Language and Cognition (2026), 18, e2, 1–23
doi:10.1017/langcog.2025.10050

ARTICLE

The test of time: experimentally recreating the reanalysis of FINISH as a recent past marker

Julia Heine, Martín Fuchs and Malte Rosemeyer

Institute for Romance Philology, Freie Universität Berlin, Germany
Corresponding author: Julia Heine; Email: julia.heine@fu-berlin.de

(Received 26 June 2025; Revised 06 October 2025; Accepted 19 November 2025)

Abstract

In grammaticalization studies, *reanalysis* is understood as the assignment of new meaning to formally unchanged elements, supported by bridging contexts compatible with the old and the reanalyzed meaning. The *source determination hypothesis* (SDH) predicts that parallel grammaticalization trajectories occur crosslinguistically, as similar source meanings give rise to similar inferences. One such pattern is the development of recent past markers from FINISH constructions. While grammaticalization pathways are well-documented cross-linguistically, the SDH has never been tested experimentally. In this study, we examine whether modern English speakers are sensitive to inferences arising from a bridging context identified as relevant to the grammaticalization of Old Spanish FINISH into a recent past marker. In a temporal distance judgment task, we examined whether the bridging context identified for Old Spanish facilitates an inference of temporal immediacy in contemporary English, where the construction has not been grammaticalized. In line with the SDH, the bridging context enhanced an inference of immediacy in contemporary English (Exp. 1), with specific grammatical features of the source determining its strength (Exp. 2). These results not only demonstrate the viability of testing hypotheses about language change using experimental pragmatics but also call for empirically refining the concept of *source determination*.

Keywords: bridging context; grammaticalization; historical linguistics; language change; psycholinguistics; reanalysis; semantic extension; source determination; temporal inference

1. Introduction

Grammaticalization refers to the process by which lexical elements or combinations of lexical and grammatical elements develop grammatical meaning (Bybee et al., 1994, p. 4). For instance, *going to* + infinitive, which originally expressed allative motion, has acquired the grammatical function of future tense in sentences like *I'm going to think about it*, where a motion reading is dispreferred. Historical linguistics assumes that grammaticalization involves *reanalysis*, the mapping of a new meaning

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

onto a superficially unchanged construction, and that reanalysis is supported by so-called *bridging contexts*, which allow for both the original meaning and the new meaning (Enfield, 2003; Evans & Wilkins, 2000; Heine, 2002). This situation in which both meanings of a given grammatical marker can coexist is supported by one of the principles of grammaticalization, referred to as 'layering' (Hopper, 1991, p. 22). For instance, in Hopper and Traugott's (2003, p. 88) classic example *I am going to be married*, both a sense of movement with intention or purpose ('I am moving towards a place with the purpose of getting married') and a sense that indicates an event that will occur in the near future ('I will be married soon') are potential sentence interpretations (see also Eckardt, 2006, Chapter 4, for an extensive discussion of the reanalysis of *going to* + infinitive as a future marker).

Most work on the relationship between bridging contexts and reanalysis stems from longitudinal diachronic analyses, which propose that bridging contexts are attested before a construction is used regularly with the new meaning (Heine, 2002, p. 86; Diewald, 2006, p. 4; Traugott & Trousdale, 2013, p. 199). Given that chronology does not necessarily imply causation, further proof of the facilitating effect of bridging contexts on reanalysis is required.

In this paper, we investigate the facilitating effect of bridging contexts on reanalysis using language comprehension experiments. We thereby test the *source determination hypothesis* (SDH) developed in diachronic typology (Bybee et al., 1994), which suggests that the source meaning of a grammatical marker or construction fundamentally constrains the grammaticalization path followed by the construction. Crucially for our study, the SDH predicts that grammaticalization processes attested in one language should be able to occur in the future in another language where a cognate lexical item has not developed that grammatical function.

We investigate this hypothesis through the grammaticalization of FINISH constructions, which frequently undergo a change from expressing a purely aspectual, completive meaning into becoming recent past markers – a change attested in numerous typologically diverse languages. To this end, we examine whether speakers of Present-Day English derive a recent past interpretation for the *finish* + gerund construction from exposure to a bridging context that previous studies have identified to play a role in the grammaticalization of the analogous construction *acabar* + *de* + infinitive in Old Spanish (Rosemeyer & Grossman, 2017, 2021). Our results suggest that participants are indeed more likely to attribute a recent past interpretation to *finish* + gerund in this bridging context than in other usage contexts. Thereby, we additionally assess the viability of testing hypotheses about language change using experimental pragmatics (Grossman & Noveck, 2015).

1.1. The source determination hypothesis

In the study of grammaticalization phenomena, the SDH, as proposed by Bybee et al. (1994), has had a major role. They argue that the (lexical) source meaning of a grammatical marker or construction fundamentally constrains the grammaticalization path – the diachronic trajectory – that the construction follows. This premise has structured our understanding of grammaticalization processes across languages, suggesting that similar source meanings tend to follow similar paths of semantic development due to the inferential patterns that in particular the lexical items constituting them can trigger.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

The SDH emerged from extensive crosslinguistic studies of grammaticalization pathways, particularly in the domains of tense, aspect and modality, where Bybee and her colleagues (1994) documented recurring patterns in how lexical items with specific semantic properties evolve into particular grammatical categories. For instance, they observed that verbs of desire (WANT)[1] commonly develop into future tense markers across unrelated languages, while verbs of obligation (MUST and OUGHT) frequently evolve into markers of epistemic necessity.

That said, grammaticalization often involves not just content words but also components that are already part of the grammar, like tense, aspect or case markers and prepositions, a process known as 'secondary grammaticalization' (Hopper & Traugott, 2003). These grammatical elements help shape the meaning of the larger construction. Moreover, the way such elements are arranged – either in relation to one another or in relation to the word or phrase they modify – can play a role in determining the construction's meaning. As Bybee et al. (1994, p. 11) put it, "for this reason, in tracing the origin of grammatical meaning, we must attend to the syntax and morphology of the source construction and not simply to the referential meaning of its lexical items".

This systematic relationship between sources and grammatical outcomes suggests that grammaticalization is not a random process but is guided by semantic constraints inherent in the source constructions themselves and by more general cognitive constraints in the minds of speakers. Thereby, the SDH became a significant theoretical advancement in our understanding of semantic change, helping to partially explain some of its (crosslinguistic) regularities (Traugott & Dasher, 2001).

1.2. The FINISH construction and its development in Spanish

As a further example of a grammaticalization process, consider the pathway by which verbal periphrases meaning FINISH develop into anterior or recent past markers – a crosslinguistically attested pattern documented by Bybee et al. (1994) in languages as diverse as Sango (Central Africa), Mwera (Tanzania), Tok Pisin (Papua New Guinea; Bybee et al., 1994, pp. 69–71) or Spanish (Olbertz, 1998; Veyrat Rigat, 1994; Yllera, 1980). This grammaticalization pathway relies on an inferential link between the completion of an event and its temporal anteriority: When a speaker indicates that an event has been 'finished', they simultaneously imply that this event occurred prior to reference time. Over time, this temporal inference can become conventionalized, allowing the FINISH construction to be reanalyzed as a marker of anterior tense or recent past.

In their 2017 study, Rosemeyer and Grossman conducted a detailed diachronic analysis of the Spanish *acabar* + *de* + infinitive construction in corpus texts from the thirteenth century to the eighteenth century, tracing its semantic evolution from a marker of completion in the thirteenth century to a grammaticalized expression of recent past after the fifteenth century. Their research demonstrated how this construction, originally meaning 'to finish doing something' (1), gradually acquired the

[1]Throughout this paper, we make use of all capitals to indicate crosslinguistic grams. For particular linguistic forms, we use the standard convention of capitalizing the first letter of the category (e.g., English Progressive). Italics indicate specific constructions in the languages of interest.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

meaning 'to have just done something' (2), exemplifying precisely the grammaticalization pathway proposed by Bybee et al. (1994) for FINISH constructions.

(1) *Otrossí far-ás la fiesta de las tienda-s*
furthermore make-FUT.2SG DET.DEF.F.SG party of DET.DEF.F.PL shop-PL
cuando acabar-edes de coger el pan e
when finish-FUT.SUBJ.2SG of get DET.DEF.M.PL bread and
el vino
DET.DEF.M.PL wine
"Furthermore, you will have the Feast of Tabernacles when you finish getting the bread and the wine" (*General Estoria*, 13th c.; example [22] in Rosemeyer & Grossman, 2017)

(2) *Juan acab-a de comprar un coche nuevo*
Juan finish-PRS.3SG of buy DET.INDF.M.SG car new
'Juan has just bought a new car'

Rosemeyer and Grossman (2017) used corpus data spanning several centuries and identified the specific bridging contexts that facilitated this reanalysis in Spanish: past-of-past contexts with informationally redundant infinitives.

In Early Spanish, the *acabar* + *de* + infinitive construction followed a distribution partially governed by the informativity of the infinitive: When the activity being finished was uninformative (i.e., predictable from context), the infinitive was typically omitted (3), whereas when it provided information that could not easily be retrieved from context, it was expressed (4).

(3) *e acab-ó la torre dal=faro que*
and finish-PST.PFV.3SG DET.DEF.F.SG tower of.the=light.house that
començ-a-ra Hercules
begin-THEME-PST.IPFV.SBJ.3SG Hercules
"And he finished (building) the tower of the lighthouse that Hercules had begun (building)" (*EDE*, 13th c.; example [31] in Rosemeyer & Grossman, 2017)

(4) *aquel anno acab-ó de cercar Toda la*
DEM.DEF.M.SG year finish-PST.PFV.3SG of surround all DET.DEF.F.SG
uilla de Roma de muro-s muy fuerte-s
city of Rome of wall-PL very strong-PL
"That year he finished surrounding the entire city of Rome with very strong walls" (*EDE*, 13th c.; example [32] in Rosemeyer & Grossman, 2017)

It can be observed in (3) that the event expressed by the *acabar* construction is stereotypical, since the preferred interpretation of *acabar la torre* ('finish the tower') involves the *building* of the tower, so that the infinitive can be omitted. On the other hand, in (4), the expectation is that *acabar la villa de Roma* ('finish the city of Rome') would also involve its *building*, but since the author wants to express that the city was *surrounded*, the infinitive is expressed overtly.

Despite this general distribution in terms of informativity – conditioning the omission of infinitives in the Spanish construction – Rosemeyer and Grossman (2017) also identified specific temporal subordinate contexts where speakers used

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

an overt infinitive despite its informational redundancy. This apparent violation of the *Maxim of Quantity* ('Do not make your contribution more informative than is required'; Grice, 1975, p. 45; Levinson, 2000, p. 37) created an implicature that highlighted the temporal sequence between events and therefore underscored its temporal immediacy. According to Rosemeyer and Grossman (2017), these contexts served as the crucial bridging contexts for the hearer-based reanalysis of the construction as an anterior marker. For example, in (5), the infinitive *fazer* ('to make') does not add information that could not be derived from the nominal involved in the *acabar* construction – *acabar la tienda* ('finish the tent') could be understood as 'finish building/putting up the tent' without an overt infinitive.

(5) *Cuando Moisén acab-ó la tienda de fazer e*
when Moses finish-PST.PFV.3SG DET.DEF.F.SG tent of make and
la alç-ó… ofrec-ie-ron… muchas
DET.DEF.F.SG raise-PST.PFV.3SG offer-THEME-PST.PFV.3PL many
ofrenda-s
sacrifice-PL
"When Moses had finished erecting the tent and raised it…, they … offered many sacrifices" (*General Estoria*, 13th c; example [34] in Rosemeyer & Grossman, 2017)

Rosemeyer and Grossman (2017) argue that in the specific sequence of events, where both Event 1 in the temporal subordinate clause with an overt infinitive (*acabó de fazer* in [5]) and Event 2 in the main clause (*alçó* in [5]) refer to the past, speakers may infer that Event 2 (immediately) followed Event 1, which is consequently conceptualized as having occurred earlier. Therefore, although in (5) *acabó* is formally a perfective preterite tense form, it is analogous to a pluperfect in terms of the temporal situation. As we shall see below, we contend that this ambiguity contributes to the function of (5) as a bridging context.

1.3. The current study

Although FINISH constructions have undergone comparable grammaticalization processes in numerous typologically diverse languages, Rosemeyer and Grossman (2021) highlight that this pattern is notably absent in many European languages, such as English. The SDH, however, would suggest that similar sources, and lexical sources in particular, should allow for parallel grammaticalization pathways in those languages.

In English, *finish* is indeed available as an aspectual verb marking completion of the expressed eventuality (6) – being part of a construction that is comparable to that of Old Spanish, even though the full equivalent construction in English includes a gerund rather than an infinitive.

(6) Jane finished mowing the lawn.

Additionally, this English construction can be part of a bridging context such as the one identified for Old Spanish. For instance, like in the Old Spanish example (5), the English example (7) highlights sequentiality between the two expressed events mowing the lawn (Event 1) and drinking coffee (Event 2).

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

(7) When Jane finished mowing the lawn, she drank coffee.

Like *acabó* in (5), *finished* is formally a (perfective) preterite but refers to a situation placed in time before Event 2. As a result, (7) could also be formed using the pluperfect *had finished*, as in (8).

(8) When Jane had finished mowing the lawn, she drank coffee.

However, the two sentences differ in the way the temporal relationship between the events is portrayed. While the version with the Simple Past presents the two events as a sequence, the temporal order of events is only implied by juxtaposition. By contrast, the version with the pluperfect makes the succession of events explicit, emphasizing that the mowing was fully completed before the coffee drinking took place. In this way, (8) highlights the temporal boundary and dependency between the events more strongly than (7).

From the perspective of grammaticalization theory, (7) provides a better possible bridging context for the reanalysis of the English FINISH construction as a tense marker than (8), since only in (7) the temporal succession remains implicit and must be inferred by the hearer, leaving enough ambiguity for reanalysis to occur and for contextual information to shape the interpretation of the temporal sequencing.

Considering that both the lexical material (a verb meaning 'to finish') and a bridging context comparable to that described for Old Spanish are available in contemporary English, the question arises of why the same grammaticalization of the FINISH construction as a recent past marker has not (yet) occurred in English. This lack of a parallel development could be attributed to a number of language-internal factors – including the presence of a strong lexical marker of temporal immediacy such as 'just' or the presence of a gerund instead of an infinitive as in the Spanish construction – as well as to language-external factors such as (lack of) geographic mobility (Thomason & Kaufman, 1988), social norms (Mazzola et al., 2022) or demographic factors (Pijpops et al., 2015). Another potential explanation – that would severely challenge the SDH – lies in the possibility that in English, the lexical material (*finish*) in combination with the suggested bridging context does not give rise to a recent past inference at all, preventing any subsequent actualization and grammaticalization processes. If the SDH holds true, at least a trace of this inference of immediacy should also be observable in contemporary English, with other language-internal or language-external factors having prevented the grammaticalization process.

This discrepancy provides a fruitful testing ground for the experimental examination of theoretical approaches to language change. Whereas previous corpus work has provided highly valuable information on possible bridging contexts, the investigation of complex theoretical approaches such as the SDH requires insights into cognitive and causal implications, which are limitations of corpus-based research. The present study aims to go beyond a descriptive perspective by enabling a falsification of the *SDH*, thereby allowing for predictive statements about potential future language changes.

To this end, we use experimental methods to determine whether the inferential patterns that drove grammaticalization in Old Spanish are potentially available to speakers of Present-Day English. By manipulating factors such as informativity and temporal subordination structures – crucial in the Spanish grammaticalization of the

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

construction – we investigate whether temporal proximity inferences might emerge under similar conditions in English. Our central research question is as follows: *Can native English speakers derive temporal proximity inferences from FINISH constructions in the same bridging contexts that proved crucial for meaning extension in Spanish?* Therefore, the experimental design of this study includes scenarios that systematically vary the informativity of the gerund complement, and the subordination status of the clause in which it appears, allowing us to isolate the precise conditions that facilitate or inhibit the emergence of temporal proximity inferences.

Through this approach, the study contributes to broader debates about the relationship between synchronic variation and diachronic change in language. If English speakers can access temporal proximity inferences from FINISH constructions under specific conditions, despite the lack of grammaticalization, this would suggest that the potential for grammaticalization along the paths outlined by the *SDH* exists synchronically across languages, with additional language-internal or language-external factors determining whether this potential is realized diachronically. Conversely, if English speakers consistently fail to derive these inferences even in contexts analogous to the bridging contexts identified for Spanish, this suggests that the inferential patterns themselves may be language-specific rather than universal, challenging stronger versions of the *SDH*.

In both experiments in this study, we therefore tested whether the same inference of temporal immediacy can be derived by native speakers of contemporary English, if they are provided with the English *finish* + gerund construction as well as the bridging context previously identified for Old Spanish, characterized by the low informativity of the nonfinite verb following '*finish*' and a temporal subordination sentence structure. In addition, considering the previous discussion, we manipulated whether the verb 'to finish' would be presented as a Simple Past (Experiment 1) form or a Past Perfect (Experiment 2) form. In light of the SDH, which highlights the importance of not only lexical material included in the source construction but also grammatical material carrying meaning within that construction, we hypothesize that the effect would be more pronounced in Experiment 1: Since the Simple Past form might be less explicit and therefore more flexible regarding the precise temporal sequencing of the events, participants might be more sensitive to the bridging context, allowing for a reanalysis of the construction. In preparation for both experiments, informative verbs and uninformative counterparts had to be identified for each of the concrete objects involved in the '*finish* + gerund + direct object' constructions that were part of the study's experimental items.

2. Pre-study

Prior to conducting the main experiments, one uninformative verb (i.e., one verb that is predictable from context) and one informative verb for each of the concrete direct objects in the experimental items of the main study were chosen in a pre-study. Eighteen objects (e.g., *cake* or *essay*) were initially preselected together with five potentially informative and five potentially uninformative verbs for each of the objects. This selection was based on *logDice* co-occurrence scores obtained from the *enTenTen21* corpus on *Sketch Engine*, indicating the typicality or strength of the collocation (Kilgarriff et al., 2014). Five verbs with high scores and five verbs with low scores were preselected for each concrete object.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

For these 10 preselected verbs per object, additional ratings were obtained from 25 monolingual English speakers (age 18–40, *M* = 31.0, standard deviation [*SD*] = 5.79; 14 female, 11 male) who were currently living in the United Kingdom and were recruited via *Prolific*. The survey was programmed on *PsyToolkit* (Stoet, 2010, 2017). Participants rated the 10 preselected verbs for each respective item according to the questions *How typical is it to do these things with [object]* (on a scale from 1 to 7)? And *how long do the following events last (approximately)?* The typicality rating scale included an additional option to indicate that the combination of the verb and the respective object was not sensible. The rating scale for the duration question included 16 options ranging from *>0 to 5 seconds* to *>3 months*, with small differences between the lower options of the scale and large differences between the options at the upper end of the scale (see Figure 1). Items were matched for duration, as the inherent duration of the first event of each trial would already alter the estimated temporal distance between the two events of that trial. Consider, for instance, an expression such as *When Sam finished writing his PhD thesis, he went to his holiday home* as opposed to *When Sam finished locking the door, he went to his holiday home*: The longer duration of the first event of the first expression would likely elicit a longer temporal distance judgment than the first event of the second expression.

All events that were judged as encompassing long durations were therefore excluded from the experimental items to make items more comparable overall (see Figure 2).

One informative verb with a low typicality rating as well as one uninformative verb with a high typicality rating were selected for each object (e.g., *finish eating the cake* [uninformative] versus *finish stealing the cake* [informative]). Furthermore, the uninformative and informative verbs for each specific item were matched more precisely regarding their duration. After selecting two verbs for each of the 18 objects, the 16 objects with the most consistently rated verbs were selected for the creation of the final stimuli (see Appendix, Table A1, for a list of selected verbs, together with

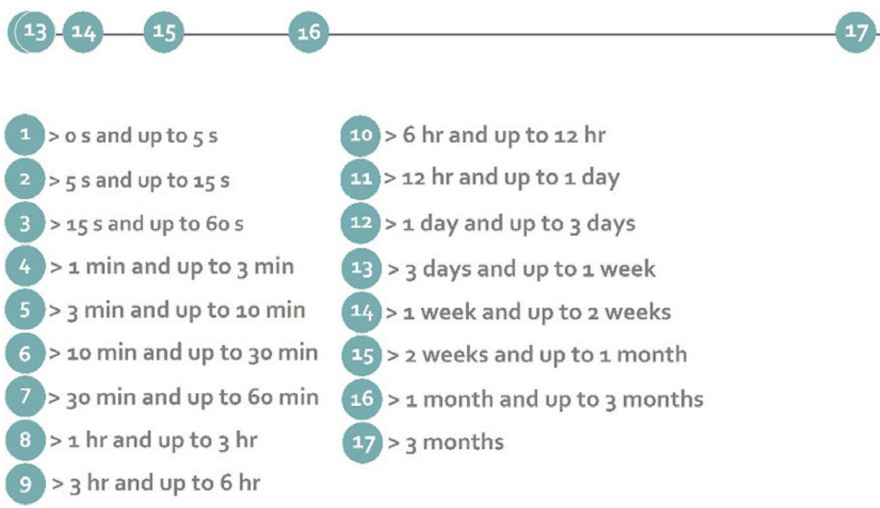


Figure 1. Visualization of the rating scale for event duration ratings in the pre-study. Temporal distances were larger at the upper end than at the lower end of the scale.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

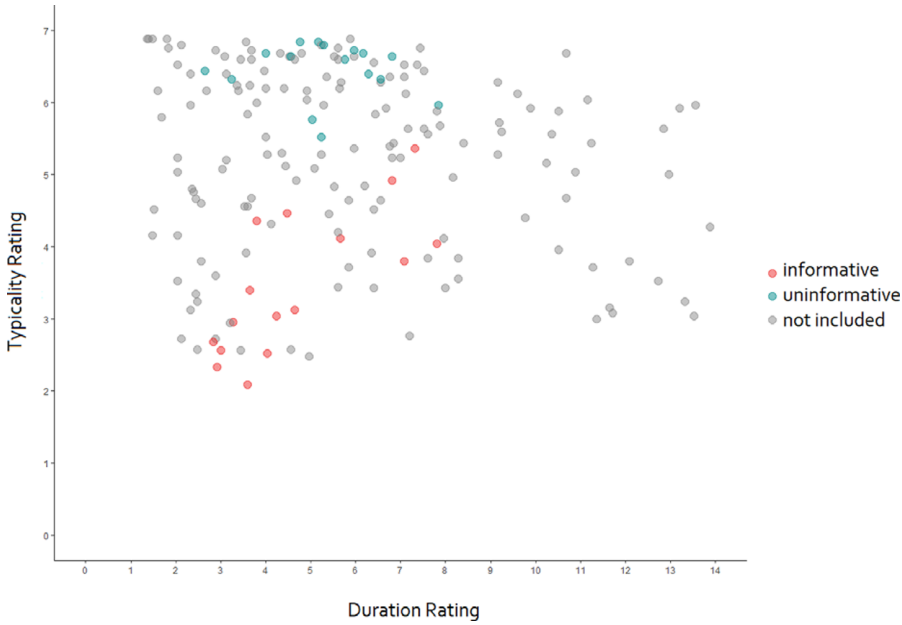


Figure 2. Typicality and duration ratings of verbs in relation to the 18 target objects of the pre-study. The informative (red) and uninformative (green) items selected for the main study (Experiment 1) were overall and individually matched for their duration, and all items at the upper end of the duration scale were not included in the main study.

their average rating scores). Informative verbs had a mean typicality rating of 3.49 (*SD* = 0.98) and a mean duration rating of 4.70 (*SD* = 1.69), whereas uninformative verbs were rated similarly regarding their duration (*M* = 5.33, *SD* = 1.33) but had a high mean typicality rating of 6.45 (*SD* = 0.39). None of the selected object–verb combinations were rated as not sensible by more than two participants.

3. Experiment 1

Experiment 1 was designed to test whether the same bridging context suggested to have supported the reanalysis of the *acabar* + *de* + infinitive construction into a recent past marker in Old Spanish would support a comparable inference of temporal immediacy for the parallel construction *finish* + gerund in Present-Day English, when the *finish* verb is presented as a Simple Past form. Our prediction is that if the *SDH* as suggested by Bybee et al. (1994) proves to be true, we should observe an effect of the bridging context – with the shortest temporal distance being inferred by participants in subordination contexts with an uninformative verb. The experiment was preregistered at https://doi.org/10.17605/OSF.IO/5HVYT.

3.1 Methods

3.1.1. Participants

One hundred and eighty monolingual English speakers (ages 18–40 years, *M* = 30.84 years, *SD* = 6.16; 103 female, 75 male, 2 diverse) were recruited online via

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

Prolific. Since no suitable analytic power analysis procedures are available for linear mixed-effects (LME) models at present, and a simulation-based power analysis as suggested by Kumle et al. (2021) could not be conducted due to the lack of comparable previous studies, it was not possible to complete an a priori power analysis to determine an appropriate sample size. Therefore, Brysbaert's (2019) general recommendations for 2 × 2 within-subjects designs were followed instead, with a sample size of 110 participants being recommended to observe a small- to medium-sized effect (d = .4) in an analysis of variance (ANOVA), including an interaction. Considering the more complex random structure in LMEs, the web-based testing, as well as a planned exclusion of participants based on preregistered control items, 70 additional participants were recruited, leading to the overall sample size of 180 English native speakers.

All participants resided in the United Kingdom at the time of their study participation. No participant had been diagnosed with dyslexia or an auditory/visual impairment that could not sufficiently be corrected by visual or hearing aids, by self-report. In addition, a prescreening setting on *Prolific* ensured that participants from the main study had not previously taken part in the stimulus selection survey. All participants were reimbursed for their participation by receiving £9 per hour via *Prolific*.

3.1.2. Materials

Each trial of the experiment consisted of two separate events. The first event was established through a *finish* + gerund construction, with the *finish* conjugated in the Simple Past (e.g., *finished eating the cake*), whereas the second event was described through a movement verb in the Simple Past with an expressed goal and implying a vague duration (e.g., *skipped to her friend's house*). The trials were further manipulated according to the four conditions of a 2 × 2 within-subjects design, with *Informativity* of the gerund (*informative* versus *uninformative*) as the first factor and *Sentence Structure* (*subordination structure* versus *main clauses*) as the second factor.

By expressing the gerund of the first event either through an informative or an uninformative verb, as specified during the pre-study, and combining the two events either through a subordination structure or by presenting two separate main clauses, each item was expressed according to one of the four conditions (Table 1).

Table 1. Example stimulus in Experiment 1, presented in all four conditions of the 2 × 2 within-subjects design with *Informativity* (*low* versus *high*) and *Sentence Structure* (*subordination* versus *main clauses*) as factors

Sentence structure	Informativity	Example
Subordination	Low	*When Mary finished eating the cake, she skipped to her friend's house.*
Subordination	High	*When Mary finished stealing the cake, she skipped to her friend's house.*
Main clauses	Low	*Mary finished eating the cake. She skipped to her friend's house.*
Main clauses	High	*Mary finished stealing the cake. She skipped to her friend's house.*

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

Items were rotated across four separate lists, so that all 16 items occurred in all four conditions without being shown more than once to any participant, while simultaneously ensuring that each participant was presented with all four conditions. Participants were assigned randomly to one of the four stimulus lists.

In addition to the 16 experimental trials per participant, each participant saw 16 filler sentences as distractors from the purpose of the study. Furthermore, these filler items included three control items implying very specific temporal distances (two short items and one long item) to ensure that participants completed the task conscientiously. The template stimulus list and the filler items are presented in the appendix (Table A2 for template stimulus list; Table A3 for filler and control items).

3.1.3. Procedure

The survey and the embedded experiment were programmed on *PsyToolkit*. Preceding the experiment, participants read the digital information sheet, completed the consent form and responded to general demographic questions. They were then introduced to the task and informed that for each trial, they would see one or two sentences at a time. Participants were instructed to indicate how much time had passed between the two events that were described in the sentences by clicking on the respective point of a blue scale (ranging from *very little time* to *a lot of time*) with the left mouse button. Furthermore, it was highlighted that there were no right or wrong answers and that the aim of the study was to investigate their first impression regarding the temporal distance, so that they would receive a maximum of 12 seconds to respond to each item. In addition, participants were presented with two example items and images of the section that had to be clicked on the scale for the respective items, indicating the shortest and longest temporal distance that they would encounter during the experiment. Finally, it was made explicit that these smallest and longest distances corresponded to events that immediately followed another event for the smallest temporal distance and to events that followed another event after 3 days for the longest temporal distance. Preceding the main experiment, participants completed four practice items.

During each trial, a fixation cross was shown above the center of the screen for 500 ms, indicating where the target sentence would occur immediately afterward. Each target sentence was shown for up to 12 seconds. Once a response was made, a cross appeared on the clicked section of the scale for 300 ms (see Figure 3). A blank screen was subsequently shown for 300 ms, followed by the fixation cross of the next trial. If no response was made for 12 seconds or more, a new trial followed a blank screen after 300 ms.

The blue scale appeared continuous to participants but was constructed out of 99 small images, resulting in a scale from 1 (shortest temporal distance) to 99 (longest temporal distance).

Once participants had reached the end of the 16th trial, they were informed that they had completed 50% of the experiment and were instructed to press the space bar once they were ready to continue. Trials were presented within four blocks that each contained one item per experimental condition as well as four filler items. Block order and item order within blocks were randomized across the experiment.

Upon completion of the main task, participants were instructed to indicate whether they had observed any patterns throughout the experiment and were asked to guess the purpose of the study. Finally, participants could indicate any potential

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

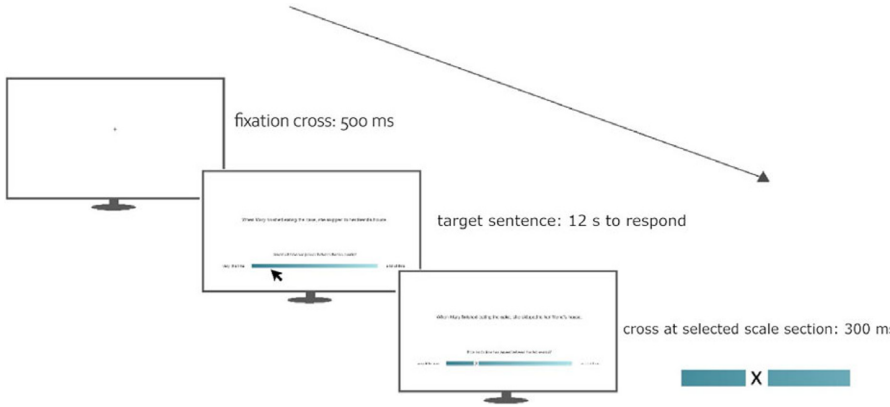


Figure 3. Sequence of events within each trial of Experiment 1.

problems they encountered during the experiment or the preceding questionnaire and were thanked for their participation.

3.2 Results

All data were analyzed in *R* (version 4.4.1). To ensure that all participants had responded to the task conscientiously, the mean ratings for the long and short control items were calculated and participants who assigned ratings of less than 2 SDs below the mean rating for the long item or more than 2 SDs above the mean ratings for each of the short control items were excluded from further analysis. Through this procedure, one participant was removed from the analysis for the first short control item (*Tina saw her friend and smiled immediately*), whereas six additional participants were excluded on the basis of the second short control item (*Arnold opened the bottle and quickly took a sip*). Finally, the inspection of the long control item – *Max received his undergraduate degree 3 days ago. Today, he begins his internship* – led to the exclusion of nine additional participants. After removing these 16 participants from the dataset, 164 participants (95 female, 67 male and 2 diverse participants) with a mean age of 30.85 years (SD = 6.03) were included in the main analysis. Mean response time of the final sample was 3769.50 ms (SD = 1705.59). With participants responding to 99.24% of the trials, less than 1% of the trials were excluded from the analysis, as participants did not respond to the stimulus sentence within 12 seconds.

Visual inspection of the graphs (see Figure 4) revealed an interaction between the factors *Informativity* and *Subordination*, with the lowest ratings observed in the *uninformative–subordination* condition (M = 19.99, SD = 17.81).

A LME model was created to test for the significance of this descriptive tendency. The model was implemented in R using the *lme4* package (Bates et al., 2015). Preceding the analysis, all categorical predictor variables were sum-to-zero coded and continuous predictors were centered. A compromise between an entirely data-driven approach and a fully hypothesis-driven approach to model selection was sought in order to balance type I and type II error risks (e.g., Barr et al., 2013; Cunnings, 2012; Matuschek et al., 2017). To this end, the random-effects structure was selected first by establishing an intercept-only model with the maximal random

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

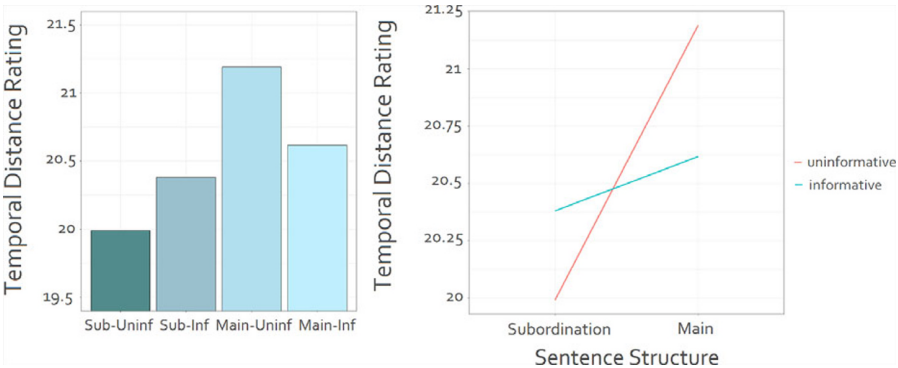


Figure 4. Temporal distance ratings for informative and uninformative verbs in the subordination conditions versus the main clause conditions. The shortest temporal distance was estimated by participants in the *subordination–uninformative* condition, in line with the suggested bridging context.

structure justified by the design that was subsequently optimized through the *step* function (*lmerTest* package; Kuznetsova et al., 2017). In a second step, a base model including the two predictors *Sentence Structure* and *Informativity* as well as their interaction was created to directly test the study's hypotheses. Additional models including *Verb Frequency* as a simple predictor or as an interaction term were established and sorted according to their Akaike information criterion (AIC) values which were obtained using the *aictab* function of the *AICcmodavg* package (Mazerolle, 2023). The model with the lowest AIC was chosen over the base model only if the additional factor significantly improved the model as indicated by the *anova()* function. Verb frequency scores were obtained from the *Wortschatz Leipzig* database (English *News 2024* corpus), where higher-frequency scores indicate that words occur less frequently. The final model was

$$\begin{aligned} &Temporal\ Distance\ Ratings \sim Informativity \times Sentence\ Structure \\ &\quad + \ Verb\ Frequency + (Informativity|Item) \\ &\quad + \ (Sentence\ Structure|Participant) \end{aligned}$$

The residual plot of this model indicated a violation of the homoscedasticity assumption, so a log-transformation was applied to the rating data. The new LME model of the log-transformed data was established through the same steps as the previous model, resulting in the following model:

$$\begin{aligned} &Log-transformed\ Temporal\ Distance\ Ratings \sim Informativity \\ &\quad \times Sentence\ Structure + VerbFrequency + (Informativity|Item) \\ &\quad + \ (Sentence\ Structure|Participant) \end{aligned}$$

P-values were provided by the *lmerTest* package, whereas random-effects estimates were obtained through the *get_variance* function (*insight* package; Lüdecke et al., 2019). The output of this model is reported in Table 2.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

Table 2. Statistical results of the linear mixed-effects analysis of Experiment 1

Fixed effects	Estimate (β)	Std. error	*t*-value	*p*-value
Intercept	2.6	.07	37.90	<.001***
Sentence structure	−.02	.01	−1.43	.154
Informativity	.03	.02	1.71	.106
Sentence structure: informativity	−.03	.01	−2.50	.013*
Verb frequency	.02	.01	2.46	.020*
Random effects				
σ^2	.72			
τ_{00}	$.70_{Participant}$			
	$.01_{Item}$			
τ_{11}	$.009_{Participant.Sentence\ Structure}$			
	$.003_{Item.Informativity}$			
ρ_{01}	$.07_{Participant}$			
	$.52_{Item}$			

Note: *p < .05, ***p < .001.

Whereas the main effects of *Sentence Structure* (β = −.02, p = .154) and *Informativity* (β = .03, p = .106) did not reach significance, a significant interaction was observed between the factors *Sentence Structure* and *Informativity* (β = −.03, p = .013), with the shortest temporal distance in the *subordination–uninformative* condition. In addition, the factor *Verb Frequency* reached significance (β = .02, p = .020), with more frequent verbs resulting in shorter temporal distance ratings[2].

To check whether the significant interaction between *Sentence Structure* and *Informativity* was driven by a varying perceived sensibility and naturalness of the experimental stimuli in the different conditions, a sensibility judgment task was administered to 25 additional monolingual English speakers currently residing in the United Kingdom (age 21 to 40 years, M = 32.04, SD = 4.21; 14 female, 11 male) via *Prolific*/*PsyToolkit.* Participants were asked to judge the experimental stimuli as well as simple sensible sentences and nonsense sentences according to the question *Do the following sentences make sense?* on a scale from *1 (not sensible at all)* to *7 (completely sensible).* As shown in Figure 5, the pattern observed in the sensibility judgment task did not correspond to the interaction reported for the main experiment. Instead, subordination sentences seemed to overall be perceived as slightly more sensible than sentences in the main clause conditions, and unsurprisingly, uninformative items were rated as more sensible than informative sentences. This tendency appeared to be additive in nature, indicating that it is unlikely that the statistically significant interaction in the main experiment was merely an effect of the varying sensibility and naturalness of the experimental items in the different conditions.

Furthermore, in line with a descriptive tendency that was preregistered, stimuli in the *subordination–uninformative* condition of the main experiment were somewhat more consistently rated than stimulus sentences in the other conditions (see Table 3), pointing toward the possibility that participants had a more pronounced intuition about the temporal inference in the *subordination–uninformative* condition.

When participants were asked to indicate the suspected purpose of the study, nobody appeared to be aware of the role of the informativity of the verb in relation to the object. Five participants indicated that the sentence structure might have played a

[2]Please note that, as indicated above, higher frequency scores in the *Wortschatz Leipzig* database indicate that words occur less frequently.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

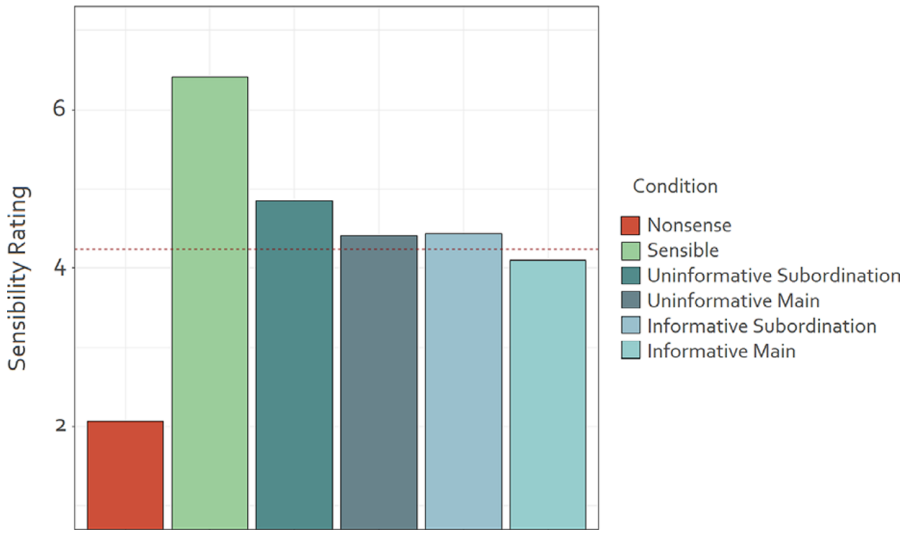


Figure 5. Follow-up sensibility judgments of experimental stimuli in Experiment 1. Nonsense (red) and simple sensible sentences (green) served as lower and upper reference points. The dashed red line indicates the center rating between nonsense and sensible sentences.

Table 3. Standard deviations for ratings within the four conditions of Experiment 1, with the smallest standard deviation observed in the subordination–uninformative condition

Condition	SD
Subordination–uninformative	17.81
Subordination–informative	18.07
Main–uninformative	18.13
Main–informative	18.79

role, without pointing out the specific contrast between main clause versus subordination constructions. Most participants simply rephrased the instructions as a purpose or pointed to the adverbs (e.g., *immediately*) in the filler sentences as well as to the lexical meaning of individual verbs (e.g., *dashed* and *skipped*).

3.3 Discussion

In line with our hypothesis, participants in Experiment 1 showed a small but significant sensitivity to an inference of temporal immediacy in the *subordination–uninformative* condition, corresponding to the bridging context that had been identified to have supported the reanalysis of Old Spanish *acabar* + *de* + infinitive into a recent past marker. The observed interaction between the sentence structure and the low informativity of the gerund suggests that the unique combination of these factors contributed to this inference. In addition, the descriptive observation that items in the *subordination–uninformative* condition were rated somewhat more consistently would similarly be in line with the assumption that the suggested

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

bridging context provided a basis for temporal inference that allowed for more systematic ratings.

Although the factor *Verb Frequency* reached significance, it did not significantly interact with the bridging context (i.e., the factors *Informativity* and *Sentence Structure*), indicating that the effect corresponded to a general cognitive bias rather than providing direct evidence for an effect of (verb) frequency on grammaticalization as such (e.g., Bybee, 2003).

4. Experiment 2

Experiment 1 of this study provided experimental evidence that speakers of contemporary English are sensitive to the recent past inference of the English *finish* + gerund construction when the main verb was presented as a Simple Past form in the suggested bridging context. To further test the precise predictions of the *SDH*, Experiment 2 was designed to examine the role of the meaning derived from formal elements of the construction by presenting participants with the verb *finish* in the Past Perfect form. Since this form tends to be less flexible regarding the event sequencing inference (i.e., it guarantees that one event follows the other one), we argue that it would partially prevent participants from being sensitive to contextual factors aiding the development of a temporal inference, including the suggested bridging context.

4.1. Methods

4.1.1. Participants

The recruitment procedure paralleled that of Experiment 1. One hundred and eighty participants (ages 18–40 years, M = 30.94 years, SD = 5.59; 97 female, 82 male, 1 diverse) completed the experiment via *Prolific* and were paid £9 per hour. All participants were monolingual speakers of English, resided in the United Kingdom at the time of their participation, had normal or corrected-to-normal vision and hearing and had not been diagnosed with dyslexia by self-report. Participants from the pre-study and from Experiment 1 did not have access to Experiment 2.

4.1.2. Materials and procedure

The four stimulus lists from Experiment 1 were adjusted, so that the first event of each trial occurred in the English Past Perfect (e.g., *When Mary had finished eating the cake, she skipped to her friend's house*) in Experiment 2. Apart from this change to the stimuli, the presentation of questionnaires and stimuli corresponded to the procedure in Experiment 1.

4.2. Results

As in Experiment 1, the data were analyzed in *R* (Version 4.4.1). Ratings from 15 participants were removed, since their responses to the long control item or one of the short control items fell 2 SDs below (long item) or above (short items) the mean response to these sentences. Therefore, data from 165 participants (age M = 30.93 years, SD = 5.48; 91 female, 74 male) were included in the main analysis.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

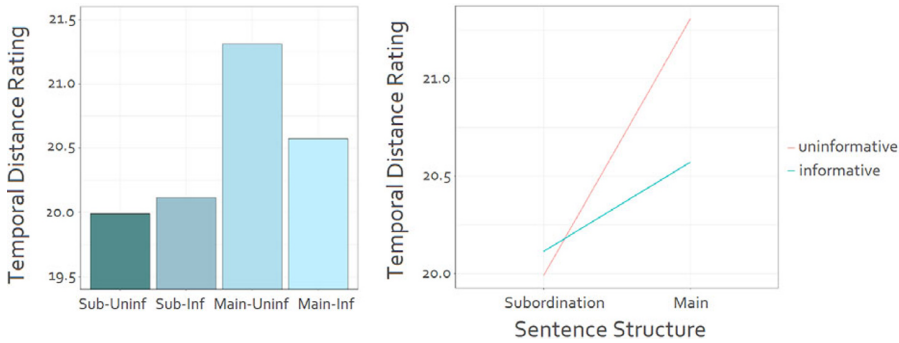


Figure 6. Temporal distance ratings for informative and uninformative verbs in the subordination conditions versus the main clause conditions in Experiment 2.

The mean response time in Experiment 2 was 3518.08 ms (*SD* = 1614.13). Participants responded to 98.98% of the trials.

Visual inspection of the graphs (see Figure 6) indicated an interaction between the factors *Informativity* and *Subordination* that was less pronounced than in Experiment 1, with the lowest ratings observed in the *uninformative–subordination* condition (*M* = 19.99, *SD* = 17.96).

Since an initial model of the raw rating data revealed heteroscedasticity as in Experiment 1, the data were log-transformed preceding the model selection procedure described for Experiment 1. The best model identified for the data in Experiment 2 was

$$Log-transformed\ Temporal\ Distance\ Ratings \sim Informativity \times Sentence\ Structure + VerbFrequency + (1|Item) + (1|Participant)$$

The output of this model is reported in Table 4. The interaction between the factors *Sentence Structure* and *Informativity* did not reach significance in Experiment 2 ($\beta = -.01$, $p = .378$). None of the main effects relevant to our hypothesis reached significance either (*Informativity* $\beta = .01$, $p = .302$; *Sentence Structure* $\beta = -.02$, $p = .111$). Only the factor *Verb Frequency* reached significance ($\beta = .01$, $p = .027$), with more frequent verbs eliciting shorter temporal distance ratings.

Table 4. Statistical results of the linear mixed-effects analysis of Experiment 2

Fixed effects	Estimate (β)	Std. error	*t*-value	*p*-value
Intercept	2.58	.07	35.94	<.001***
Sentence structure	−.02	.01	−1.59	.111
Informativity	.01	.01	1.03	.302
Sentence structure: informativity	−.01	.01	−0.88	.378
Verb frequency	.01	.01	2.22	.027*
Random effects				
σ^2	.76			
τ_{00}	$.76_{Participant}$ $.01_{Item}$			

Note: *p < .05, ***p < .001.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

Table 5. Standard deviations for ratings within the four conditions of Experiment 2, with smaller standard deviations for subordination conditions but not for uninformative conditions

Condition	SD
Subordination–uninformative	17.96
Subordination–informative	17.61
Main–uninformative	18.78
Main–informative	18.04

In contrast to the descriptive tendency that was preregistered and observed in Experiment 1, stimuli in the *subordination–uninformative* condition were not the most consistently rated stimulus sentences in Experiment 2: Overall, only the sentence structure appeared to positively affect how systematically the items were rated (see Table 5).

As in Experiment 1, none of the participants appeared to be aware of the informativity and specific sentence structure conditions. Suspected patterns and purposes corresponded to those reported for Experiment 1 (e.g., concerning the lexical meaning of verbs or the use of adverbs in filler sentences).

4.3. Discussion

In Experiment 2, we tested the prediction of the SDH that not only the lexical meaning of FINISH should determine whether an inference of recency can be obtained, but also that grammatical features of the source construction additionally affect the strength of this inference. In line with the hypothesis, no significant interaction was observed when the construction was presented in Past Perfect – despite the final sample size of 165 participants that should have allowed for the detection of a small- to medium-sized effect (d = .4). This suggests that due to the overt specification of the temporal sequencing through the Past Perfect form, participants were less sensitive to contextual information – and the bridging context specifically – when computing the temporal ordering of events.

5. General discussion

We tested predictions of Bybee et al.'s (1994) SDH, a seminal approach within research on language change, in two experiments. The SDH predicts that similar source construction meanings – including both the meaning of their lexical constituents and the meaning of their grammatical elements – should give rise to crosslinguistically parallel grammaticalization processes.

Focusing on the grammaticalization of FINISH constructions into recent past markers, we examined the role of a hypothesized bridging context for this grammaticalization pathway, in line with previous corpus research conducted by Rosemeyer and Grossman (2017, 2021), who identified a crucial role of the low informativity of the nonfinite verb and of a subordination structure in the diachronic trajectory of this construction in Old Spanish.

In our experiments in contemporary English, we therefore manipulated the informativity of the gerund in a *finish* + gerund + direct object construction (i.e., how easily the meaning of the gerund could be predicted from the direct object)

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

as well as the subordination versus main clause structure in which the construction was embedded. Accordingly, a stimulus sentence such as *When Mary finished eating the cake, she skipped to her friend's house* in the *uninformative–subordination* condition would correspond to the sentence *When Mary finished stealing the cake, she skipped to her friend's house* in the *uninformative–subordination* condition. Stimuli in the main clause conditions were expressed as two main clauses instead of the subordination structure. Participants were asked to estimate the temporal distance between the two events presented as part of each trial of our study. Results showed a significant interaction of the factors *Informativity* and *Sentence Structure* in Experiment 1. Specifically, the temporal distance was rated to be shortest when the informativity of the verb was low and the construction was presented in a subordination context. This finding is in line with the assumption that this bridging context aided the analogous grammaticalization process in Spanish. Even though the effect was small, this suggests that speakers of Present-Day English do show sensitivity to the hypothesized bridging context and the resulting temporal inference. These findings thereby replicate the pattern found in corpus data under tightly controlled conditions, indicating the cognitive plausibility of the bridging context mechanism.

Moreover, a descriptive tendency in Experiment 1 indicated a smaller variation of ratings for this condition than for all other conditions, suggesting that participants had a more pronounced and consistent temporal intuition when presented with the construction in the bridging context. Post hoc sensibility judgments confirmed that this significant interaction was not merely driven by the different naturalness of the stimuli in the different conditions. To our knowledge, this is the first study to experimentally assess the role of such complex bridging contexts in grammaticalization processes.

By contrast, no significant interaction was observed in Experiment 2, where the verb *finish* was presented as a Past Perfect form instead of a Simple Past form, despite a similar descriptive tendency of the rating data. A possible explanation for the lack of an effect in Experiment 2 was outlined in the introduction and is supported by the outcome of this experiment. Whereas the Simple Past form in Experiment 1 is relatively ambiguous regarding the precise temporal relationship between the two events, the Past Perfect form in Experiment 2 overtly specifies this relationship. Arguably, this leads participants to pay less attention to other contextual information to compute the temporal ordering of events, including the bridging context.

Overall, this confirms the assumption implied by the SDH that not only the lexical meaning of the source construction – of the verb *finish* in this case – itself determines whether the construction allows for an inference that can drive the grammaticalization process, but also that the meaning expressed by morphosyntactic and grammatical features of the source construction – such as the tense marking of the verb or the nature of the nonfinite form – additionally determines how likely a construction is to be grammaticalized.

Moreover, the findings suggest that the lack of a grammaticalization of the verb 'to finish' into a recent past marker in English is not mainly caused by the absence of suitable source material that can give rise to relevant inferences – thereby supporting the SDH. Future research should test further language-internal and language-external factors that could determine why the construction undergoes grammaticalization in some languages but not in others, despite the seemingly universal inference that it facilitates.

Another possibility that could be explored in future research is that even more pronounced effects than in Experiment 1 might be observed in languages such as

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

Italian and French where the source construction resembles the one described for Old Spanish even more closely. Note that the source construction in Old Spanish included an infinitive, whereas the analogous construction present in contemporary English involves a gerund. If a stronger version of the SDH holds, parallel reanalysis pathways should be even more likely to occur the more the meanings of the respective source constructions resemble one another. Testing whether the effect observed in this study would indeed be more pronounced in Italian and French could further shed light on the specific role of the grammatical meaning of the source construction during grammaticalization processes, compared to the role of lexical semantic information alone. In addition, comparing data from English–Spanish or English–French bilinguals with the present data could inform our understanding of the role of individuals' preexisting linguistic knowledge and associated biases in grammaticalization processes.

In light of the circumstances under which reanalysis usually takes place – where individuals would repeatedly be exposed to the constructions in linguistic bridging and congruent pragmatic contexts – it would also be relevant to test whether such repeated exposure to these constructions in relevant pragmatic contexts could strengthen the inference that facilitates the subsequent grammaticalization. Testing this hypothesis would also be relevant from a theoretical point of view, considering the causal role that both token and type frequency have been suggested to play in grammaticalization processes – by supporting processes such as semantic bleaching, phonological reduction and the spread of grammatical constructions across the lexicon (e.g., Bybee, 2003).

More generally, our study demonstrates how experimental methods can be employed not only to address questions of synchronic language variation, processing and learning, but to additionally investigate phenomena of language change, especially in combination with corpus methods. Beyond the information targeted and provided by the latter methods, experimental methods allow for a more direct falsification of hypotheses about the inferential processes that trigger and drive grammaticalization. Experimental historical linguistics can thus advance our understanding of how and under which precise circumstances languages change over time.

Data availability statement. The data, analysis codes and experiment codes that support the findings of this study are openly available in OSF at https://osf.io/mu6vg/overview?view_only=6cd508f7f2864794b561a80994c1b07a.

Acknowledgements. We thank Carmen Notarangelo and Dereck Bobran for their help with the proofreading of the experimental stimuli.

Author contribution. J.H. developed the study idea, J.H., M.F. and M.R. created and discussed the stimuli, J.H. programmed the surveys and experiments, J.H. conducted the data analysis, and J.H., M.F. and M.R. wrote the first draft of the manuscript. All authors discussed the results and read and approved the submitted version.

Funding statement. This work was supported by the European Research Council (ERC) under the European Union's Horizon Europe Research and Innovation Program (EXREAN, Grant Agreement No. 101123544). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Competing interests. The authors declare none.

Ethical statement. In compliance with the university's guidelines, the ethics checklist provided by the university was completed. The checklist confirmed that no extended ethical approval by the ethics committee

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

of the Freie Universität Berlin was required, as behavioral data from healthy adult participants were collected anonymously, and appropriate information and consent forms were provided.

References

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.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2(1), 1–38. https://doi.org/10.5334/joc.72.

Bybee, J. L. (2003). Mechanisms of change in Grammaticization: The role of frequency. In B. D. Joseph and J. Janda (Eds.), *The handbook of historical linguistics* (pp. 711–733). Blackwell Publishing. https://doi.org/10.1002/9780470756393.ch19.

Bybee, J. L., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world*. University of Chicago Press.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382. https://doi.org/10.1177/0267658312443651.

Diewald, G. (2006). Context types in grammaticalization as constructions. *Constructions*, 1(9), 1–29.

Eckardt, R. (2006). *Meaning change in grammaticalization. An inquiry into semantic reanalysis*. Oxford University Press.

Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76, 546–592.

Enfield, N. J. (2003). *Linguistic epidemiology: Semantics and grammar of language contact in mainland Southeast Asia*. Routledge Curzon.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). Academic Press.

Grossman, E., & Noveck, I. (2015). What can historical linguistics and experimental pragmatics offer each other? *Linguistic Vanguard*, 1, 145–153. https://doi.org/10.1515/lingvan-2015-1005.

Heine, B. (2002). On the role of context in grammaticalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 83–101). John Benjamins.

Hopper, P. (1991). On some principles of grammaticization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization Volume 1. Focus on theoretical and methodological issues* (pp. 17–35). John Benjamins.

Hopper, P., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge University Press.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, 1(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9.

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, 53, 2528–2543. https://doi.org/10.3758/s13428-021-01546-0.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Lüdecke, D., Waggoner, P., & Makowski, D. (2019). Insight: A unified interface to access information from model objects in R. *Journal of Open Source Software*, 4(38), 1412. https://doi.org/10.21105/joss.01412.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. https://doi.org/10.1016/j.jml.2017.01.001.

Mazerolle, M. J. (2023). *AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c) (Version 2.3.3)*. https://cran.r-project.org/package=AICcmodavg.

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

Mazzola, G., Rosemeyer, M., & Cornillie, B. (2022). Syntactic alternations and socio-stylistic constraints: The case of asyndetic complementation in the history of Spanish. *Journal of Historical Sociolinguistics*, 8(2), 197–235. https://doi.org/10.1515/jhsl-2021-0026.

Olbertz, H. (1998). *Verbal periphrases in a functional grammar of Spanish*. De Gruyter Mouton. https://doi.org/10.1515/9783110820881.

Pijpops, D., Beuls, K., & Van de Velde, F. (2015). The rise of the verbal weak inflection in Germanic: An agent-based model. *Computational Linguistics in the Netherlands Journal*, 5, 81–102.

Rosemeyer, M., & Grossman, E. (2017). The road to auxiliariness revisited: The grammaticalization of FINISH anteriors in Spanish. *Diachronica*, 34(4), 516–558. https://doi.org/10.1075/dia.16024.ros.

Rosemeyer, M., & Grossman, E. (2021). Why don't grammaticalization pathways always recur? *Corpus Linguistics and Linguistic Theory*, 17(3), 653–681. https://doi.org/10.1515/cllt-2020-0053.

Stoet, G. (2010). PsyToolkit – A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31. https://doi.org/10.1177/0098628316677643.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.

Traugott, E. C., & Dasher, R. B. (2001). *Regularity in semantic change*. Cambridge University Press.

Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes*. Oxford University Press.

Veyrat Rigat, M. (1994). La perífrasis verbal acabar de + infinitive y la resolución de su ambigüedad. *Vox Romanica*, 53, 238–252.

Yllera, A. (1980). *Sintaxis histórica del verbo español: Las perífrasis medievales*. Universidad de Zaragoza.

Appendix

Table A1. Verb–object combinations selected for Experiments 1 and 2, together with their duration and typicality ratings in the pre-study

	Uninformative			Informative		
Object	Verb	Duration	Typicality	Verb	Duration	Typicality
cake	*eat*	4.76	6.84	*steal*	3.00	2.56
letter	*write*	6.28	6.40	*rearrange*	4.04	2.52
guitar	*strum*	2.64	6.44	*smash*	2.84	2.68
sword	*sharpen*	5.24	5.52	*hide*	3.28	2.96
dress	*buy*	5.76	6.60	*dye*	7.80	4.04
carpet	*vacuum*	5.16	6.84	*discard*	5.65	4.12
pumpkin	*carve*	6.80	6.64	*halve*	3.80	4.36
computer	*get*	7.84	5.96	*hack*	7.08	3.80
loaf	*slice*	3.24	6.32	*reheat*	4.64	3.12
essay	*read*	6.16	6.68	*annotate*	7.32	5.36
soup	*cook*	6.56	6.32	*chill*	6.80	4.92
sculpture	*unveil*	5.04	5.76	*vandalize*	4.24	3.04
pasta	*boil*	5.28	6.80	*microwave*	4.48	4.46
cabbage	*chop*	4.00	6.68	*squeeze*	2.92	2.33
bicycle	*ride*	5.96	6.72	*steal*	3.64	3.40
eggs	*scramble*	4.56	6.64	*wrap*	3.60	2.09

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press

Table A2. Stimulus sentence list for Experiment 1 (finished) and Experiment 2 (had finished) in the subordination condition with one informative (e.g., stealing) and one uninformative (e.g., eating) verb per item

	Stimulus sentence
1	When Mary *finished/had finished* [**eating/stealing**] the **cake**, she skipped to her friend's house.
2	When David *finished/had finished* [**writing/rearranging**] the **letter**, he commuted to his job.
3	When Linda *finished/had finished* [**strumming/smashing**] the **guitar**, she walked to the embassy.
4	When Christopher *finished/had finished* [**sharpening/hiding**] the **sword**, he cycled to the lake.
5	When Karen *finished/had finished* [**buying/dyeing**] the **dress**, she drove to the airport.
6	When George *finished/had finished* [**vacuuming/discarding**] the **carpet**, he skated to the dentist.
7	When Elizabeth *finished/had finished* [**carving/halving**] the **pumpkin**, she hiked to the woods.
8	When Ronald *finished/had finished* [**getting/hacking**] the **computer**, he raced to the swimming pool.
9	When Jennifer *finished/had finished* [**slicing/reheating**] the **loaf**, she ran to the city centre.
10	When John *finished/had finished* [**reading/annotating**] the **essay**, he jogged to the nature reserve.
11	When Dorothy *finished/had finished* [**cooking/chilling**] the **soup**, she rambled to her colleague's flat.
12	When Richard *finished/had finished* [**unveiling/vandalizing**] the **sculpture**, he went to his aunt's cabin.
13	When Patricia *finished/had finished* [**boiling/microwaving**] the **pasta**, she dashed to the hotel.
14	When Charles *finished/had finished* [**chopping/squeezing**] the **cabbage**, he travelled to his sister's new home.
15	When Susan *finished/had finished* [**riding/stealing**] the **bicycle**, she jaunted to the beach.
16	When Anthony *finished/had finished* [**scrambling/wrapping**] the eggs, he strolled to the museum.

Table A3. Filler and control sentences for Experiment 1 and Experiment 2. Control sentences are underlined

	Filler/control sentence
1	Peter raised his hand and hit the fruit fly.
2	Evelyn spent two days preparing for this competition. Today, she will compete against her major rival.
3	Nick had locked the door. But had he switched off the lights?
4	Leo apologized over and over again but the day after he had forgotten about the whole thing.
5	Max received his undergraduate degree three days ago. Today he begins his internship.
6	Tina saw her friend and smiled immediately.
7	Kamala is buying a new pair of shoes because she lost her old ones yesterday.
8	Paula is reading her acceptance letter. Tomorrow, she will call her grandfather.
9	Once the teacher mentioned Oliver's name, he became silent.
10	Two days ago, Mandy began her student exchange. She is already returning home now.
11	Yesterday, Liz would not have believed it but now she could see it with her own eyes!
12	Had Alex changed his hairstyle from yesterday? He looked different today.
13	Arnold opened the bottle and quickly took a sip.
14	Carol entered the office at 9 am and left at 6 pm.
15	Thomas opened the wardrobe. He pulled out a pullover.
16	Emma shouted: > > Someone stole my bike! < <

Cite this article: Heine, J., Fuchs, M., & Rosemeyer, M. (2026). The test of time: experimentally recreating the reanalysis of FINISH as a recent past marker, *Language and Cognition*, 18, e2, 1–23. https://doi.org/10.1017/langcog.2025.10050

https://doi.org/10.1017/langcog.2025.10050 Published online by Cambridge University Press