Why don't grammaticalization pathways always recur?

Malte Rosemeyer and Eitan Grossman
Accepted to *Corpus Linguistics and Linguistic Theory*

December 18, 2020

Abstract

Many grammaticalization pathways recur across languages. A prominent explanation for this is that the properties of lexical items determine their developmental pathways. However, it is unclear why these pathways do not always occur. In this article, we ask why English did not undergo a cross-linguistically common grammaticalization pathway, FINISH > ANTERIOR. We operationalize this question by testing a theory proposed on results regarding a language that did undergo this change, Spanish, on corpus and experimental data. While English FINISH constructions are associated with *some* of the distributional properties of Early Spanish FINISH, speakers do not show evidence of conventionally associating FINISH constructions with a particular type of inference crucial for the grammaticalization of the Spanish anterior. We propose that the non-conventionality of this inference blocks the grammaticalization of 'finish' constructions, demonstrating that some of the black box of language change currently attributed to chance can be explored empirically.

Keywords: grammaticalization; reanalysis; tense; anterior

1 Introduction

It is well known that grammaticalization pathways recur across languages: future tense markers often develop out of constructions containing verbs of volition, definite articles often develop out of demonstratives, and anteriors often develop out of constructions containing a verb meaning 'finish' (Bybee et al. 1994). Two main types of explanation have been offered for this recurrence. One is 'event-based triggers' (Bickel 2015), historical events that brought speakers of different languages into contact and thereby led to contact-induced grammaticalization (Heine and Kuteva 2005). For example, the development of perfect constructions out of possessive constructions across Europe has been attributed to language contact (Drinka 2017), and the grammaticalization pathways involving verbs meaning 'give' recur throughout Southeast Asia (Enfield 1994). The other type of explanation involves 'functional factors,' which are principles

> grounded in the biological/cognitive or social/communicative conditions of language, such as specific processing preferences ... or specific sociolinguistic constellations ... that systematically bias the way linguistic structures evolve. The defining property of functional triggers is that they affect transition probabilities universally, independent of concrete historical events (Bickel 2017: 42).

The functional factors that have been proposed include the inherent semantics of lexical expressions and the inferences to which they give rise, whether conventionally or on the fly (Bybee et al. 1994); a tendency for grammars to be simplified by language acquisition (Roberts and Roussou 2003); the tendency of speakers to talk about their subjective point of view (Traugott and Dasher 2003) hearer-based mechanisms such as presupposition failure (Eckardt 2009), and more.

Curiously, there have been relatively few discussions about why a particular grammatical pathway does not recur in all languages, or even in one language. This question, which Weinreich et al. (1968) called the actuation question (see also de Smet (2012), Walkden (2017) and the papers in Andersen (2001)), is at the frustrated heart of historical linguistics. It is possible that most linguists assume that whether or not a particular grammaticalization pathway has occurred in a particular language can be chalked up to the fundamentally probabilistic nature of language change.

However, a very few proposals have been made about the *non*-recurrence of grammaticalization pathways in some languages. For example, Ansaldo and Lim (2004) proposes that some commonly-observed aspects of grammaticalization, e.g., coalescence and univerbation, do not occur in some Asian languages due to language-specific phonological and phonotactic properties. Burridge (2006) suggests that highly culture-specific factors inhibit some pathways, thereby leading to grammaticalization byroads, in Pennsylvania German. Finally, Eckardt (2006: 105) assumes that although the German *gehen* 'go' + infinitive construction can sometimes give rise to a futurity implicature, this inference has not been conventionalized as in English or Spanish, resulting in a lack of grammaticalization of this construction in German. However, this non-conventionalization is itself left unexplained. Other proposals for reasons that a language did or didn't undergo a particular developmental pathway are found in, e.g., Fischer (1997, 2007); Verstraete (2008); De Vogelaer (2010); De Smet and Van de Velde (2013); Gafter et al. (2019).

In this article, we engage with one functional factor, asking why it does not lead to a particular, cross-linguistically common grammaticalization pathway. The functional factor is Bybee et al. (1994)'s proposal that similar lexical meanings give rise to the same implicatures, thereby leading to the same grammaticalization pathways. We call this the Lexical Determinism Hypothesis (LDH). Several components of this hypothesis can be teased apart. The first component is the hypothesis that lexical items with similar meanings undergo similar developmental trajectories. This seems to be well-established in the grammaticalization literature. The second component, with which we engage in

this article, is the hypothesis that these similar developmental trajectories are the result of the similar implicatures that lexical items with similar meanings give rise to. The latter is not an unexpected causal theory, given that the non-detachability of inferences is considered a basic Gricean principle of communication (Grice 1975).

Our case study deals with the development of recent past or anterior constructions out of constructions containing a verb meaning 'finish', in which an active agent brings a process to an end. Bybee et al. (1994) document this grammaticalization pathway in languages as diverse as Sango (Central Africa), Mwera (Tanzania), Tok Pisin (Papua New Guinea) and Palaung (Burma). In Palaung, for example, (1), the verb cluster "signals a situation prior to and relevant to the situation in the next clause" (Bybee et al. 1994: 72).

(1) Palaung (Austroasiatic, Burma; cited in Bybee et al. (1994: 72))
mī hwǫ-i hǫm yǫ pǫm vēng
2SG finish eat EMP rice return
'Come after you have eaten [rice]'

This particular pathway is common in languages, including Tobela, Daga, Coptic, Thai, and numerous creoles, such as Tok Pisin and Rodrigues Creole French, and it is attested in sign languages such as American Sign Language.

Interestingly, it is not common in Europe, with the exception of Ibero-Romance. In Modern Spanish, for example, the *acabar* 'finish' + *de* 'of' + infinitive construction has developed into a marker of recent past (2).

(2) Spanish (Indo-European; (Veyrat Rigat 1994: 239))
Juan acab-a de le-er un libro estupendo
Juan finish-PRS.3SG of read-INF DET.INDF.M.SG book great
'Juan just read a great book'

Since the Lexical Determinism Hypothesis seems to explain much of the variance in the marking of grammatical categories in many of the world's languages, it is interesting to ask why it doesn't cover even more of the variance. For example, if the Lexical Determinism Hypothesis predicts that lexical items with similar meanings are associated with similar implicatures and thereby lead to similar developmental trajectories, then why didn't English develop a FINISH anterior? We suspect that the devil is in the details. In particular, we hypothesize that the Gricean notion of non-detachability–the idea that synonymous lexical items invoke similar implicatures–is not language-independent, and inferences associated with lexical items with comparable meanings–whether implicatures or presuppositions–may be language-specific. Moreover, we propose that the likelihood of a particular type of inference to be conventionalized and semanticized may be facilitated or inhibited by aspects of language structure and usage.

In this study, we present preliminary results from a study that aims at testing the predictions of the Lexical Determinism Hypothesis experimentally, and thus "experiments on the past" (Grossman and Noveck 2015; De Smet and

Van de Velde 2017; Hilpert and Saavedra 2018). The structure of the paper is as follows. In Section 2, we provide a historical account of a FINISH anterior in Spanish, focusing on the type of inferences that were proposed to be crucial for its grammaticalization. In Section 3, we turn to English, which did not undergo a similar developmental trajectory, and present in Section 3.1 a corpus study of constructions containing the verb 'finish.' The goal of this section is to evaluate whether the specific predictions formulated on the basis of the Spanish study were borne out in a corpus of Present-Day English. Section 3.3 describes a follow-up psycholinguistic experimental study on speakers of English. In Section 4, we discuss our findings and their broader relevance.

In general, our results support the Lexical Determinism Hypothesis to an extent, but also suggest that speakers of different languages may exploit different implicatures to different degrees. Similarly to Early Spanish speakers, Modern English speakers use FINISH + gerund sequences to highlight the narrative progression from the finished event to a subsequent event in the text. However, whereas this implicature was conventionalized in Early Spanish, leading to a hearer-based reanalysis, we find no evidence for such a process in English. In other words, the lexically similar constructions in Spanish and English have different *salient discourse profiles* (Ariel 2008), which are are 'bundlings of lexical, syntactic and pragmatic features that tend to be repeatedly interpreted in a specific way' (Gafter et al. 2019). In our view, this explains why FINISH + gerund has not been grammaticalized into an anterior or recent past marker in English.

2 The role of implicatures in grammaticalization

In this section, we briefly review some proposals regarding the role of implicatures, and inferential mechanisms more broadly, in grammaticalization. In Section 2.1, we review some proposals in prior art, and in Section 2.2, we present a worked example from Spanish.

2.1 A brief survey of inferential mechanisms in language change

Utterances usually convey more meaning than can be explained on the basis of the semantics of the words alone. As described in Ehmer and Rosemeyer (2018), this 'surplus' of meaning arises on the basis of processes of *inferencing*, namely "process of accepting a statement or proposition (called the conclusion) on the basis of the (possibly provisional) acceptance of one or more other statements or propositions (called the premises)" (Huang 2011: 397). We can distinguish between two types of inference: entailment and pragmatic inference. Entailment refers to a type of logical connection between propositions, where the truth of one proposition follows from the truth of another proposition. For instance the sentence *I have drunk a liter of milk* entails *I have drunk half a liter of milk*. In contrast, pragmatic inference is based on default logic, i.e. "reasoning on the basis of stereotypes and prototypes" (Eckardt 2006: 86). For instance, in

the correct context and with the correct intonation the sentence *I have drunk half a liter of MILK* might cause the listener to infer that she was incorrect in assuming that the speaker has drunk half a liter of beer. If this inference was intended by the speaker, it can be described as an implicature. Because pragmatic inferences, as well as implicatures, are based on probabilistic logic, they can be canceled, whereas entailments cannot.

Many early studies of grammaticalization, in particular those that paid attention to the motivations for meaning change associated with grammaticalization, proposed that meaning change results from the conventionalization of conversational implicatures (cf., e.g., Dahl (1985); Traugott and König (1991); Bybee et al. (1994); Traugott and Dasher (2003); Hopper and Traugott (2003). A well-known example is the development of future reference from allative or motion-cum-purpose constructions, such as the *going to* + *V* constructions that developed into future tenses. Crucial to such accounts is the claim that at early stages, the motion-cum-purpose constructions were used by speakers to implicate *intention*, which Bybee et al. (1994) and others argued is the crucial bridge to the development of future reference. Due to the conventionalization of this implicature, the literal meaning of these constructions was then reanalyzed in terms of future reference. Note that this reanalysis did not lead to the replacement of the original allative meaning, which is still possible in Present-Day English (e.g., *I'm going to fetch some potatoes* is a licit answer to *Where are you going?*).

Crucially to our approach, more recent research has highlighted the relevance of the perspective of the hearer for the description of grammaticalization (see Detges and Waltereit (2002), Eckardt (2009), Grossman and Polis (2014), and Rosemeyer and Grossman (2017). These studies argue that the notion of conventionalization is only a partial explanation of the processes of meaning change undergone in grammaticalization. Rather, models of grammaticalization have to take into account hearer-based reanalysis, i.e., a mechanism by which speakers come to associate a previously inferred meaning as the principal meaning of a construction. Hearer-based reanalysis has been modeled in terms of the 'Principle of Reference' (Detges and Waltereit 2002) or 'Avoid Pragmatic Overload' (Eckardt 2009), among others. Such approaches assume that hearer-based reanalysis occurs when a construction is used innovatively in a context that is no longer compatible with the original meaning of that construction. Consequently, a hearer may be coerced into reanalyzing the construction in terms a meaning that can be inferred from the specific use of this construction in this specific context. An obvious example for such a context in terms of the *going to* + *V* construction is the sentence *I'm going to go to Paris*, which excludes a literal interpretation of the first instance of 'go'. In such approaches, it is by repeated processes of hearer-based reanalysis that an implicature can be conventionalized across time.

Having surveyed some approaches to the role of inferences and implicatures in grammaticalization processes,[1] we are now in a position to turn to our case

[1] A fuller survey would also take into account formal approaches, such as those in Bar-

study of 'finish' anteriors in Spanish.

2.2 Case study: Spanish

As noted above, Spanish developed a recent past construction based on the lexical verb *acabar* 'finish', illustrated in 2. A recent longitudinal study (Rosemeyer and Grossman 2017) analyzed the discursive conditions for the emergence of the implicature that enabled the reanalysis of the aspectual 'finish' meaning as a recent past anterior. In this section, we briefly report on the findings and conclusions of this study, focusing on the proposed inferential mechanisms and contexts that led to the eventual grammaticalization of the construction.

In Early Spanish, *acabar* (+ V) constructions showed an interesting but not unexpected distribution: when the implied activity that was finished was uninformative, the event was left unexpressed by a lexical verb (3)[2]. When the activity was informative, the lexical verb was overtly expressed as an infinitive (4). Informativity is understood here as the probability for the finished event to be inferred from the direct object; whereas the event 'building' is stereotypically associated with meaning 'finish (a/the) tower', the event 'surrounding' is not likely to be inferred from the meaning 'finish (a/the) city'.[3] In other words, the distribution followed the principle "If you don't have something interesting to say, don't say anything at all" (Rohde 2020). This distribution was analyzed as two constructions, the infinitiveless construction coding the uninformativity of the finished activity, and the construction with the overt infinitive coding its informativity. However, at this stage, neither construction was associated with a recent past or anterior meaning.

(3) Spanish

e	**acab-ó**	la	torre	dal=faro	que
and	finish-PST.PFV.3SG	DET.DEF.F.SG	tower	of.the=lighthouse	that
comença-a-ra		hercules			
begin-THEME-PST.IPFV.SBJ.3SG		hercules			

'And he finished the tower of the lighthouse that Hercules had begun' (*Estoria de Espanna*, 13th c., GRADIA)

Asher Siegal (2020); Hale (1998); Kiparsky (2012); van Gelderen (2004), just to name a few.

[2] All Spanish examples are taken from the GRADIA corpus of historical Spanish (Grupo Gramática y Diacronía 2015)

[3] There is an ongoing debate about the mechanism causing the inference that leads to the default interpretation 'finish building the tower' for the syntagm *finish the tower* (Lauwers and Willems 2011). In particular, while ellipsis accounts such as Kleiber (1999) assume that in such sequences the abstract predicate 'building' is inferred in the final semantic representation, coercion accounts such as Pustejovsky (1995) and Michaelis (2004) propose that the direct object is type-shifted from an artifact ('tower') to a complex event ('building the tower'). The difference between the two accounts largely has to do with the question of the ontological category of the direct object. Given that the two accounts agree that an inference by the hearer has to take place in order to reach an interpretation in which the finished event is indeed represented, we believe that that our account is compatible with any of the different approaches.

(4) Spanish
aquel anno **acab-o** **de cercar** toda la uilla de
that year finish-PST.PRFV.3SGM of surround.INF all the city of
Roma de muros muy fuertes
Rome of walls very strong
'That year he finished surrounding the city of Rome with very strong walls'
(*Estoria de Espanna*, 13th c., GRADIA)

It turned out that the recent past meaning first occurred in past-of-past contexts in which, contrary to the distribution described above, an overt infinitive expressed an uninformative event. For instance, in example (5), the infinitive *hacer* 'make' could have easily been omitted without loss of information, given that the meaning 'to finish a bridge' implicates that the full meaning of the sentence is 'to finish building/making a bridge.' It was proposed that in contexts such as (5), the apparent 'too much information' led to the highlighting of the event itself, and consequently the highlighting of the temporal sequence between the two events ('making the bridge' and 'passing over the bridge'). Note that in (5), the building of the bridge is a crucial precondition for the following event of crossing the bridge.

(5) Spanish
Y en este tiempo como **fuese**
and in DET.DEF.M.SG time when be.PST.IPFV.SBJ.3SG
acab-a-da **de hacer** la puente
finish-THEME-PTCP.F of make.INF DET.DEF.F.SG bridge
pas-ó la Infantería española
pass-PST.PFV.3.SG DEF.DET.F.SG infantry Spanish
'And then, when they had finished building the bridge
(lit. the bridge had been finished building), the Spanish
infantry passed over it' (*Crónica del Emperador Carlos V*, 16th c., GRADIA)

Rosemeyer and Grossman (2017) proposed that anterior temporal clauses were the crucial context for the reanalysis of the *acabar* construction as an anterior construction. The argument is, briefly, as follows. Innovative speakers used the *acabar de* INFINITIVE construction in contexts not licensed by its earlier meaning, in order to highlight a discourse progression. Listeners, unable to accommodate the presupposition that the finished activity is highly informative, were coerced into reanalyzing the construction (Eckardt 2009; Schwenter and Waltereit 2010). Because the *acabar de* INFINITIVE construction was frequently used in relative past contexts, listeners reanalyzed the construction as marking relative past, which in main clauses amounted simply to locating the event prior to speech time.

More broadly, the authors argued that this particular grammaticalization pathway was shaped by the lexical semantics of 'finish', by syntactic context, and by general pragmatic mechanisms, such as implicature and presupposition accommodation, as well as an overarching principle of relevance.

3 'Finish' constructions in Present-Day English

Historical linguists often offer functional and/or event-based explanations for observed changes. However, it is difficult to know if these are simply stories that we are telling ourselves. Since we do not have access to the minds of speakers or to interaction between speakers and listeners in the past, it is hard to evaluate the mentalistic claims made by historical linguists, beyond the consistency of the claims with the data.

In order to address this problem, we take the analysis of Rosemeyer and Grossman (2017) as a causal theory which can be tested on other languages. Since their claims are not about historical forces, but rather about panchronic principles of online communication, we test the validity of the particular predictions on speakers of a distantly-related language, English. We tackle this in two ways. First is a corpus study (3.1), in which we operationalize the notion of 'informativity,' used in an impressionistic fashion in Rosemeyer and Grossman (2017), as a quantitative measure known as Pointwise Mutual Information. Second is an experimental study (3.3), in which we measure the extent to which the inferential mechanisms claimed to be operative in Early Spanish are operative in Present-Day English as well.

3.1 Goals and predictions

In this section, we present a corpus study of Present-Day English. The goal of the study was to test the prediction that, like in Early Spanish, the use of an overt lexical verb in Present-Day English FINISH constructions depends on two factors. First, we test the prediction that the distribution of overt vs. omitted lexical verbs in Present-Day English is subject to the same constraint as that identified for Early Spanish. Specifically, we test whether the use of an overt verb (in a form we call here by the traditional term 'gerund') is more likely when the action described by the lexical verb is informative.

By 'informative' we mean that the interpretation differs from the default interpretation of a specific 'finish' + direct object collocation. For example, we hypothesize that, averaging across contexts, the collocation 'finish dinner' (6a) has the default interpretation 'finish eating dinner' and not, e.g., 'finish transporting dinner in a car' or 'finish instagramming dinner.' Conversely, we expect that when the finished event is informative, an overt verb will occur, as in (6b). This parameter of course only affects transitive uses of FINISH (see Section 3.2).

(6) English
   a. John finished dinner.
   b. John finished instagramming dinner.

Second, the Lexical Determinism Hypothesis would lead us to expect that like Early Spanish speakers, Present-Day English speakers can exploit the FINISH + gerund construction to implicate the highlighting of a narrative progression in

a text. We would therefore expect the use of FINISH + gerund to be likelier in temporal subordination contexts than in other syntactic contexts such as main or relative clauses.

Finally, since we hypothesize the 'highlighting' function to be more likely to arise with *uninformative* than informative gerunds, we would expect the effect of temporal subordination to be stronger with uninformative than informative gerunds. Thus, we predict that the syntactic context (temporal subordination vs. the rest) will interact with the lexical semantic context (uninformative vs. informative finished activity).

3.2 Data extraction and annotation

We extracted all instances of the syntagm *had finished* from two corpora of Present-Day English: the Corpus of Contemporary American English (COCA) (Davies 2017b) and the British National Corpus (BNC) (Davies 2017a). The decision to focus on pluperfect uses of FINISH was made on the basis of the fact that these data would allow for a comparison of FINISH in temporal subordination contexts and other syntactic contexts.

After elimination of false positives and duplicates, this procedure yielded a total of n=2326 tokens of *had finished*. Our hypotheses target the use of FINISH in contexts in which a direct object is present, as no default interpretation can be derived from uses of FINISH without a direct object. For instance, in example (7) the action that was finished ('playing the song') has to be inferred from context. We consequently eliminated all tokens of FINISH in which no direct object was present.

(7) When I turned the radio station back up, Blind Willie **had finished** and Big Joe Williams was carrying on about his "Little Leg Woman". (Bk:DesertWindLena, 2012, COCA)

Likewise, when the direct object refers to an event or a state, sometimes only a default interpretation is available and the use of an informative gerund is impossible. For instance, in example (8), the only possible gerund that could be used is the semantically light *doing*, which makes the default interpretation overt. This means that such uses escape the variable context (Tagliamonte 2012: 10-11).

(8) Other than Winston and Buddy Lockridge, he'd only given the number to Brass Doran in Quantico and he **had finished his business** with her. (Bk:DarknessMoreThan, 2001, COCA)

We therefore eliminated all such cases from our data, leading to a final dataset of n=660 tokens of transitive FINISH. Of these tokens, n=249 involve the use of a gerund, whereas n=411 do not.

We then coded these tokens according to four parameters targeting the hypotheses established at the beginning of Section 3.1. First, we coded the

tokens instantiating the FINISH + gerund + DO construction in terms of the variable OVERT. This variable received the three levels (a) "NoGerund", (b) "InformativeGerund", and (c) "UninformativeGerund". The distinction between informative and uninformative gerunds (levels (b) and (c)) was made on an intuitive basis and had an entirely heuristic value for our analysis.

Second, in order to establish a data-driven and more objective estimate of the informational value of the use of the gerund in a specific context involving FINISH and a direct object, we established the variable PMIFINISHOBJ. This variable measures the Pointwise Mutual Information Score (PMI) between FINISH and the lemma of the direct object.[4] It provides a quantitative measure of the attraction between these two constituents in the corpus. This is done by contrasting their actual co-occurrence rate with the co-occurrence rate corresponding to a chance distribution (Oakes 1998; Manning and Schütze 1999: 63–65,66–68). Pointwise Mutual Information was calculated on the basis of the formula in (1), where *p(do)* represents the probability of occurrence of the lemma of the direct object, *p(f)* represents the probability of occurrence of the lemma FINISH, and *p(fdo)* represents the probability of co-occurrence of f and do, within a four-word span. The probabilities p were calculated on the basis of maximum likelihood estimates; for instance, *p(f)* was calculated by dividing the number of occurrences of f by corpus size (as implemented in, e.g., (Blaette 2020)).

$$PMI(f; do) = log_2(\frac{p(fdo)}{p(f) * p(do)}) \tag{1}$$

In terms of research design, the PMIFINISHOBJ variable can be said to measure the availability of default readings for a given FINISH + DO sequence. Thus, a high value of the FINISH + DO sequence on the PMI variable would imply that a default reading is easy to derive from the FINISH + DO sequence. This expectation is borne out in the data. As can be seen in Figure 1 below, there is indeed a significant correlation between the OVERT and PMIFINISHOBJ variables. Thus, the mean Pointwise Mutual Information score between FINISH and the object lemma is lowest for cases involving informative gerunds, somewhat higher for cases involving uninformative gerunds, and highest for cases that do not involve a gerund. These differences reach statistical significance.[5] These findings clearly confirm our premise: informativity of the gerund can be measured in terms of the Pointwise Mutual Information score for the FINISH + object sequence.

[4] While our goal here is clearly related to those investigated in the framework of collostructional analysis and collexeme analysis (Stefanowitsch and Gries 2003, 20 May. 2005; Hilpert 2014), and there are other measures of attraction between parts of collocations (e.g., ΔP, Lexical Gravity G, and others, (Schneider 2018)), PMI seemed appropriate for our goals.

[5] Since the data violated the assumption of normality, U-tests were computed (Gries 2009: 209-210). The U-tests showed that the difference in mean Pointwise Mutual Information between cases without a gerund and cases involving an informative gerund is significant (W=1531, $p_{\text{two-tailed}}$<.001), whereas the difference between cases involving an uninformative gerund and cases involving an informative gerund reached marginal statistical significance (W=250.5, $p_{\text{two-tailed}}$<.1).

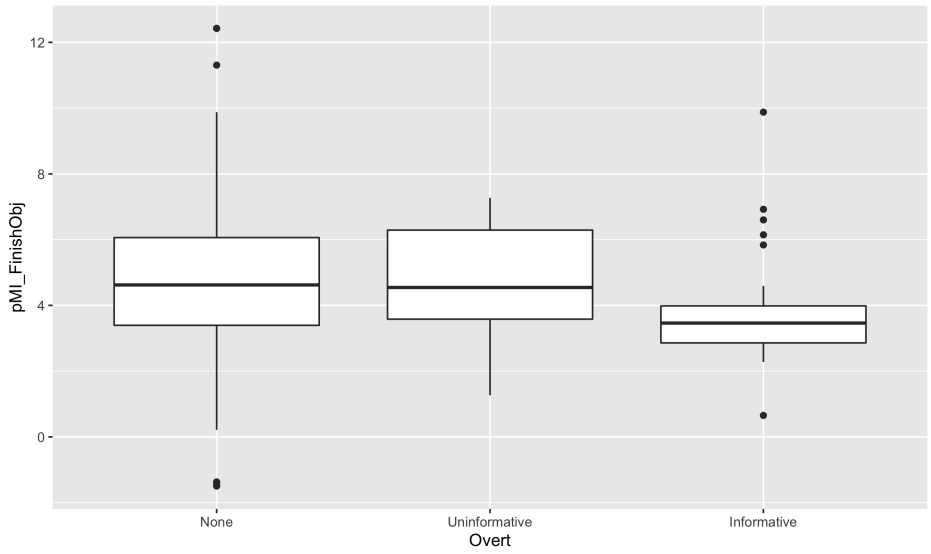

Figure 1: Mean Pointwise Mutual Information score between FINISH and the object, by OVERT

Third, we coded each token in the corpus for the variable AFTER, which describes the syntactic context. We distinguished two levels: (a) 'False', representing main clauses and subordinate clauses other than temporal clauses that indicate that the action in the main clause happened after the action in the subordinate clause (as in example 9), (b) 'True', representing temporal clauses that indicate that the action in the main clause happened after the action in the subordinate clause (as in example 10). The most frequent types of 'After' complementizers were *when*, *after*, *by the time that*, and *once*.

(9) In fact, at one point, she thought she **had finished** the model's head, but revised it when she saw him again. (BkSF:MovingMars, 1993, COCA)

(10) Henry sat down out of sight and weeded geraniums for a few moments. [...] When he **had finished the flower bed**, he thought he heard the young doctor moving in the office. (ChildLife, 1998, COCA)

Finally, we coded each token according to the nature of determination of the object because we assumed determination to be indicative of the informational status of the event. The variable DETERMINATION received the levels 'None,' 'Definite', 'Indefinite', and 'Possessive'. In examples such as (10), the omission of the gerund might be due to anaphoricity, i.e., the fact that the action has been explicitly mentioned in the preceding context and consequently, *flowerbed* has a specific referent. Likewise, we assumed that possessive marking leads to a

familiarity effect that can be interpreted in terms of our notion of informativity. Consider, for instance, example (11), in which Darcy and Booth are having brunch.

(11) On Saturday, Darcy met Box 1143, Albert Booth, for brunch at the Victoria Café. [...] By the time she **had finished her coffee**, she also doubted just about everything else he'd claimed except computer expert. (Bk:UnderBeetles, 1991, COCA)

The context of having brunch itself strongly suggests the possibility of them drinking coffee, which is why using the gerund would be even less informative in this example than in other contexts. In fact, possessive markers are commonly considered to be presupposition triggers, such that the expression 'her coffee' invites the addressee to accommodate the presupposition that there was coffee; furthermore, it invites the addressee to accommodate the presupposition that Darcy had coffee, and there is therefore no need to assert such a proposition. Such presupposition triggers work well when the presupposed proposition is uncontroversial, but they can lead to presupposition failure when the proposition cannot be reasonably assumed to be in the common ground.[6]

Finally, we can assume a similar effect for cases with bare objects. In examples such as (12), the lack of determination itself suggests a strong mutual informativity between FINISH and *lunch*, which is why using the gerund seems unnecessary here. These considerations lead us to expect an interaction between the degree of informativity of the event and the determiner of the direct object.

(12) When they **had finished lunch**, the Rex asked her grandmother if he could show Julia what lay inside (FantasySciFi, 2002, COCA)

3.2.1 Results

We conducted a mixed-effects logistic regression model that predicted whether or not a gerund was used in the FINISH construction from the variables described in Section 3.2, namely PMIFINISHOBJ, AFTER, and DETERMINATION.[7] The hypotheses established at the beginning of Section 3.1 would lead us to expect the following three effects:

- Gerund use is less likely if the FINISH + object syntagm scores high on the PMIFINISHOBJ variable, and when the object NP is definite,
- Gerund use is more likely in subordinate temporal clauses indicating anteriority of the event expressed in the subordinate clause to the event expressed in the main clause (variable AFTER),

[6] For the operationalization of presupposition accommodation and failure in historical linguistics, see Eckardt (2009) and Schwenter and Waltereit (2010)

[7] All data handling was conducted in R (R Development Core Team 2019), using the following packages: effects (Fox and Weisberg 2019), ggplot2 (Wickham 2016), lme4 (Bates et al. 2015), stargazer (Hlavac 2018)

- Gerund use is even more likely in such subordinate temporal clauses indicating anteriority if the FINISH + object syntagm scores high on the PMIFINISHOBJ variable. This interaction might be moderated even further by the determiner of the direct object (interaction between PMIFINISHOBJ, AFTER and DETERMINATION)

As random intercepts we included the variables GENRE and SOURCE, taking the labels from the corpora. These random intercepts were modeled as a nested random effect (GENRE:SOURCE) because of the hierarchical dependency relation between the two variables. This modeling strategy significantly increases the explained degree of variation (Gries 2015).

Whereas AFTER and DETERMINATION are categorical variables, PMIFINISHOBJ is a numeric variable. In line with our predictions, we also included an interaction effect between PMIFINISHOBJ, AFTER and DETERMINATION.

With a c index of concordance of 0.83, the model reached a good fit to the data. It correctly predicted 76 percent of gerund use in the data. The model was also tested for multicollinearity by using Variance Inflation Factors (VIF scores), which did not show any sign of possible correlation of fixed effects. Table 4 in the appendix summarizes the full results from the regression analysis.

The model found all predictor variables to have a significant main effect on the likelihood of gerund usage in FINISH + DO constructions. First, a higher PMI score between FINISH and the object lemma significantly decreases the likelihood of gerund use. Second, gerund use is significantly more likely in 'after' contexts than other syntactic contexts. Third, in comparison to definite determination, there is a lower likelihood of gerund use when the object is modified by a possessive or when the object is a bare noun.

Our analysis also found a significant effect of the interaction between PMIFINISHOBJ, AFTER and DETERMINATION. Figure 2 visualizes the trend of the interaction effect. It demonstrates that, as expected, in subordinate clauses introduced with a complementizer of the 'after' type, gerund use is relatively more likely with high PMI scores between FINISH and the direct object. However, this effect is restricted to tokens in which the object is not modified by a determiner ('Determination: none') or by a possessive determiner ('Determination: poss'). Note that the plot also illustrates nicely the main effect for PMIFINISHOBJ, in that in general, the likelihood of gerund use declines with an increasing Pointwise Mutual Information score.

3.2.2 Discussion

The fact that our corpus study found a higher PMI score between FINISH and the object lemma to decrease the likelihood of gerund use can be interpreted in terms of our first prediction. Thus, the use of a gerund is more likely when the interpretation obtained by leaving out the gerund differs from the interpretation obtained by using the FINISH + gerund construction. In other words, the gerund is typically used when it is informative, i.e., when the use of the gerund leads to an interpretation of the finished event that differs significantly from the default

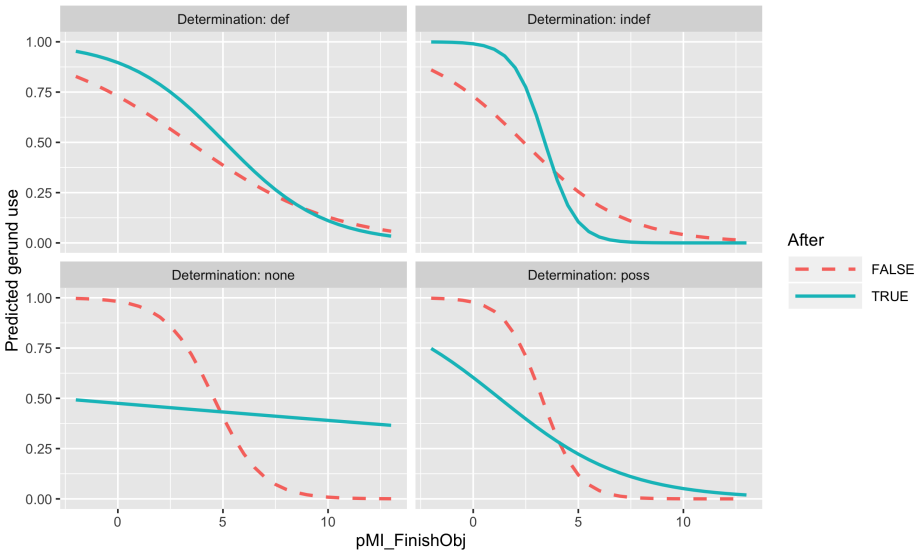

Figure 2: Predicted probabilities of gerund use by Pointwise Mutual Information score between FINISH and object lemma, temporal clause usage, and determination of the direct object

interpretation that would be expected on the basis of the combination of FINISH and the direct object.

Likewise, the results from the corpus study confirm the second prediction, namely that speakers exploit the use of the gerund in order to highlight discourse progression. Gerund use is significantly more likely in subordinate temporal clauses in which the complementizer is *after* or a complementizer with similar meaning.

The analysis found our third and most important prediction, i.e., that while gerund use is less likely in sentences where the nature of the event cannot be deduced from the FINISH + direct object sequence, uninformativity actually favors the usage of gerunds in temporal subordinate clauses indicating anteriority, to only hold in very specific contexts. Speakers do resort more frequently to uninformative than informative gerunds when exploiting the use of gerunds in 'after' contexts in order to highlight discourse progression. However, this effect is restricted to tokens in which the object is either not modified by a determiner or modified by a possessive determiner, i.e., contexts in which the determination of the direct object already indicates the low informativity of the event.

It thus appears that by using an overt gerund, the speaker or writer's highlighting discourse progression implicature can only arise in Present-Day English when there is additional contextual support for the hearer's assumption that the use of the gerund is not due to informativity. This suggests that in contrast to Old Spanish, there is no evidence for the conventionalization of this

implicature in English. This might explain why no hearer-based reanalysis of FINISH constructions as recent past markers has arisen.

It is important to highlight the potential importance of this finding for our study, since it is the first point in which our predictions, derived from a combination of linguistic theory and empirical findings from a historical case study, were not borne out for English. In essence, this finding means that while some of the inferences typically associated with 'finish' events are shared by English and Spanish, the two languages differ in a particular and potentially crucial aspect of usage. While for Early Spanish speakers, there was a tight statistical association between anterior temporal clauses and uninformative events, English speakers exploit constructions like 'After finishing eating dinner' both to highlight a discourse progression and to mark the informativity of the action described by the gerund. In other words, there is no close one-to-one correspondence between this construction and uninformative events. We return to this finding in Section 3.3.

3.3 Experimental study

The results from the corpus study presented in the last section suggest that although English speakers exploit gerund use in FINISH constructions to implicate discourse progression, this implicature is restricted to certain syntactic contexts — especially tokens in which the direct object is modified by a possessive determiner — and consequently has not been generalized. We tentatively interpret this result as raising the possibility that the use of both informative and uninformative gerunds in anterior temporal clauses blocks the one-to-one association between anterior temporal clauses and uninformative events that was hypothesized to be a main trigger of the grammaticalization of FINISH constructions in Early Spanish.

However, the data in our corpus study can only be used to look at language production, i.e., to evaluate the use of FINISH constructions by language speakers. In order to determine with more certainty whether or not a hearer-based reanalysis has occurred and the discourse progression use of FINISH + gerund sequences has been conventionalized, it is necessary to analyze the perception of such constructions.

In order to investigate whether and/or to what extent English hearers reanalyze the use of gerund in anteriority-indicating contexts in terms of discourse progression, we conducted a web-based experiment in the non-cumulative self-paced reading moving-window paradigm.

3.3.1 Materials

54 participants were asked to read 18 invented target stories consisting of an introductory sentence and a target sentence. Consider Table 1 for an example of the makeup of the target stories. The introductory sentence established the protagonist of the story (here: Elizabeth) and introduces an activity carried out by the protagonist (getting ready for the Medieval fair). The target sentence is

Table 1: An example stimulus from the reading-time experiment

Introduction	Elizabeth was getting ready for the Medieval fair.
Target	[immediacy = F, overt = none] A while after she had finished the sword, she heard her phone ring.
	[immediacy = T, overt = none] As soon as she had finished the sword, she heard her phone ring.
	[immediacy = F, overt = uninf] A while after she had finished forging the sword, she heard her phone ring.
	[immediacy = T, overt = uninf] As soon as she had finished forging the sword, she heard her phone ring.
	[immediacy = F, overt = inf] A while after she had finished choosing the sword, she heard her phone ring.
	[immediacy = T, overt = inf] As soon as she had finished choosing the sword, she heard her phone ring.

composed of an anteriority-expressing temporal subordinate clause and a main clause. In the subordinate clause, the protagonist finishes an action that is understood as part of the more general activity introduced in the introductory sentence. This action involves an artifact codified as the direct object of the action, whose existence can be inferred from the scenario depicted in the introductory sentence via a bridging inference. The main clause describes a subsequent action that takes place after the action in the subordinate clause.

As shown in the example stimulus, the subordinate clause in the target sentence was modified according to two variables: (a) IMMEDIACY, with the levels 'True' and 'False', and (b) GERUND, with the three levels 'None', 'Uninformative' and 'Informative'.[8] The variable IMMEDIACY referred to whether the complementizer expresses that the action in the main clause instantly succeeded the action in the subordinate clause (complementizer AS SOON AS, variable level 'True') or there was a lapse between the two actions (complementizer A WHILE AFTER, variable level 'False'). The three levels of the variable GERUND referred to the parameter of presence and informativity of the gerund; thus either no gerund was present (variable level 'None'), an uninformative gerund was present (variable level 'Uninformative') or an informative gerund was present ('Informative'). The two variables lead to a total of six conditions.

The dependent variable in the experimental setup was the reading time of the direct object (in the example stimulus, *sword*). This was due to the fact

[8]The distinction between the levels 'Uninformative' and 'Informative' was based on our own classification.

that the informativity of the gerund (if expressed) can only be evaluated by the readers post-hoc, i.e., upon reading the direct object. In contrast, measuring the reading times of the gerunds (when present) would not lead to an accurate representation of the impact of the parameter of informativity on the perception of the target sentences.

Reading time was understood as a proxy of the expectedness of the direct object in the specific context of the target sentence. It is well known in psycholinguistic studies that content that is plausible in terms of the current situation model by the listener eases processing in general (Hagoort et al. 2004), leading to lower reading times (Troyer and Kutas 2018). Consequently, we predicted the reading times of the direct object to be lower for target sentences in which no gerund was present than for target sentences in which an informative gerund was present because the presence of the definite determiner would lead the readers to anticipate a direct object related to the general scenario established in the introductory sentence. In other words, the referent 'sword' is already pre-activated in the reader's mind when reading the introductory sentence.

Recall that we operationalized the notion of informativity in relation to the degree of collocational association and expectedness between the direct object and the verb form. Consequently, we expected reading times of the direct object to be lower for target sentences in which an uninformative gerund was present than for target sentences in which an informative gerund was present due to the high degree of mutual informativity between uninformative gerunds and their preferred direct objects. Thus, the referent 'sword' is more likely to be expected after mention of the gerund *forging* than after *choosing*.

Finally, in light of the results of the corpus study, we expected this informativity effect to be moderated by the variable IMMEDIACY: the complementizer *as soon as* is more likely to trigger a discourse progression reading than the complementizer *a while after*. Consequently, if a hearer-based reanalysis of FINISH constructions as a marker of discourse progression has taken place we would expect reading times on the object to be even lower for target sentences with an uninformative gerund and the complementizer *as soon as*.

3.3.2 Experimental design

In conducting the experiment, a Latin square design was employed. Thus, all 54 participants were exposed to each one of the six conditions three times, but without reading the same target sentence more than once. In addition, the participants were exposed to 18 filler stories of approximately the same length. Participants were informed that the experiment aimed at testing the comprehension of these stories. Consequently, after each story, participants were asked a simple comprehension question about the contents of the story, whose real purpose was to maintain the participants' attention.

The experiment was programmed on the online experiment platform Ibex (Drummond 2020)[9] as a non-cumulative self-paced reading moving-window

[9]https://spellout.net/ibexfarm/, last access 24 June 2020.

experiment. Thus, the participants read each story word by word, seeing only one word at a time and progressing by pressing the space bar. In this manner, the precise reading times for each word in the story could be monitored.

3.3.3 Participants

The experiment was distributed on the recruitment platform Prolific[10]. Participants were paid £9.94 per hour and on average took about 15 minutes to complete the experiment.

3.3.4 Results

The n=972 sentences read by the 54 participants in the experiment were filtered according to standard experimental practices. In a first step, we defined outlier cases as cases falling below Q1-1.5*IQR, where Q1 refers to the first 25 percent of the distribution and IQR to the interquartile range, or above Q3+1.5*IQR, where Q3 refers to the upper 75 percent of the distribution. We subsequently deleted these outliers from the data, leading to a new total of n=866 responses. In a second step, we checked the correct answer responses by the participants. Overall, participants replied correctly in 90.2 percent of the cases. With 67.6 percent, one participant scored significantly lower than the average participant, and the reading times of this participant were excluded. This second exclusion process led to a final dataset of n=848 responses.

This final dataset was submitted to a statistical analysis using linear mixed-effect regression modeling. As already mentioned in Section 3.3.1, the dependent variable was the reading time (in milliseconds) of the direct object, excluding the determiner. Table 2 summarizes the predictor variables in the regression model. Note that by including random intercepts for TextID and ID, it was possible to filter out systematic variation due to these two parameters. Finally, in order to test the prediction of an influence of the parameter of Immediacy on the effect of Gerund, we tested for an interaction effect between these two variables.

The regression analysis was carried out in R (R Core Development Team 2020), using the lme4 package (Bates et al. 2016). Table 2 summarizes the results from the regression analysis. The regression analysis reached a corrected R^2 of .51, corresponding to a correlation of .71 between expected and observed values. Table 4 in the appendix gives the full results from the regression model.

The analysis found significant and expected effects for our control variables Age, WordFreq, and Trial (for instance, higher word frequency leads to reduced reading times), demonstrating that the model is generally able to explain reading times on the direct objects. The analysis also found a significant effect for the main effect of the variable Gerund, whereas the main effect of Subordination, as well as the interaction between Gerund and Subordination was not found to affect the reading times of the direct objects in the target sentences.

[10]https://www.prolific.co/, last access 24 June 2020.

Table 2: Predictor variables in the linear mixed-effects regression model predicting reading time of the direct object

Variable	**Variable type**	**Description**	**Levels/Range**
GERUND	Fixed	Presence and informativity of gerund	None
IMMEDIACY	Fixed	Type of temporal complementation	Awa ("a while after"), Asa ("as soon as")
AGE	Fixed	Age of participant	-0.9-5.6 (scaled z-value)
SEX	Fixed	Sex of participant	Female (RL), Male
WORDFREQ	Fixed	Frequency of the direct object in the COCA/BNC corpus (z-value)	-0.6-3.5 (scaled z-value)
TRIAL	Fixed	Position of story in the experiment (z-value)	-1.6-1.8 (scaled z-value)
TEXTID	Random	ID of target story	
ID	Random	Participant ID	

Figure 3 visualizes the effect of the interaction between GERUND and SUBORDINATION on the reading times of the direct objects. It demonstrates that reading times of the objects are lowest for target sentences in which no gerund was present and highest for target sentences with an informative gerund, with target sentences with an uninformative gerund taking a middle position. However, pair-wise testing revealed the difference between target sentences with informative and uninformative sentences not to reach statistical significance. While the plot seems to indicate an interaction between the two variables in the condition of uninformative gerunds, this effect is much too weak to be considered significant.

3.3.5 Discussion

The analysis of the data from the reading-time experiment confirmed our prediction regarding the reading times on the direct object with respect to the role of the presence and informativity of the gerund. Thus, reading times for the object were lower for both target sentences involving an uninformative gerund and in particular, target sentences involving no gerund, than for target sentences involving an informative gerund. This is due to the difference in informativity of the action associated with these different conditions. The low reading times for sentences without a gerund are due to the fact that the referent of the direct object can easily be established from the general scenario introduced in the introductory sentence via bridging inference (e.g., 'medieval fair' - 'sword'). The relatively low reading times of sentences with an uninformative gerund are due to the high degree of mutual informativity between the gerund and the direct object (e.g., 'to forge' - 'sword'). In contrast, an informative action such as 'to choose' leads to a greater uncertainty in the reader's expectation regarding the referent of the direct object. This uncertainty is reflected in longer reading times

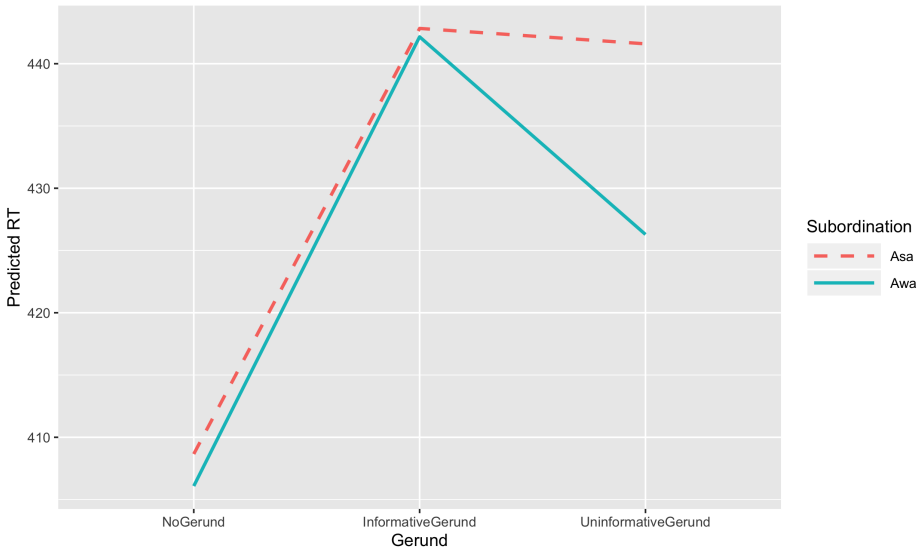

Figure 3: Predicted reading times of the direct object in the target sentence, by GERUND and SUBORDINATION

of target sentences involving informative gerunds.

However, our analysis did not show any evidence for an influence of the variable IMMEDIACY, reflecting the difference between immediate succession ('as soon as') or time-lapsed succession ('a while after') of the events in the subordinate and main clause on the reading times of the direct objects. Even more crucially, no evidence for an interaction effect between GERUND and IMMEDIACY was found, indicating that the effect by informativity is not moderated by the difference between immediate and time-lapsed succession.

Consequently, our analysis does not find evidence for the assumption that a hearer-based reanalysis of the use of gerunds in English FINISH constructions has taken place. While our corpus study has shown that speakers routinely employ gerunds in order to highlight discourse progression and that this effect is stronger for uninformative gerunds at least in some syntactic contexts, hearers have not reanalyzed this usage of gerunds as indicating *solely* discourse progression. Rather, they are still mainly sensitive to the parameter of informativity, failing to pay attention to the parameter of immediacy. Due to this lack of a hearer-based reanalysis of the construction, it seems that unlike Early Spanish FINISH constructions, it is unlikely for English FINISH constructions to grammaticalize into markers of recent past.

4 Conclusions

Historical linguistic theory, together with a corpus study of Spanish, supports an informativity-based account of the acquisition of an overt verbal complement by the *acabar* + direct object and its subsequent grammaticalization. However, this account is based strongly on assumptions related to cognition and communication; moreover, these assumptions are made with respect to long-dead speakers based on written texts. As such, it is of interest to investigate whether this account applies to present-day users of other languages. In doing so, our paper offers answers to the question of why similar lexical items undergo particular grammaticalization processes in some languages, but not others. In essence, we are exploring some of the variance not covered by the Lexical Determinism Hypothesis.

The corpus study described above (Section 3) showed that our hypotheses were largely borne out. First, the use of FINISH + direct object constructions is governed by informativity. Speakers are more likely to use a gerund in contexts in which the action described by the gerund cannot be predicted from the referent of the direct object. Second, whether or not a gerund is used in FINISH + direct object constructions is due to narrative progression contexts: speakers are more likely to use a gerund in temporal subordination clauses indicating anteriority to the event in the main clause.

However, the use of these FINISH constructions differs in one important point in Early Spanish and English. Like Early Spanish speakers, English speakers exploit FINISH constructions in order to highlight narrative progression, but unlike Early Spanish *acabar* + *de* + infinitive, these constructions are still mainly sensitive to the informativity of the finished event. In other words, and crucially, we do not find evidence for the assumption that English has developed a close and systematic association between anterior temporal contexts and uninformative overt verbal complements.

The perception experiment presented above (Section 3.3) confirmed the findings from the corpus study in terms of processing: English-speaking addressees are sensitive to the presence and informativity of gerunds, but not to the narrative progression function. Since this aspect was proposed to be crucial for the grammaticalization of the Spanish anterior, it makes sense that English did not grammaticalize an anterior out of 'finish' and is unlikely to do so. The motivation for the use of English FINISH + gerund constructions can always be attributed to considerations of informativity even in contexts that are prone to the discourse progression reading. As a result, the presumed preconditions for a coerced reanalysis are not present in English; listeners do not need to attribute a new meaning to FINISH + gerund constructions.

Our study asks a type of question not often asked (or answered) in historical linguistics: why a particular linguistic property did not develop in a particular language. Our results suggest that the black box in which answers to our 'why' questions reside can be explored in an empirical fashion. Specifically, we propose that a particular aspect of language structure and language use can block or inhibit a cross-linguistically frequent developmental pathway; the fact that

English speakers do not conventionally associate the use of FINISH constructions with discourse progression contexts ultimately blocks the grammaticalization of these constructions as anteriors. Moreover, our study shows that such 'blocking' effects can be identified using an experimental approach.

Consequently, in order to explain why certain grammaticalization paths do *not* recur, we clearly have to look beyond the Lexical Determination Hypothesis. In other words, the universality of lexical determination resides in the fact that the use of certain lexical items may lead to specific implicatures that may in a second step become conventionalized, leading to grammaticalization processes in a third step. While our results contribute to explaining this third step, we do not have a solution to the second step, i.e., why English speakers did not conventionalize the use of overt gerunds in FINISH constructions to highlight discourse progressions. However, we hope to have shown that the part of language change that must be attributed to chance can be reduced to some extent.

More broadly, the implications of this study are as follows. Historical linguists are interested in explaining how and why language change occurs. There are many and diverse explanations on offer for nearly every type of historical change. Typically (and ideally), these explanations are consistent with the data. But this, we think, is not enough. Historical linguists should treat these explanatory accounts involving speakers, listeners, children, communities, and so on, as hypotheses to be tested by other research methods, such as corpus research, simulations, psycholinguistic experiments and more. Our study joins others that have taken an experimental approach to the testing of the hypotheses of historical linguistics (De Smet and Van de Velde 2017; Hilpert and Saavedra 2018).

Acknowledgments

The authors thank the reviewers for the constructive and helpful comments, as well as Haim Dubossarsky, Uta Reinöhl, Shira Tal and Freek van de Velde for their detailed comments on earlier drafts. We are also grateful to the audiences at the workshop 'Beyond Time' (Brussels, Belgium, 7 February 2020), the Academia Grammaticorum Salensis Septima Decima (Salos, Lithuania, 28 July 2020) and the Center for the Study of Language and Society (Bern, Switzerland, 28 October 2020), for their valuable feedback. Malte Rosemeyer thanks Albert Wall for his help with the implementation of the experimental paradigm. Eitan Grossman thanks the Mandel Scholion Research Center of the Hebrew University for its support of the Historical Linguistics and Formal Semantics research group, as well as the members and guests of the group, in particular Omri Amiraz, Elitzur Bar-Asher Siegal, Noa Bassel, Nora Boneh, and Shira Tal, for many hours of inspiring discussions.

References

Andersen, Henning (ed.). 2001. *Actualization: Linguistic change in progress. papers from a workshop held at the 14th International Conference on Historical Linguistics, Vancouver, B.C., 14 August 1999.* Amsterdam: Benjamins.

Ansaldo, Umberto and L. Lim. 2004. Phonetic absence as syntactic prominence: Grammaticalization in isolating tonal languages. In Olga Fischer, Muriel Norde and Harry Perridon (eds.), *Up and down the cline: the nature of grammaticalization*, 345–362. Amsterdam: John Benjamin.

Ariel, Mira. 2008. *Pragmatics and grammar*. Cambridge: Cambridge University Press.

Bar-Asher Siegal, Elitzur A. 2020. A formal approach to reanalysis: The case of a negative counterfactual marker. In *Formal Approaches to Grammaticalization, LSA Proceedings*, vol. 5 2, .

Bates, Douglas, Martin Mächler, Ben Bolker and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:10.18637/jss.v067.i01.

Bickel, Balthasar. 2015. Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine and Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis, 2nd edition*, 901–923. Oxford: Oxford University Press.

Bickel, Balthasar. 2017. Areas and universals. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 40–54. Cambridge: Cambridge University Press.

Blaette, Andreas. 2020. polmineR: Verbs and Nouns for Corpus Analysis. R package version 0.8.2.

Burridge, Kathryn. 2006. Language contact and convergence in Pennsylvania German. In R. M. W. Dixon and Alexandra Y. Aikhenvald (eds.), *Grammars in contact. A cross-linguistic typology*, 179–200. Oxford: Oxford University Press.

Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.

Davies, Mark. 2017a. British National Corpus (BNC). Available online at https://www.english-corpora.org/bnc/. Last access 13 May 2019.

Davies, Mark. 2017b. Corpus of Contemporary American English (COCA). Available online at https://www.english-corpora.org/coca/. Last access 13 May 2019.

De Smet, Hendrik and Freek Van de Velde. 2013. Serving two masters: form-function friction in language change. *Studies in Language* 371(3). 534–565.

De Smet, Hendrik and Freek Van de Velde. 2017. Experimenting on the past: A case study on changing analysability in English *ly*-adverbs. *English Language and Linguistics* 21. 317–340.

De Vogelaer, Gunther. 2010. Morphological change in continental West Germanic: Towards an analogical map. *Diachronica* 27(1). 1–31.

Detges, Ulrich and Richard Waltereit. 2002. Grammaticalization vs. reanalysis: a semantic-pragmatic account of functional change in grammar. *Zeitschrift für Sprachwissenschaft* 21(2). 151–195.

Drinka, Bridget. 2017. *Language contact in Europe: The periphrastic perfect through history*. Cambridge: Cambridge University Press.

Drummond, Alex. 2020. Ibex. Internet Based EXperiments. Available online at https://spellout.net/ibexfarm/. Last access 24 June 2020.

Eckardt, Regina. 2006. *Meaning change in grammaticalization. An enquiry into semantic renalysis*. Oxford: Oxford University Press.

Eckardt, Regina. 2009. APO: Avoid pragmatic overload. In Maj-Britt Mosegaard Hansen and Jacqueline Visconti (eds.), *Current trends in diachronic semantics and pragmatics*, 21–41. Bingley: Emerald.

Ehmer, Oliver and Malte Rosemeyer. 2018. Inferences in interaction and language change. *Open Linguistics* 4. 536–551.

Enfield, Nicholas J. 1994. *Linguistic epidemiology: Semantics and grammar of language contact in mainland Southeast Asia*. London: Routledge Curzon.

Fischer, Olga. 1997. The grammaticalisation of infinitival *to* in English compared with German and Dutch. In Raymond Hickey and S. Puppel (eds.), *Language history and linguistic modelling, a Festschrift for Jacek Fisiak on his 60th birthday*, 265–280. Berlin: Mouton de Gruyter.

Fischer, Olga. 2007. *Morphosyntactic change. Functional and formal perspectives*. Oxford: Oxford University Press.

Fox, John and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Thousand Oaks; CA. http://tinyurl.com/carbook> (last access 31 August 2020): SAGE 3rd edn.

Gafter, Roey J., Scott Spicer and Mira Ariel. 2019. How does bring (not) change to give? *Folia Linguistica* 53(2). 443–477.

van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: Benjamins.

Grice, H. P. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan (eds.), *Syntax and semantics: Vol. 3: Speech acts*, 41–58. New York: Academic Press.

Gries, Stefan Thomas. 2009. *Quantitative Corpus Linguistics with R. A Practical Introduction*. New York, London: Routledge.

Gries, Stefan Thomas. 2015. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10(1). 95–125. doi:10.3366/cor.2015.0068.

Grossman, Eitan and Ira Noveck. 2015. What can historical linguistics and experimental pragmatics offer each other? *Linguistics Vanguard* 1(1). 145–153.

Grossman, Eitan and Stéphane Polis. 2014. On the pragmatics of subjectification: the grammaticalization of verbless allative futures (with a case study in Ancient Egyptian). *Acta Linguistica Hafniensia* 46(1). 1–40.

Grupo Gramática y Diacronía. 2015. *Corpus GRADIA*. http://gradiadiacronia.wixsite.com/gradia/corpus-gradia. Last access 18 September 2016.

Hagoort, Peter, Lea Hald, Marcel Bastiaansen and Karl Magnus Petersson. 2004. Integration of word meaning and world knowledge in language comprehension. *Science* 304(5669). 438–441. doi:10.1126/science.1095455.

Hale, Mark. 1998. Diachronic syntax. *Syntax* 1. 1–18.

Heine, Bernd and Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Hilpert, Martin. 2014. Collostructional analysis. measuring associations between constructions and lexical elements. In Dylan Glynn and Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 391–404. Amsterdam: Benjamins.

Hilpert, Martin and David Correia Saavedra. 2018. The unidirectionality of semantic changes in grammaticalization: an experimental approach to the asymmetric priming hypothesis. *English Language and Linguistics* 22(3). 357–380.

Hlavac, Marek. 2018. Stargazer: Well-formatted regression and summary statistics tables. R package version 5.2.2.

Hopper, Paul J. and Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press 2nd edn.

Huang, Yan. 2011. Types of inference: entailment, presupposition, and implicature. In Wolfram Bublitz and Neal R. Norrick (eds.), *Foundations of pragmatics*, 397–421. Berlin: De Gruyter.

Kiparsky, Paul. 2012. Grammaticalization as optimization. In Dianne Jonas, John Whitman and Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcomes*, 15–51. Oxford: Oxford University Press.

Kleiber, Georges. 1999. *Problèmes de sémantique. La polysémie en questions*. Paris: Presses universitaires du Septentrion.

Lauwers, Peter and Dominique Willems. 2011. Coercion: Definition and challenges, current approaches, and new trends. *Linguistics* 49(6). 1219–1235.

Manning, Christopher D. and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Michaelis, Laura A. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15. 1–67.

Oakes, Michael P. 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, Mass.: MIT Press.

R Development Core Team. 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available online at http://www.R-project.org. Last access 15 November 2017.

Roberts, Ian and Anna Roussou. 2003. *Syntactic change. A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.

Rohde, Hannah. 2020. If you don't have anything nice interesting to say, don't say anything at all. Talk given at CUNY 2020. Available online at https://osf.io/e3dxq/. Last access 1 September 2020.

Rosemeyer, Malte and Eitan Grossman. 2017. The road to auxiliariness revisited: The grammaticalization of FINISH anteriors in spanish. *Diachronica* 34(4). 516–558.

Schneider, Ulrike. 2018. ΔP as a measure of collocation strength. *Corpus Linguistics and Linguistic Theory* .

Schwenter, Scott and Richard Waltereit. 2010. Presupposition accommodation and semantic change. In Kristin Davidse, Lieven Vandelanotte and Hubert Cuyckens (eds.), *Subjectification, intersubjectification, and grammaticalization*, 75–102. Berlin: De Gruyter.

de Smet, Hendrik. 2012. The course of actualization. *Language* 88(3). 601–633.

Stefanowitsch, Anatol and Stefan Th. Gries. 20 May. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1). 1 – 43.

Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.

Tagliamonte, Sali. 2012. *Variationist sociolinguistics. Change, observation, interpretation*. Malden, MA: Wiley-Blackwell.

Traugott, Elizabeth and Richard Dasher. 2003. *Regularity in semantic change*. Cambridge: Cambridge University Press.

Traugott, Elizabeth C. and Ekkehard König. 1991. The semantics-pragmatics of grammaticalization revisited. In Elizabeth C. Traugott and Bernd Heine (eds.), *Approaches to grammaticalization, i: Focus on theoretical and methodological issues*, 189–218. Amsterdam: John Benjamin.

Troyer, Melissa and Marta Kutas. 2018. Harry Potter and the Chamber of What?: the impact of what individuals know on word processing during reading. *Language, Cognition and Neuroscience* 35(5). 641–657. doi:10.1080/23273798.2018.1503309. https://doi.org/10.1080/23273798.2018.1503309.

Verstraete, Jean-Christophe. 2008. Same source, same target, different paths: from 'person' to reflexive in Umpithamu and other Paman languages. *Language Sciences* 30. 599–620.

Veyrat Rigat, Montserrat. 1994. La perífrasis verbal *acabar de* + infinitivo y la resolución de su ambigüedad. *Vox Romanica* 53. 238–252.

Walkden, George. 2017. The actuation problem. In Adam Ledgeway and Ian Roberts (eds.), *The Cambridge handbook of historical syntax*, 403–424. Cambridge University Press.

Weinreich, Uriel, William Labov and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Winfred Lehamn and Yakov Malkiel (eds.), *Directions for historical linguistics*, 95–195. Austin: University of Texas Press.

Wickham, Harald. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

5 Appendix

Table 3: Results from the regression analysis over the use of FINISH constructions in the corpus data

	Dependent variable:
	Gerund_bin
After	1.170
	(0.760)
Determinationindef	0.003
	(1.601)
Determinationnone	3.001**
	(1.398)
Determinationposs	2.798***
	(0.999)
PMI_FinishObj	−0.290***
	(0.103)
AfterTRUE:Determinationindef	2.445
	(3.239)
AfterTRUE:Determinationnone	−5.261***
	(1.871)
AfterTRUE:Determinationposs	−4.538***
	(1.334)
AfterTRUE:PMI_FinishObj	−0.134
	(0.176)
Determinationindef:PMI_FinishObj	−0.123
	(0.377)
Determinationnone:PMI_FinishObj	−0.583**
	(0.271)
Determinationposs:PMI_FinishObj	−0.867***
	(0.241)
AfterTRUE:Determinationindef:PMI_FinishObj	−0.803

	(0.809)
AfterTRUE:Determinationnone:PMI_FinishObj	0.973***
	(0.361)
AfterTRUE:Determinationposs:PMI_FinishObj	0.957***
	(0.310)
Constant	0.990**
	(0.439)
Observations	660
Log Likelihood	−352.647
Akaike Inf. Crit.	739.293
Bayesian Inf. Crit.	815.661
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Full results from the linear mixed-effects regression model predicting the reading times of the direct object in the reading-time experiment

	Dependent variable:
	RT
GerundInformativeGerund	34.178**
	(15.871)
GerundUninformativeGerund	32.941**
	(15.494)
SubordinationAwa	−2.573
	(15.838)
Age	28.350*
	(16.569)
Sexmale	−62.572
	(40.205)
WordFreq	−15.799***
	(4.714)
Trial	−45.004***
	(4.431)
GerundInformativeGerund:SubordinationAwa	1.896
	(22.555)
GerundUninformativeGerund:SubordinationAwa	−12.735
	(21.982)
Constant	458.906***
	(37.072)
Observations	848
Log Likelihood	−5,336.686
Akaike Inf. Crit.	10,699.370
Bayesian Inf. Crit.	10,761.030
Note:	*p<0.1; **p<0.05; ***p<0.01