations. In western society, left and right are conceptually associated with the past and future, respectively; hence, keys placed along a horizontal line might have affected participants' performance (Ulrich & Maienborn, 2010; Bonato et al., 2012). Moreover, the position of past and future response keys was counterbalanced across participants. The experiment was run with Psychopy (version 1.73.04).

Data analysis

Eye-tracking

Eye-movement data from each participant and within each interest area were preprocessed using Data Viewer software (SR Research, Ontario, Canada). Individual fixations that were longer than 800 msec and shorter than 80 msec were removed by an automatic procedure or merged with the preceding/following fixation if the distance was within 0.5 degrees.

Of the 5 regions of interest in each experimental sentence, the reading time data of 3 regions were considered for data analysis: the verb region (target), the post-verb region, and the adverb region. We analyzed all relevant measures of the target region: first-pass, go-past, total time, and the probability of regression in(side) and out(side) the target region. Conversely, we selected which measure to analyze in the other regions based on our predictions (von der Malsburg & Angele, 2017). The post-target region was included, since previous behavioral literature showed that temporal violations gave rise to sustained effects in spillover areas such as the one following the target (Biondo, 2017; Biondo et al., 2018). We decided to analyze all relevant measures except the probability of regression inside the post-target region, since we did not have any hypothesis regarding the probability of going back to reread the post-target region from the end of the sentence region. The adverb region was also included, to assess whether our experimental conditions gave rise to different re-reading patterns in this area, as measured by the total reading time and the probability of regression variables. For this reason, we only analyzed the total time and the probability of regression inside the adverb region.

We conducted the analysis by fitting linear mixed-effect models to our data, using the R package *lme4* (Bates et al., 2014). We included *time reference* (past, future), and *verb match* (match, mismatch) as fixed-effect factors in the models used to investigate the processing of past-future violations (Q1), by adopting sum contrast coding (Schad et al., 2020): past and match conditions were coded as -0.5; while future and mismatch conditions were coded as 0.5. Given an interaction, we fitted a second linear mixed-effect model with separate nested contrasts for tense mis/match in the past and future conditions. In this model, match and mismatch were also coded as -0.5 and 0.5, respectively. We included *time reference* (past, future) and *adverb type* (deictic, non-deictic) as fixed-effect factors sum coded as -0.5 (past, deictic) and 0.5 (future, non-deictic) in the models used to investigate the role of adverbs (Q2). Moreover, we included crossed random intercepts and random slopes for all fixed-effect parameters for subject and item grouping factors (Barr et al., 2013) in all models.

We reduced the complexity of the random effect structure of the maximal model by performing a Principal Component Analysis so as to identify the most parsimonious model properly supported by the data (Bates et al., 2015). The best-fitting model for each research question, region, and measure is reported in Appendix A. We used the Box-Cox method to determine the transformation of the reading time data (Box & Cox, 1964), applying the *boxcox* function of the *MASS* package (Venables & Ripley, 2002); all reading time data were log transformed before performing the analyses. Logit mixed-effect models were employed (Jaeger, 2008) for the analysis of the probability of regression measure. Given that eye-tracking latency measures are all correlated to some degree and can lead to an inflated rate of false positives, we applied the Bonferroni correction to correct p-values for multiple comparisons (von der Malsburg & Angele, 2017). We divided the α threshold for determining significance based on the number of tests that were performed. A fixed effect was then considered significant if its p-value was equal or smaller than .0045. P-values were derived by using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017).

Temporal decision task

The analyses of the (log-transformed) reaction times and accuracy data were carried out by adopting a linear-mixed effect model analysis similar to the one adopted for the analysis of data from the eye-tracking task. In these analyses, the fixed-effect factor was *time reference* (past, future) and the default treatment contrast coding was used (Schad et al., 2020).

Results

We collected data from 66 participants, but only 60 were entered in the analyses. Data from five participants were discarded because of technical problems during data acquisition (e.g., poor calibration), while data from one participant were excluded for task-related reasons (the participant misunderstood button selection when answering questions in the first block of the experiment). All participants reached at least 75% accuracy on the comprehension questions. Mean reading times in the region of interests are reported in Appendix B and illustrated in Figure 3 (Q1) and Figure 4 (Q2). Finally, mean accuracy and reaction time data for the temporal decision task are illustrated in Figure 5.

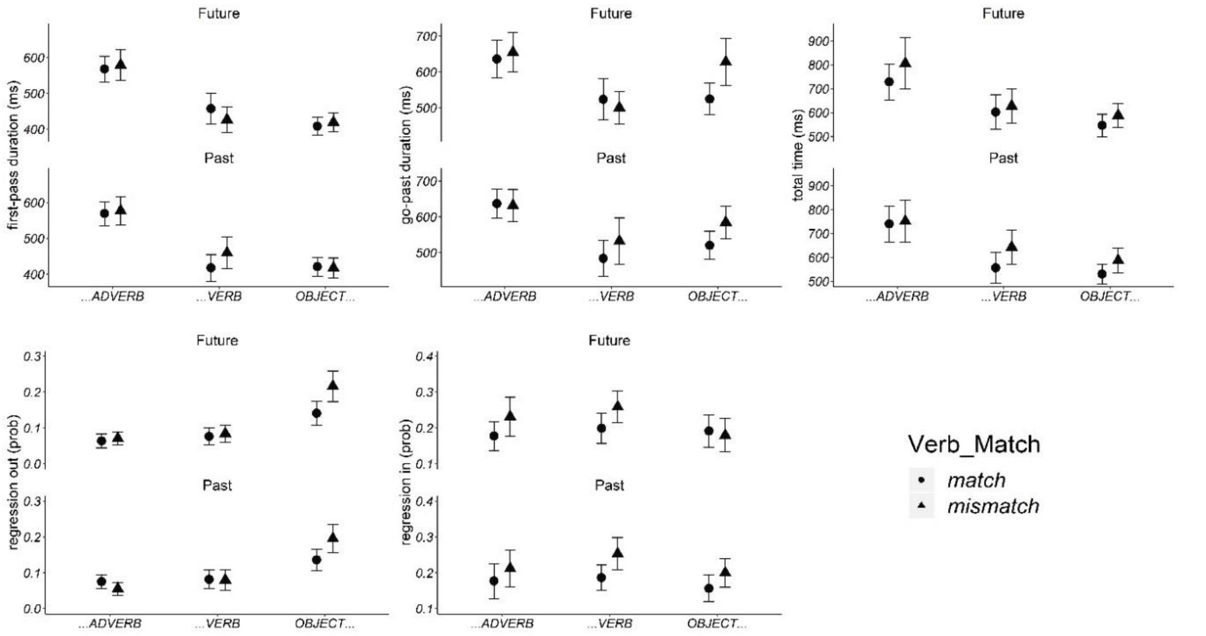

Figure 3. Average reading times and probability of regression (bars represent standard errors) related to Question1

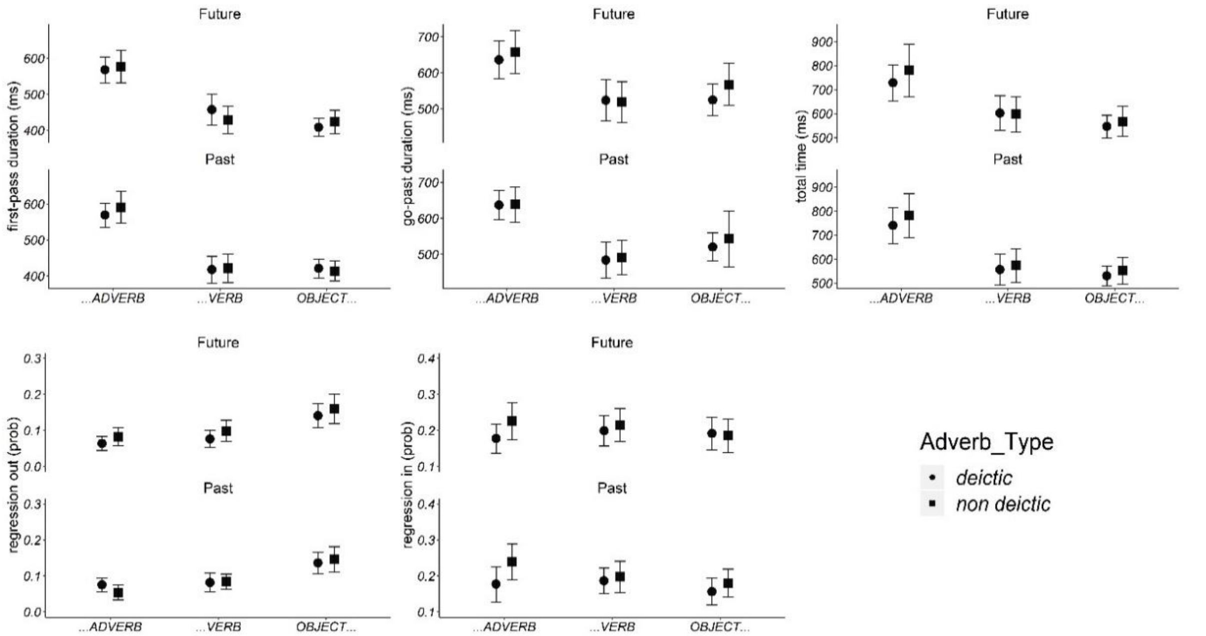

Figure 4. Average reading times and probability of regression (bars represent standard errors) related to Question2.

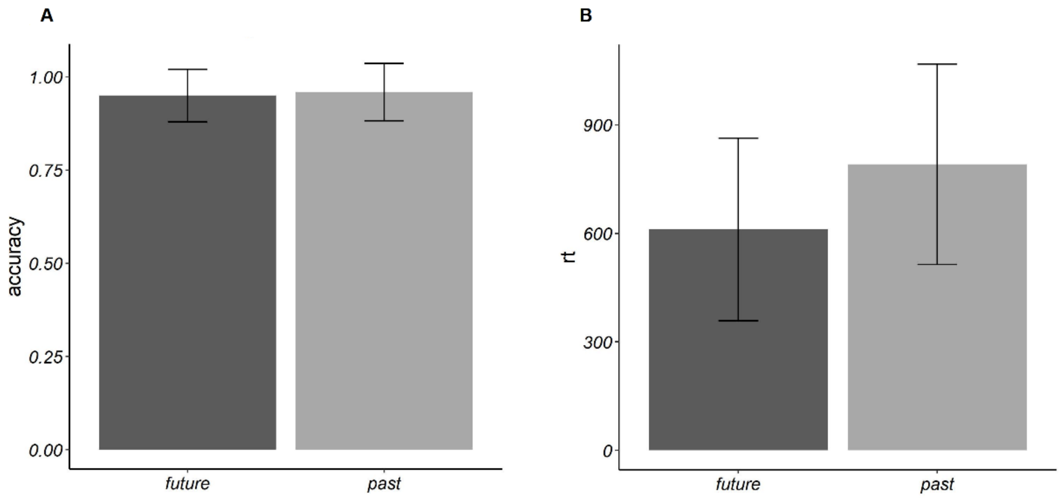

Figure 5. Average accuracy and response times (bars represent standard errors) in the temporal decision task

Eye-tracking data (Q1). The model estimates, standard errors, t- and p-values for each analysis are reported in Table 5. Estimates from the models fitted to resolve an interaction (interaction models, henceforth) are reported within the text.

In the verb region. The analysis of the first-pass duration revealed a time reference x verb match interaction. The interaction model showed that while past time reference violations triggered the expected mismatch effect, i.e., longer reading times in the violation than the control condition (Intercept: 5.96; Estimate: .07, SE: .02, $t = 3.55$, $p < .0045$), the opposite effect was found for future time reference conditions, that is, longer reading times in the match than the mismatch condition (Intercept: 5.96; Estimate: -.06, SE: .02, $t = -3$, $p < .0045$). The analysis of the go-past measure also showed a significant time reference x verb match interaction. The interaction model showed that, in this measure, only past conditions showed a significant mismatch effect (Intercept: 6.07; Estimate: .02, SE: .03, $t = 3.62$, $p < .0045$), while the effect of mismatch in future conditions did not reach significance after applying Bonferroni correction (Intercept: 6.07; Estimate: -.05, SE: .02, $t = -2.42$, $p = .018$). A marginal time reference x verb match interaction was also found in the analysis of total time. This marginal interaction was arguably related to the fact that the effect of mismatch was numerically larger for past than for future conditions. Finally, the analysis of total reading time and probability of regression in revealed an effect of verb match, that is longer reading times and higher probability inside the verb region in presence of a mismatch.

In the post-verb region, a main effect of match was found in the analysis of go-past, total time and probability of regression out. In other words, readers spent more time rereading the post-target region, and previous parts of the sentence in the presence of past/future mismatches compared to their control counterparts.

In the adverb region, the analysis of total reading times in the adverb region showed a significant effect of verb match and a marginal time reference x verb match interaction. The marginal interaction was driven by a numerically larger mismatch effect for future than past conditions. The analysis of the probability of regression inside the adverb region also revealed an effect of verb match, so a higher probability of regression inside the adverb region given a temporal violation.

Eye-tracking data (Q2). The model estimates, as well as the standard errors, t-values, and p-values of each analysis are reported in Table 6.

In the verb region, the analysis of the first-pass duration showed only a marginal effect of time reference. This marginal effect was arguably driven by the numerically longer reading times in the verb region when the time reference was to the future than the past. No interaction with adverb type was found. A significant effect of time reference was instead found in the analysis of go-past duration and total reading time, i.e., larger reading times on the verb for future time reference compared to past time reference. No interaction with adverb type was found.

The other regions did not show any effects, except for a significant effect of adverb type in the probability of regression inside the adverb region. In particular, a larger probability of regressions was found when the adverb was non-deictic compared to deictic.

Temporal decision task. Accuracy data showed ceiling performance in the categorization of past and future verbs (above 95% for both conditions). The best-fitting logit model showed only a marginal effect of time reference, with slightly better accuracy for past compared to future categorization (Intercept: 3.58; Estimate: .26, SE: .14, $z = 1.9$, $p = .06$). The analysis of the reaction time data gave stronger results, showing that readers needed more time to categorize a verb as past than as future (Intercept: .06; Estimate: .11, SE: .02, $t = 6.94$, $p < .05$).

	Adverb				**Verb**				**Post-verb**			
	Estimate	SE	*t / z*	*p*	Estimate	SE	*t / z*	*p*	Estimate	SE	*t / z*	*p*
First-pass												
TR	-	-	-	-	.001	.001	.52	.60	-.001	.001	-.45	.65
VM	-	-	-	-	.0003	.001	.25	.80	.001	.002	.32	.75
TR x VM	-	-	-	-	-.01	.003	-4.45	**2.34e-05**	.01	.003	1.63	.11
Go-past												
TR	-	-	-	-	.0002	.001	.18	.86	.002	.002	1.12	.26
VM	-	-	-	-	.001	.001	.89	.37	.01	.002	7.49	**8.66e-14**
TR x VM	-	-	-	-	-.01	.004	-3.50	**.001**	.01	.004	1.57	.12
Pr. Reg. out												
TR	-	-	-	-	-.01	.13	-.09	.93	.10	.09	1.05	.29
VM	-	-	-	-	.05	.13	.39	.70	.52	.11	4.84	**1.3e-06**
TR x VM	-	-	-	-	.12	.37	.32	.75	.12	.21	.55	.59
Pr. Reg. in												
TR	.06	.10	.64	.53	.07	.08	.80	.43	-	-	-	-
VM	.32	.10	3.36	**.001**	.43	.10	4.31	**1.64e-05**	-	-	-	-
TR x VM	.13	.17	.76	.45	-.06	.19	-.31	.76	-	-	-	-
Total time												
TR	.01	.02	.64	.53	.02	.01	1.20	.23	.001	.002	.78	.44
VM	.04	.01	2.96	**.004**	.09	.02	5.53	**8.53e-07**	.01	.001	6.40	**2.97e-08**
TR x VM	.07	.03	2.76	.01	-.09	.03	-2.89	.01	-.001	.003	-.19	.85

Table 5. Summary of model estimates, standard errors, t-values (first-pass, go-past, total) and z-scores (probability of regression in, probability of regression out) for the data in the three regions of interest, related to question 1. TR stands for Time Reference, VM stands for Verb Match. Bolded values indicate significant effects at $\alpha = .0045$.

	Adverb				**Verb**				**Post-verb**			
	Estimate	SE	*t / z*	*p*	Estimate	SE	*t / z*	*p*	Estimate	SE	*t / z*	*p*
First-pass												
AT	-	-	-	-	-.03	.01	-2.19	.03	-.003	.02	-.18	.86
TR	-	-	-	-	.04	.02	2.68	.01	-.007	.02	-.47	.64
AT x TR	-	-	-	-	-.06	.03	-1.99	.05	.05	.03	1.52	.13
Go-past												
AT	-	-	-	-	-.01	.01	-.42	.67	.003	.002	1.59	.12
TR	-	-	-	-	.05	.02	3.17	**.002**	.001	.002	.87	.39
AT x TR	-	-	-	-	-.04	.03	-.96	.34	.01	.003	1.51	.13
Pr. Reg. out												
AT	-	-	-	-	.19	.14	1.29	.20	.13	.11	1.21	.23
TR	-	-	-	-	.05	.12	.40	.69	.17	.13	1.32	.19
AT x TR	-	-	-	-	.28	.25	1.12	.27	.05	.19	.25	.80
Pr. Reg. in												
AT	.39	.11	3.60	**.0003**	.10	.10	1.08	.28	-	-	-	-
TR	-.05	.10	-.52	.60	.11	.09	1.23	.22	-	-	-	-
AT x TR	-.09	.17	-.52	.60	.03	.17	.17	.86	-	-	-	-
Total time												
AT	.04	.02	1.44	.15	.01	.02	.50	.62	.02	.01	1.65	.11
TR	-.02	.01	-1.10	.28	.05	.02	3.25	**.002**	.02	.02	1.28	.20
AT x TR	.02	.03	.60	.55	-.03	.03	-.86	.39	.01	.03	.37	.71

Table 6. Summary of model estimates, standard errors, t-values (first-pass, go-past, total) and z-scores (probability of regression in, probability of regression out) for the data in the three regions of interest, related to Question 2. TR stands for Time Reference, AT stands for Adverb Type. Bolded values indicate significant effects at $\alpha = .0045$.

General Discussion

In this study we aimed to investigate the mechanisms underlying the processing of past and future time reference. In particular, we tested whether past and future violations are processed differently during online sentence reading; whether the presence of adverbials, which either do (deictic) or do not (non-deictic) provide a specific time reference, affects past and future processing; and whether past and future time reference processing differs in different linguistic scenarios, i.e., when verbs are presented in isolation or within a sentence. We address each of these points in the following sections.

Past and future time reference violations are processed differently, at early stages

We presented two alternative predictions about how different reference times (past and future) are processed in the presence of a temporal mismatch. If both the past and future are interpreted as temporally dislocated from the time of utterance, we expected to find similar parsing costs (from early reading measures onwards) for both past and future reference violations. Alternatively, if only past-time reference requires a discourse link from the time of utterance to the event time, while the future is locally bound, as hypothesized by PADILIH, we expected parsing costs to differ. More specifically, we expected to find early costs for past but not for future reference violations, suggesting the past poses greater processing difficulty than the future.

Our findings showed that responses to past and future violations differed for some measures and regions, but lead to similar reading patterns for others. On the one hand, violations of past and future temporal context by an anomalously inflected verb generated different effects on early measures at the target region. Past time reference violations gave rise to longer reading times compared to the control condition from the first pass through the verb onwards. Conversely, we found an effect of future mismatch in the predicted direction only during the re-reading of the verb (i.e., total time). On the other hand, past and future mismatches yielded similar re-reading and regressive effects at the verb (total time and probabilities of regression to the verb), in the spillover region (go-past, total time and probability of regression out) and at the adverb (total time, probability of regression in). Similar rereading mechanisms may thus be at play when dealing with a past/future reference violation regardless of verb tense, and these mechanisms involve both the adverb and the verb. We found that the mismatch of past time reference resulted in an early effect, while the mismatch of future time reference resulted in a later effect in line with previous ERP studies testing past and non-past violations (e.g. Baggio, 2008; Dragoy et al., 2012), and with the PADILIH predicting early effects for past violations due to the ease of processing non-past mismatching verbs that are locally bound. It should be noted, however, that this finding can also be interpreted as an effect of the temporal frame provided by the adverb, and not only of the verb tense (as postulated by PADILIH). It is possible that the time course of temporal adverb processing differs for the two types of time reference. On the one hand, past time frame processing is more complex because of discourse-linking and gives rise to immediate mismatch effects. On the other hand, the establishment of a future time frame may be shallower, because of the more abstract and alterable temporal representation associated with the future tense (see also Fyndanis et al., 2018), which could give rise to delayed mismatch effects. More research is needed to disentangle these two interpretations, but our findings on the role of adverbs presented in the following section suggest that time processing does not rely solely on the information provided by the verb.

Temporal adverbs play a role during time processing

The second aim of this study was to investigate the contribution adverbs make to the interpretation of temporal information on the verb, by contrasting the processing of tensed verbs preceded by deictic and non-deictic adverbs. While deictic adverbs specify a reference time that is unambiguously anchored either in the past or future, non-deictic adverbs do not provide any cues about the past or future location of the reference time. We argued that if deictic adverbs provided a facilitatory effect (i.e., shorter reading times at the verb) relative to non-deictic adverbs, this would constitute evidence that adverbs played a relevant role in time processing.

Our data did not provide evidence that deictic adverbs provided a more facilitatory effect at the verb position than non-deictic adverbs. Nevertheless, the current dataset did clearly show that the type of adverb affected regressive movements to the adverb region in the expected direction, i.e., there was a higher probability of regression to adverbs that provided undefined/ambiguous reference time compared to unambiguous deictic adverbs[3]. It should be noted that a similar regression effect was found during the processing of subject-verb "unagreement" relations in Spanish (Mancini et al. 2014, Experiment 1; Hurtado, 1986). Unagreement is the apparent mismatch between the person feature expressed on the verb and the subject noun, which is resolved by adopting an overall 1st person plural interpretation (*Los periodistas*$_{3.pl}$ *escribimos*$_{1.pl}$ *un artículo*, 'We journalists wrote an article'). In this context, the authors interpreted the greater probability of regressions to the subject that unagreement elicited (compared to standard agreement configurations where subject and verb's feature matched) as the parser's attempt to superimpose the marked 1st person plural onto the unmarked 3rd person plural feature on the subject noun and thus derive an overall interpretation of the relation. Regressions may thus mirror an attempt to assign a value/feature to a previously read constituent, when this contains unmarked (e.g. 3rd person) or undefined (non-deictic) information. Under this hypothesis, we interpret the greater probability of regressions inside

[3] In the naturalness judgement task, non-deictic sentences were rated as less natural compared to deictic sentences, so one licit question is whether naturalness affected regressions. We ran post-hoc analyses to add naturalness (by-item mean rating values) as a predictor of the best-fitting regression model. The goodness of fit of the model did not improve neither by adding naturalness ($\chi^2(1) = 2.32$, $p > 0.05$) nor by adding the naturalness x adverb type interaction ($\chi^2(1) = 2.64$, $p > 0.05$). We can thus conclude that this effect was not simply driven by sentences naturalness.

the non-deictic adverb region as an attempt to attribute a specific temporal value to an otherwise underspecified element, by overwriting the value extracted from the verb on the adverb. Interestingly, reinspection of the adverbial area occurs also after detection of an incongruence on the verb (following a deictic adverb). The fact that the regressions to the adverbial regions occur both when adverbs present undefined temporal specifications and when they explicitly specify future/past times suggests that the parser is indeed sensitive to the featural setup of the adverb, and that it uses this piece of information to guide mechanisms of reinspection and reanalysis.

Past/future time reference processing on verbs in isolation and within a sentence, and the role played by modality

The temporal decision task aimed at verifying whether a potential processing difference between past and future could be detected with verbs presented in isolation. In null-subject languages such as Spanish, isolated verbs provide a basic sentential context. Under the PADILIH, we expected the categorization of past verbs to require longer reaction times than the categorization of future verbs. Conversely, if both past and future required the dislocation of an event (to a time period either before or after the time of utterance), we hypothesized that past and future processing would rely on similar mechanisms and thus would exhibit no differences.

The analysis of accuracy data showed only a marginal difference between the past and future condition, due to slightly higher accuracy for the former compared to the latter condition. In other words, we did not find strong evidence for a past-future dissociation in our offline data. Conversely, the analysis of the reaction time data did support a past-future dissociation, in the direction predicted by PADILIH (e.g. Bastiaanse et al., 2011; Bastiaanse, 2013), as past verbs required more categorization time than future verbs. It thus seems that the ease/difficulty of processing the future/past manifests during the online interpretation of the tense features of the verb. Once the temporal information has been interpreted, it is equally easy to categorize a verb as past or future.

Interestingly, a future verb presented within a correct sentence and preceded by either a deictic or non-deictic adverb resulted in longer reading times than a past verb in rereading measures (go-past, total, probability of regression in). This finding goes against what is predicted by PADILIH. Given the results of the temporal decision task, however, we cannot claim that future tense processing is *per se* more difficult than past tense processing. This difficulty seems to be specifically related to processing the future within a sentential context. Why? One possible explanation could be the intrinsic nature of future time reference and its modal component. Cross-linguistic data show that this modal component is present not only in languages with transparent modal morphology, such as English (where the modal component is encoded in the auxiliary verb *will*), but may be a semantic universal across languages (Bochnack, 2019 and references therein), including Spanish. It is possible that the modal properties of future time reference are particularly active in a sentential context. When the verb is presented in isolation, readers are only asked to categorize an event as past or future; there is no context provided that would trigger the analysis of modality-related properties. Conversely, when the future verb is embedded in a sentential context, readers must comprehend the time of the event to evaluate the role and the intentions of the individuals participating in this event (subject, object) and the context in which this event is embedded (adjunct phrase provided at the beginning of the sentence). It is thus likely that modality plays a prominent role in a sentential context, adding complexity to future (compared to past) time reference processing. Modality was accorded a similar role by Fyndanis et al. (2018) in their interpretation of the past and future time reference production exhibited by aphasic patients. The authors suggested that patients' impairments related to future time reference production might be related to the difficulty of making reference to possible worlds and the involvement of more abstract representations in future compared to past time reference.

The different costs associated with the analysis of future reference, and the potential role of modality in shaping these effects, suggests that the parser flexibly adapts to different linguistic contexts for the analysis of the same linguistic aspect. Nevertheless, modality is a linguistic phenomenon that entails a complex set of information about the speaker's attitude towards the event described in the sentence, as well as a spectrum of truth possibilities related to that same event. Further research is thus necessary to clarify the potential interaction of modality with tense and other temporal elements within the sentence.

An unexpected result was the longer reading time during the first pass through the verb region found in the future control compared to the mismatch condition. This reversed effect could be due to an interaction between the modal properties of the future tense and temporal adverbs that provide deictic information in sentential contexts. We can positively exclude the conclusion that this effect was related to (i) processing a future verb *per se*, based on the data from the temporal decision task; and (ii) to processing future verbs in a sentence, independently of the type of adverb. A closer inspection of the data from both Analysis 1 and 2 reveals that processing future reference in correct sentences triggers an increase in reading times only if it is preceded by a deictic adverb. Indeed, compared to the non-deictic future and deictic/non-deictic past, deictic future conditions show numerically longer first-pass reading times (see Figure 4), which suggests that the modal properties of future time reference and the presence of a deictic adverb may interact leading to greater reading costs. Nevertheless, further investigation is needed to corroborate this hypothesis, for example by testing different types of future modality (expressed by different realizations of the future) in the presence of different types of temporal adverbs.

Time course of temporal processing during sentence reading

By tracking eye movements during the reading of sentences where the temporal specifications of adverbs and verbs were manipulated to match or mismatch, we were able to observe the time course of temporal information processing during sentence reading. Based on previous ERP and eye-tracking studies (Baggio, 2008; Biondo et al., 2018), we expected increased reading time at a verb whose tense mismatched the temporal specification of the preceding adverb, compared to a verb with matching information. In line with this prediction, we showed that a temporal mismatch between adverb and verb generates reading disruptions already beginning from the first pass through the verb.

The effect generated by this temporal mismatch is not only early, but also long-lasting. Mismatch effects were also reported in the analysis of the post-verb region, in particular in rereading measures (go-past duration, total reading time, probability of regression out), as well as in the analysis of the adverb region (probability of regression in, total reading time). Temporal violations were found to have similarly long-lasting effects in previous behavioral (e.g. Biondo, 2017; De Vincenzi, Rizzi, Portolan, Di Matteo, Spitoni & Di Russo, 2006) and event-related potential (e.g. Dragoy et al., 2012; Bos et al., 2013) studies. When the information provided by the verb does not match the time frame provided by the temporal adverb, the reading process is rapidly disrupted and the complex set of information that the parser needs to reconsider for reanalysis—from the morphosyntactic to the semantic and discourse information conveyed by the verb tense and the adverb—leads to a persistent mismatch effect.

Critically, the earliness of the temporal mismatch effect is relevant for existing models of sentence comprehension. The rapidity with which this effect emerges is at odds with traditional syntax-first models, which consider temporal information to be semantic in nature and thus managed by the parser only at later stages of processing (e.g. Frazier, 1978; Friederici, 2002).

Conversely, the earliness of time processing is compatible with models that predict both syntactic and extra-syntactic (e.g., semantic) factors play a role either during the early stages of processing of any type of linguistic relation (e.g. Hagoort, 2013; 2016) or specifically in the processing of non-primary relations, such as the one between the temporal adverb and the verb (e.g. Frazier & Clifton, 1996). Indeed, it is possible that both syntactic and non-syntactic information is considered during the association/construal stage, in which the non-primary constituent (adverb) is just associated with the current thematic domain, while waiting for the theta assigner (the verb) to be finally attached, as predicted by the Construal model (Frazier & Clifton, 1996). How this non-/syntactic processing unfolds over time is not specified in the Construal model. Biondo and colleagues (2018) proposed that the time course of the interpretation of temporal information is affected by the distance between the adverb and the verb. More specifically, the temporal information provided by the adverb is used from early stages of verb processing if the parser has had enough time to "unpack" and link the information provided by the adverb with the discourse, as happens in distal adverb-verb configurations like those adopted in this study.

Our finding is also potentially compatible with models of semantic composition, such as the Enlightened Incrementality Conjecture (EIC; Beck & Tiemann, 2018), which predicts that the time course of language comprehension depends on the domains that are involved in the composition of sentence meaning, namely the lexical, inflectional, and discourse layers. In this account, the processing of units that belong to the same domain happens early on (incrementally). This would be the case for temporal adverbs and verbs, which the EIC locates in the inflectional layer of the sentence.

Conclusion

The past is immutable and beyond our control, while the future encompasses events that have yet to occur and are therefore still subject to alterations (Comrie, 1985). In this study we investigated how readers comprehend sentences and categorize verbs that express these two temporal and conceptual dimensions, showing that both similar and different mechanisms are at work. Overall, our results add to existing psycholinguistic studies and bridge gaps left by techniques that relied on a less ecological presentation of materials or a less precise temporal resolution. By combining a behavioral detection task with eye movement recording during reading, we showed that the analysis of past and future temporal reference is a multifaceted process that arguably is not limited to the presence/absence of discourse linking, as suggested by PADILIH.

Consistent with PADILIH, past time processing was found to be more complex than future time processing, both when past verbs were presented in isolation and when past temporal mismatches were introduced within the sentence. However, we also found that future reference processing within a congruous sentential context required more effort than past time reference processing, differing from predictions based on PADILIH. We speculated that the extra time needed for future processing in sentences could be due to the modal properties of future reference and the possible, alterable worlds these imply; such aspects are clearly active in sentences but not in verbs in isolation. However, we note that more work is needed to explore the role of modality during sentence processing.

Moreover, eye movement tracking allowed us to examine the relative role of verbs and adverbs, which previous ERP studies could not do. In particular, we were able to show that (i) temporal information on verbs is accessed and integrated early during processing, and that (ii) in spite of their structural optionality, adverbs contribute information that guides the parser's reinspection and reanalysis mechanisms when ambiguities or incongruences occur.

In sum, this study has provided a wide overview of the time course and mechanisms involved in temporal processing. Future eye-tracking studies could further investigate past and future processing from the perspective of language development and language disorders, to assess whether and how similar dissociations emerge.

Acknowledgments

This research is supported by the Basque Government through the BERC 2018-2021 program, by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490. S.M. was supported by grants RYC-2017-22015, FFI2016-76432-P_LAMPT (Ministerio de Economía y Competitividad, Agencia Estatal de Investigación & Fondo Europeo de Desarrollo Regional) and partially by grants IN[18_HMS_LIN_0058 (BBVA Foundation) and PIBA_2020_I_0024 from the Basque Government.

References

Alemán Bañón, J., & Rothman, J. (2019). Being a participant matters: Event-related potentials show that markedness modulates person agreement in Spanish. *Frontiers in psychology*, *10*, 746.

Alexiadou, A. (1997). *Adverb placement: A case study in antisymmetric syntax* (Vol. 18). John Benjamins Publishing.

Anelli, F., Peters-Founshtein, G., Shreibman, Y., Moreh, E., Forlani, C., Frassinetti, F., & Arzy, S. (2018). Nature and nurture effects on the spatiality of the mental time line. *Scientific reports*, *8*(1), 1-9.

Aronson, H. I. (1977). Interrelationships between aspect and mood in Bulgarian. *Folia slavica*, *1*(1), 9-32.

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in cognitive sciences*, *5*(12), 533-539.

Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's aphasics. In *Language and the brain* (pp. 295-313). Academic Press.

Baggio, G. (2008). Processing temporal constraints: An ERP study. *Language Learning*, *58*, 35-55.

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.

Bastiaanse, R., Bamyaci, E., Hsu, C. J., Lee, J., Duman, T. Y., & Thompson, C. K. (2011). Time reference in agrammatic aphasia: A cross-linguistic study. *Journal of Neurolinguistics*, *24*(6), 652-673.

Bastiaanse, R. (2013). Why reference to the past is difficult for agrammatic speakers. *Clinical linguistics & phonetics*, *27*(4), 244-263.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Beck, S., & Tiemann, S. (2018). Towards a model of incremental composition. In *Proceedings of Sinn und Bedeutung* (Vol. 21, No. 1, pp. 143-162).

Biondo, N. (2017). *The Syntactic Side of Time: Processing Adverb-Verb Temporal Agreement* (Doctoral dissertation, University of Trento).

Biondo, N., Vespignani, F., Rizzi, L., & Mancini, S. (2018). Widening agreement processing: a matter of time, features and distance. *Language, Cognition and Neuroscience*, *33*(7), 890-911.

Bochnak, M. R. (2019). Future reference with and without future marking. *Language and Linguistics Compass*, *13*(1), e12307.

Boland, J. E., & Blodgett, A. (2006). Argument status and PP-attachment. *Journal of psycholinguistic research*, *35*(5), 385-403.

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: evidence for a mental time line. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2257-2273.

Bos, L. S., Dragoy, O., Stowe, L. A., & Bastiaanse, R. (2013). Time reference teased apart from tense: Thinking beyond the present. *Journal of Neurolinguistics*, *26*(2), 283-297.

Bos, L. S., Hanne, S., Wartenburger, I., & Bastiaanse, R. (2014). Losing track of time? Processing of time reference inflection in agrammatic and healthy speakers of German. *Neuropsychologia*, *65*, 180-190.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211-243.

Boyer, P. (2008). Evolutionary economics of mental time travel?. *Trends in cognitive sciences*, *12*(6), 219-224.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, *1*(1).

Clifton Jr, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In *Eye movements* (pp. 341-371). Elsevier.

Collart, A., & Chan, S. H. (2021). Processing past time reference in a tenseless language: An ERP study on the Mandarin aspectual morphemes-le and-guo. *Journal of Neurolinguistics*, *59*, 100998.

Comrie, B. (1985). *Tense* (Vol. 17). Cambridge university press.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in psychology*, *8*, 522.

De Vincenzi, M., Rizzi, L., Portolan, D., Di Matteo, R., Spitoni, G., & Di Russo, F. (2006). Mapping the language: A reading time and topographic ERP study on tense, agreement, and Aux-V violations. *Ms., Univerity of Chieti*.

Dragoy, O., Stowe, L. A., Bos, L. S., & Bastiaanse, R. (2012). From time to time: Processing time reference violations in Dutch. *Journal of Memory and Language*, *66*(1), 307-325.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior research methods*, *45*(4), 1246-1258.

Enç, M. (1987). Anchoring conditions for tense. *Linguistic inquiry*, 633-657.

Faroqi-Shah, Y., & Dickey, M. W. (2009). On-line processing of tense and temporality in agrammatic aphasia. *Brain and Language*, *108*(2), 97-111.

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. *Doctoral dissertation, University of Connecticut*.

Frazier, L., & Clifton, C. (1996). *Construal*. Mit Press.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, *6*(2), 78-84.

Fyndanis, V., Arcara, G., Capasso, R., Christidou, P., De Pellegrin, S., Gandolfi, M., Messinis, L., Panagea, E., Papathanasopoulos, P., Smania, N., Semenza, C., & Miceli, G. (2018). Time reference in nonfluent and fluent aphasia: A cross-linguistic test of the PAst DIscourse LInking Hypothesis. *Clinical Linguistics & Phonetics*, *32*(9), 823-843.

Gibson, Edward, Steve Piantadosi, and Kristina Fedorenko. "Using Mechanical Turk to obtain and analyze English acceptability judgments." *Language and Linguistics Compass* 5.8 (2011): 509-524.

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), 493-498.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in psychology*, *4*, 416.

Hagoort, P. (2016). MUC (Memory, Unification, Control): A model on the neurobiology of language beyond single word processing. In *Neurobiology of language* (pp. 339-347). Academic Press.

Hurtado, A. (1985). The unagreement hypothesis. In *Selected papers from the thirteenth linguistic symposium on romance languages* (pp. 187-211). Amsterdam: John Benjamins.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434-446.

Kizach, J. (2014). Analyzing Likert-scale data with mixed-effects linear models: a simulation study. Poster presented at meeting Linguistic Evidence 2014, Tübingen, Germany.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(13), 1-26.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, *44*(2), 325-343.

Mancini, S., Molinaro, N., Davidson, D. J., Avilés, A., & Carreiras, M. (2014). Person and the syntax–discourse interface: an eye-tracking study of agreement. *Journal of Memory and Language*, *76*, 141-157.

Martínez-Ferreiro, S., & Bastiaanse, R. (2013). Time reference in Spanish and Catalan non-fluent aphasia. *Lingua*, *137*, 88-105.

Martorell, J. (2018). Merging generative linguistics and psycholinguistics. *Frontiers in psychology*, *9*, 2283.

McCormack, T., & Hoerl, C. (2017). The development of temporal concepts: Learning to locate events in time. *Timing & Time Perception*, *5*(3-4), 297-327.

McCormack, T. (2014). Three types of temporal perspective: characterizing developmental changes in temporal thought. *Annals of the New York Academy of Sciences*, *1326*(1), 82-89.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain research*, *1164*, 81-94.

Partee, B. H. (1973). Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy*, *70*(18), 601-609.

Qiu, Y., & Zhou, X. (2012). Processing temporal agreement in a tenseless language: An ERP study of Mandarin Chinese. *Brain research*, *1446*, 91-108.

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, *67*(8), 603.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038.

Smith, C. S. (1981). Semantic and syntactic constraints on temporal interpretation. In *Tense and Aspect* (pp. 213-237). Brill.

Szpunar, K. K. (2010). Episodic future thought: An emerging concept. *Perspectives on Psychological Science*, *5*(2), 142-162.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans?. *Behavioral and brain sciences*, *30*(3), 299-313.

Tutunjian, D., & Boland, J. E. (2008). Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass*, *2*(4), 631-646.

Ulrich, R., & Maienborn, C. (2010). Left–right coding of past and future in language: The mental timeline during sentence processing. *Cognition*, *117*(2), 126-138.

Vasishth, S., von der Malsburg, T., & Engelmann, F. (2012). What can eye-tracking tell us about sentence processing. *WIREs Cognitive Science.*

Von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of memory and language*, *94*, 119-133.

Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review*, *15*(2), 426-430.

Zagona, K. (2003). Tense and anaphora: Is there a tense-specific theory of coreference. *Anaphora: A reference guide*, 140-171.

Zagona, K. (2013). Tense, aspect and modality. *The Cambridge handbook of generative syntax*, 746-792.

Zhang, M., & Hudson, J. A. (2018). Children's understanding of yesterday and tomorrow. *Journal of experimental child psychology*, *170*, 107-133.

Appendix A

Best-fitting models for each research question, region and measure.

Question 1

Adverb

log(tt) ~ 1 + timeref * vmatch + (1 + timeref:vmatch || sj) + (1 + timeref * vmatch || item)

ri ~ 1 + timeref * vmatch + (1 + vmatch || sj) + (1 + timeref + vmatch || item)

Verb

log(fp) ~ 1 + timeref * vmatch + (1 + timeref || sj) + (1 + timeref + timeref:vmatch || item)

log(gp) ~ 1 + timeref * vmatch + (1 + timeref:vmatch || sj) + (1 + timeref || item)

log(tt) ~ 1 + timeref * vmatch + (1 + vmatch + timeref:vmatch || sj) + (1 + timeref || item)

ri ~ 1 + timeref * vmatch + (1 + vmatch || sj) + (1 + vmatch + timeref:vmatch || item)

ro ~ 1 + timeref * vmatch + (1 + timeref:vmatch | sj) + (1 + vmatch + timeref:vmatch || item)

Post-verb

log(fp) ~ 1 + timeref * vmatch + (1 + vmatch || sj) + (1 + timeref:vmatch || item)

log(gp) ~ 1 + timeref * vmatch + (1 | sj) + (1 + timeref + timeref:vmatch || item)

log(tt) ~ 1 + timeref * vmatch + (1 + vmatch || sj) + (1 + timeref || item)

ro ~ 1 + timeref * vmatch + (1 + vmatch || sj) + (1 + vmatch + timeref:vmatch || item)

Question 2

Adverb

log(tt) ~ 1 + advtype * timeref + (1 + advtype + advtype:timeref | sj) + (1 + advtype * timeref || item)

ri ~ advtype * timeref + (1 | sj) + (1 + advtype + timeref || item)

Verb

log(fp) ~ 1 + advtype * timeref + (1 + advtype + advtype:timeref || sj) + (1 + timeref + advtype:timeref || item)

log(gp) ~ 1 + advtype * timeref + (1 + advtype * timeref || sj) + (1 | item) log(tt) ~ 1 + advtype * timeref + (1 + advtype + advtype:timeref || sj) + (1 + timeref + advtype:timeref || item)

ri ~ advtype * timeref + (1 + advtype || sj) + (1 | item)

ro ~ advtype * timeref + (1 + advtype || sj) + (1 + advtype + advtype:timeref || item)

Post-verb

log(fp) ~ 1 + advtype * timeref + (1 + advtype + advtype:timeref | sj) + (1 + timeref || item)

log(gp) ~ 1 + advtype * timeref + (1 + advtype + timeref || sj) + (1 + timeref || item)

log(tt) ~ 1 + advtype * timeref + (1 + advtype + advtype:timeref || sj) + (1 + timeref || item)

ro ~ advtype * timeref + (1 + advtype + timeref || sj) + (1 + timeref | item)

Appendix B

		Adverb	Verb	Post-verb
First-pass				
Future	Match	568 (37)	457 (42)	409 (25)
Future	Mismatch	580 (43)	427 (36)	420 (26)
Past	Match	569 (34)	417 (37)	421 (26)
Past	Mismatch	577 (39)	459 (44)	418 (28)
Go-past				
Future	Match	636 (52)	523 (58)	525 (44)
Future	Mismatch	655 (56)	500 (46)	628 (65)
Past	Match	637 (40)	484 (50)	520 (39)
Past	Mismatch	632 (45)	532 (65)	585 (46)
Total time		-	-	-
Future	Match	728 (75)	602 (73)	546 (47)
Future	Mismatch	806 (106)	627 (73)	588 (50)
Past	Match	740 (76)	556 (65)	531 (41)
Past	Mismatch	752 (87)	643 (71)	588 (53)

Table A1. Mean reading times and standard errors (in brackets) of the eye-tracking measures related to Question 1.

		Adverb	Verb	Post-verb
First-pass				
Deictic	Future	568 (37)	457 (42)	409 (25)
Deictic	Past	569 (34)	417 (37)	421 (26)
Non-deictic	Future	578 (45)	429 (38)	423 (33)
Non-deictic	Past	591 (44)	421 (40)	414 (28)
Go-past		-	-	-
Deictic	Future	636 (52)	523 (58)	525 (44)
Deictic	Past	637 (40)	484 (50)	520 (39)
Non-deictic	Future	657 (60)	518 (57)	567 (58)
Non-deictic	Past	638 (49)	491 (48)	543 (77)
Total time				
Deictic	Future	728 (75)	602 (73)	546 (47)
Deictic	Past	740 (76)	556 (65)	531 (41)
Non-deictic	Future	781 (109)	597 (73)	567 (63)
Non-deictic	Past	781 (92)	573 (71)	552 (55)

Table A2. Mean reading times and standard errors (in brackets) of the eye-tracking measures related to Question 2.

Supplementary Material

Sample size was determined a-priori using the *simR* package (Green & MacLeod, 2016) to run simulations (Brysbaert & Stevens, 2018). Power was calculated over a range of different sample sizes through 200 simulations on real first-pass data from Biondo et al. (2018), who manipulated and tested similar adverb-verb mis/match configurations. The output showed that more than 50 participants were necessary to achieve the 80% power threshold required to reliably detect a violation effect from early measures onwards, as shown in the plot below.