

# The Grammaticalization of Catalan *anar* ('to go') + Infinitive for the Expression of Perfective Past: A Diachronic, Corpus-Based Perspective

Marc Gandarillas<sup>1</sup>

<sup>1</sup> Department of Spanish and Portuguese Studies, University of Florida, FL, United States

Correspondence: Marc Gandarillas, Department of Spanish and Portuguese Studies, University of Florida, Daurer 170, PO Box 117405, Gainesville, FL, 32611-7405, United States. Tel: 1-352-392-2016. E-mail: m.gandarillascot@ufl.edu

Received: December 10, 2021; Accepted: January 9, 2022; Online Published: January 12, 2022

*The research is financed by La Caixa Foundation.*

## Abstract

This study constitutes an exploratory analysis of the grammaticalization cline of *anar* ('to go') + infinitive in Catalan to express perfective past (e.g., *va arribar* 's/he arrived'). Our research interest primarily lies in diachronically tracing the evolution of this grammatical change, which appears to be unprecedented in other Romance languages (e.g., Spanish, French), in which the construction has instead led to the expression of a near and/or intentional future. A gap in research is found in the fact that there have been few corpus-based, pragmatic approaches to the matter. We base our theoretical framework on the definition of grammaticalization by Hopper and Traugott (2003) and a number of related publications (Alturo 2017, Pérez-Saldanya & Hualde 2003). Critical items (N=346) were retrieved from the diachronic corpus CICA (11th-18th c.) and subsequently analyzed in the light of pragmatic factors, establishing a three-stage cline based on Segura (2012). Results show how informative bridging contexts are in shaping grammaticalization processes, as they highlight the challenges of tracing a grammaticalization process based on corpora of literary texts. A discussion follows on the identification of potential next steps that might be useful in complementing our own research.

**Keywords:** corpus linguistics, grammaticalization, language change, pragmaticalization, morphophonology, Romance languages

## 1. Introduction

This study constitutes an exploratory analysis of the grammaticalization path followed by the Catalan bigram *anar* ('to go') + infinitive between the earliest documentation of the language in the 11<sup>th</sup> century until the 18<sup>th</sup> century. We adopt here the narrow definition of grammaticalization proposed by Hopper and Traugott (2003: Ch. 1), according to which grammaticalization would be the process universally found in world languages by which a lexical item eventually loses its semantic characterization and becomes a functional morph specializing in expressing certain grammatical features (e.g., aspect, tense).

Grammaticalization has been broadly studied in both non-Romance (e.g., English, German) and Romance contexts (e.g., French, Italian), including the field of Catalan Studies. Focusing on the latter, however, most studies have traditionally adopted the broad definition of grammaticalization, which encompasses related processes that are more precisely characterized in terms of pragmaticalization or discoursivization (Alturo & Chodorowska-Pilch 2009, Cuenca & Massip 2005). (Note 1)

Synchronic evidence shows that the grammaticalization of 'to go' + infinitive is widespread among Romance languages, and, since no instances thereof are documented in Classical Latin, we need to assume that this has entirely been an intra-Romance process. That being said, we observe a relevant contrast between outcomes in Catalan and other Romance languages. On the one hand, in Spanish, Portuguese, Italian, and French the grammaticalization of—respectively—*ir/ir/andare/aller* 'to go' + infinitive has resulted in a periphrasis that specializes in the expression of intentional and/or near future:

- (1) Sp. *va a comer*, Port. *va comer*, It. *va a mangiare*, Fr. *il/elle va manger* 's/he is going to eat'

In contrast, the Catalan outcome shows that the homologous construction has specialized in the expression of perfective past (Note 2), which is unprecedented within the Romance scene (Note 3). Research on this contrast has to date been as scarce as it has been controversial (Juge 2005), with a majority of studies heavily relying on qualitative analysis (Juge 2005, Pérez-Saldanya & Hualde 2003). Based on such state of affairs, the goal of the present study is twofold: (1) to trace the grammaticalization cline of Catalan *anar* + infinitive using a corpus-based methodology; (2) to adopt and/or validate a particular stance on the contrast observed for Catalan within Romance.

## 2. Literature Review

Over the course of the last two decades, grammaticalization processes found in a variety of world languages have received increased scholarly attention—especially from theory-based, qualitative perspectives (Cuenca 2001, Montserrat i Buendia 2004) (Note 4). Research on grammaticalization has traditionally been comprised of case-by-case studies, although in more recent times some studies are found that focus comprehensively on a particular language and—even more recently—comparative analysis across languages (Fagard & Mardale 2012, Iacobini 2012). Along these lines, a significant publication that has paved the path for further, more in-depth research within the field of research in grammaticalization is found in the handbook written by Hopper and Traugott (2003). This work needs to be credited with providing a picture of the grammaticalization process through a variety of world language examples at a time in which the term grammaticalization was being used rather imprecisely.

As mentioned earlier, Hopper and Traugott (2003: Ch. 1) provide a narrow—yet comprehensive—definition of the term, regarding grammaticalization as a process by which a lexical item (e.g., noun, verb) eventually loses its semantic meaning and becomes a functional morph used to express a specific grammatical notion. Hopper and Traugott's (2003) work provides an inclusive, yet detailed overview of the grammaticalization process by delving into its pragmatic underpinnings (e.g., semantic bleaching, metaphor, metonymy) and providing abundant, carefully analyzed examples from multiple—related and unrelated—world languages. In more recent times, a number of publications have followed on Hopper and Traugott's (2003) view of grammaticalization (e.g., Bisang 2017).

Within the area of Catalan studies, grammaticalization has received considerable scholarly attention, although the term has usually resorted to its narrow definition, including instances of what Hopper and Traugott (2003) would rather consider as examples of pragmatization or discoursivization (Note 5). An example of this is provided by Alturo and Chodorowska-Pilch (2009), who trace the pragmatization process of *si us plau* from its origin as a conditional protasis (lit. 'if it pleases you') to a courtesy discourse marker. Although it might seem more appropriate to regard this as a case of pragmatization (Note 6), it must be noted that this kind of processes share multiple characteristics with more prototypical grammaticalization processes (i.e., generalization of meaning and semantic bleaching, increase in frequency and in number of contexts of usage, phonological simplification in popular speech (Note 7)). Similarly, Cuenca and Massip (2005) examine the pragmatization of a selection of conjunctive phrases such as *encara que* or *†jassia que* ('although'), which show virtually identical correspondences in other world languages, with similar underlying processes.

In recent years, the grammaticalization of the construction *anar* + infinitive as a functional past perfective has received considerable scholarly attention, especially within the field of Catalan Studies. Specifically, the *anar* + infinitive grammaticalization has been significantly studied either from morphophonological or morphosyntactic approaches (Pérez-Saldanya & Hualde 2003, Pérez-Saldanya 1998, Vallduví 1989) or in connection with pragmatic-discourse factors (Alturo 2017, Segura 2012). In the broader context of world language studies, mentions of this phenomenon are usually shallow or even anecdotal, yet some remarkable exceptions exist (e.g., Juge 2005 (Note 8)). However, potential biases might have been introduced by building upon previous theoretical models, as multiple approaches appear to be heavily theory-based—whereas grammaticalization, as a gradual process, seems to demand a more diachronic, corpus-based perspective.

Overall, little comprehensive research has been conducted on the process encompassing all language fields and from a comprehensive synchronic and diachronic approach. Alturo (2017) compares instances of the construction under study from medieval literature and modern usage examples as she establishes interesting connections between these and ad-hoc hypotheses based on previous literature (e.g., reanalysis, syncretism, language contact with Occitan). Pérez-Saldanya (1998) traces the evolution of the Catalan verbal system back to its Latin origin by approaching this inclusively—and functionally—in terms of diatopic varieties. From a corpus-based perspective, Pérez-Saldanya and Hualde (2003) find that, in most tokens, the infinitive corresponds to a verb of accomplishment (Note 9). Finally, Segura (2012) establishes the multistep evolutionary cline of the periphrasis *anar* + infinitive that we adopt in our study—specifically regarding its semantic characterization in terms of movement, directionality/purpose, and emphasis.

The pragmatic-discourse component appears to be a key addition to explaining the outcome(s) of every grammaticalization process. Even though some studies have been published on the matter, these are usually limited to enriching preexisting theoretical frameworks rather than conducting corpus-based research (Hopper & Traugott 2003: Ch. 4). Heine (2002) provides a relevant proposal that harmonizes a variety of pragmatic factors crucial in shaping grammaticalization processes—specifically, frequency of use, reasoning or inferential processes, transfer mechanisms (i.e., metaphor and metonym (Note 10)), directionality (Note 11), and semantic implications (e.g., bleaching, generalization). When eventually focusing on the context, Heine (2002) is especially interested in determining the contextual requirements that allow for—or, at least, do not hinder—the evolution of grammatical meanings. Within this context, the author uses terms such as *source vs. target meaning*, *conventionalization*, and *bridging context*. The latter is especially relevant to our own research, since it would appear that a corpus-based approach to a grammaticalization process needs to account for both pre-grammaticalized, grammaticalized structures, and—relevantly—the “bridging contexts” that provide a plausible explanation on how the transfer occurred.

### 3. Current Study: Rationale and Research Questions

As earlier described in the Introduction and Literature Review sections, the present study intends to shed light onto the research gaps found in the grammaticalization process followed by *anar* + infinitive, which eventually resulted in a periphrasis used for the expression of perfective past. Grammaticalization theories have been carefully rethought and refined over the last two decades, especially by constraining the definition of the process (Hopper & Traugott 2003: Ch. 1). Grammaticalization processes appear to be universally present in world languages, even though these are operationalized differently according to language typology, pragmatic factors (Heine 2002), and sociolinguistic variables (Hopper & Traugott 2003: Ch. 1, Traugott, 1995).

The Catalan outcome of *anar* + infinitive is scientifically relevant to the field of grammaticalization studies, both within the sphere of Catalan studies and also more broadly (i.e., Romance and world languages). In the words of Pérez-Saldanya and Hualde (2003), this outcome is “highly anomalous” within the Romance sphere, with a controversial multistep cline that appears to demand a comprehensive approach. A number of studies have been conducted on this particular process, yet following a heavily theory-based approach that relies more on introspection than actual data of language use. While acknowledging the importance of introspection as a tentative approach, the use of prototypical examples can often lead to a reductionistic view of the grammaticalization process as a whole. Introspection-based examples tend to be simpler and more systematic than their corpus-based counterparts (e.g., *en Joan va arribar ahir a les tres* ‘John arrived at three yesterday’ is often found with less-prototypical syntactic structures, usually due to expressive needs (Note 12)—some of these less-prototypical structures are unlikely to arise from introspective analysis).

Regardless of the type of critical items that the researcher decides to adopt as their own, morphophonological and/or syntactic approaches definitely appear to be a logical, necessary first step (Pérez-Saldanya & Hualde 2003, Pérez-Saldanya 1998, Vallduví 1989). However, research has shown that such approaches are usually insufficient, opaque, and—to a large extent—inconclusive. In our view, the aforementioned state of affairs motivates the need for further analysis that—concurrently with building on previous theoretical frameworks—incorporates pragmatic-discourse factors and, in so doing, provides a comprehensive approach to the phenomenon under study. Following Hopper and Traugott (2003: Ch. 4) and Heine (2002) as some of the few studies that comprehensively tackle grammaticalization from a pragmatic-discourse perspective, we adhere to the multistep cline model, which has been a constant trend for years in the study of grammaticalization (Norde 2019, Bisang 2017). Additionally, we intend to harmonize more theoretical, synchronic approaches with more corpus-based, diachronic ones.

Last but not least, we need to acknowledge that the process of grammaticalization occurs in spontaneous speech. The lack of access to the oral language of medieval Catalan certainly encourages us to tell a cautionary tale before reaching hurried conclusions, since written samples of the language are usually more conservative than popular speech in embracing language change. Based on the research gaps found in the literature reviewed, the following research questions were established:

1. How could a corpus-based approach inform the grammaticalization process of Catalan *anar* + infinitive?
2. Based on the findings from the previous research question, is it possible to support a particular stance regarding the contrast between Catalan and other Romance language in the evolution of the periphrasis?

### 4. Methods: Materials, Data Management, and Analysis

A corpus-based search was conducted using the electronic version of the Corpus Informatitzat del Català Antic (CICA (Note 13)), designed by members of the Institute for Catalan Studies (Note 14) and other academic

institutions from Catalan-speaking territories (Note 15). This corpus is comprised of a variety of literary texts (e.g., fiction prose, chronicles, religious and administrative works, poetry) that range from the earliest documentation of the language (11<sup>th</sup> century) until the 1750s. All *anar* + infinitive bigrams (N=346) were retrieved for the 3<sup>rd</sup> person singular and plural (Note 16) (i.e., auxiliaries *va* (N=173) and *van* (N=101) or *varen* (N=72), respectively, + infinitive ending in *-r(e)*). Next, all non-complying tokens were manually filtered out and not considered further into the research. Every critical item was expanded to a minimum surrounding context of 8L-8R for further analysis of contextual and pragmatic cues (e.g., co-occurrence with other past tenses, most relevantly the *perfet simple*).

Once all relevant bigrams—along with their surrounding contexts—were retrieved, a threefold classification was established based on previous theoretical frameworks (especially Segura 2012; but also Juge 2005, Pérez-Saldanya 1998, Pérez-Saldanya & Hualde 2003), after which all critical items underwent a manual labeling process according to the following distinctive stages of grammaticalization: (1) non-grammaticalized (i.e., infinitives of motion; cf. Eng. *I am going to the store*); (2) bridging context (i.e., infinitives expressing directionality/purpose; cf. Eng. *I am going to travel*); (3) grammaticalized (i.e., deriving from a merely emphatic use; cf. Eng. *I am going to buy that book*). However, it is important to note that in some instances an expression that appears to have completed its grammaticalization cline can become even more grammaticalized by undergoing a morphophonological simplification process (e.g., *I am going to buy that book* > *I'm gonna buy that book* > *I'mma buy that book*). Here are some examples that illustrate this multistep grammaticalization path (Note 17):

- (2) Non-grammaticalized (13<sup>b</sup>): *E el rey [...] va prendre ses armes, e muntà a caval e comensà a córrer* ‘And the king [...] grabbed his weapons, rode his horse, and started to flee’ (*Crònica [B. Desclot]*, p. II.43, l. 6) → This example is especially enlightening, since the periphrastic construction is here coordinated with two traditional preterits, *mundà* (‘he rode’) and *començà* (‘he started’). From this, one can infer that there is a clear underlying contrast between the use of the two.
- (3) Bridging context (14<sup>b</sup>): *E alsunes vegades aquells qui-s van combatre no-s combaten* ‘And sometimes those who fought each other (or who went to fight each other) no longer fight each other’ (*Epistolari de Pere III*, 2, Carta 20, l. 58) → In this context, the directionality/purpose notion is still present in the semantic load of the verb, which allows for a twofold interpretation.
- (4) Grammaticalized (18<sup>a</sup>): [...] *en una mà tan destre que a tots gran gust va donar* ‘[...] with such a skilled hand that this brought [lit. gave] great pleasure to everyone’ (*Memòries d'un capellà del segle XVIII*, p. 81, l. 30) → Here, the auxiliary *anar* appears followed by the infinitive *donar* ‘to give,’ which no longer seems to allow for an interpretation in terms of movement. This shows that the construction has reached the end of the grammaticalization cline.

With the CICA corpus examples at hand, the theoretical frameworks described in the Literature Review section were considered and contrasted with this data, especially regarding the idea that the pragmatic component is relevant to the study of this grammaticalization process (Hopper & Traugott 2003: Ch. 4, Heine 2002)—potentially allowing for drawing preliminary conclusions that account for the contrast in terms of outcomes between Catalan and other Romance languages.

### 5. Results and Discussion

In order to tentatively approach our first research question, Figure 1 provides an overview of how the *anar* + infinitive construction quantitatively evolves over the course of the centuries. No instances of the construction were found during the 11<sup>th</sup> and 12<sup>th</sup> centuries, while the construction consistently appears henceforth. Especially relevant is the number of tokens found for the 15<sup>th</sup> and 17<sup>th</sup> centuries, which respectively account for 44.80% and 33.53% of the total number of critical items. This alone suggests that the grammaticalization cline followed by *anar* + infinitive is not fully completed in the most recent examples from the CICA corpus, which is most likely due to the fact that there was another competing form for the expression of perfective past (i.e., the *perfet simple*, as shown above in example (2)).

| cent. | 11 | 12 | 13   | 14    | 15   | 16   | 17    | 18   | TOTAL |
|-------|----|----|------|-------|------|------|-------|------|-------|
| n     | 0  | 0  | 19   | 155   | 30   | 18   | 116   | 8    | 346   |
| %     | 0  | 0  | 5.49 | 44.80 | 8.67 | 5.20 | 33.53 | 2.31 | 100   |

Figure 1. Raw frequencies and percentages of 3<sup>rd</sup> person *anar* + infinitive in the CICA corpus (11<sup>th</sup>-18<sup>th</sup> c.)

A synchronic look at modern Catalan shows that the *perfet simple* (e.g., *arribà* vs. *va arribar* ‘s/he arrived’) has been long lost in most spoken varieties of the language; however, it has been maintained in archaizing literature. Again, it is essential to note that corpus-based diachronic data does not reflect—for the most part—popular speech,

yet it is an indispensable tool in searching for trends and patterns. An instant implication of this is that data in Figure 1 needs to be interpreted cautiously. For instance, the fact that there are more tokens of the construction in the 17<sup>th</sup> century than in the 18<sup>th</sup> century does not imply that the expression became less used, but rather that, during the latter century, it became a trend to use the *perfet simple* as an alternative in writing (probably because *anar* + infinitive had by then become so widespread in popular speech that authors tended to avoid it in formal writing). In other words, we know for a fact that the construction *anar* + infinitive eventually superseded the *perfet simple* in spontaneous speech, yet data in Figure 1 just shows the figures regarding the use of the construction in literary texts.

Given the scope of our research, Figure 2 below is more informative, since it reflects the evolution of *anar* + infinitive in terms of grammaticalization, according to the three degrees indicated in the Methods section (Note 18). Specifically, Figure 2 shows a quantitative trend of the construction towards becoming more grammaticalized over time, with non-grammaticalization instances gradually disappearing and bridging contexts (i.e., twofold interpretations) possibly occurring until at least the 18<sup>th</sup> century.

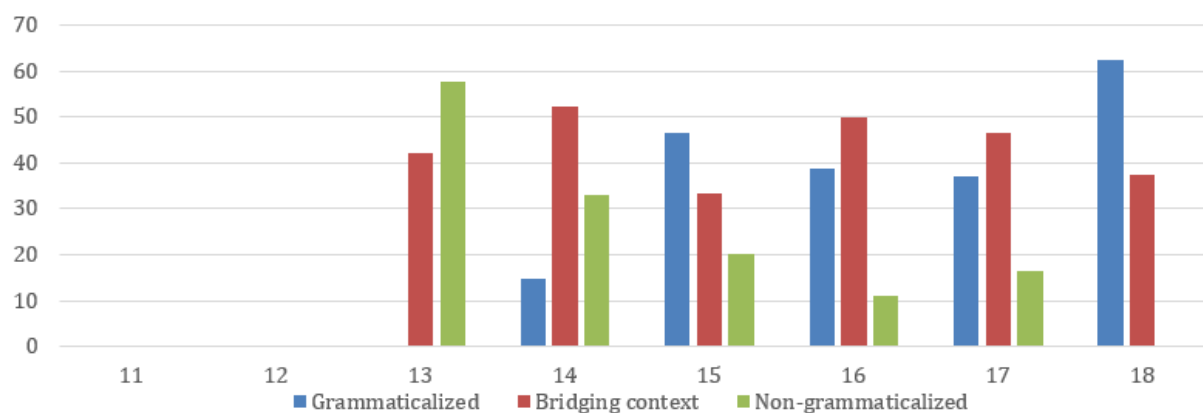


Figure 2. Percentages of 3<sup>rd</sup> person *anar* + infinitive in the CICA corpus according to the three degrees of grammaticalization (i.e., grammaticalized (G), bridging context (BC), and non-grammaticalized (NG))

Grammaticalized expressions tend to appear in contexts of motion in which *anar* can be analyzed as a “narrative/atemporal present” (Juge 2005), whereas in non-grammaticalized expressions motion is no longer a possibility based on the nature of the infinitive. Bridging contexts primarily admit a twofold interpretation; however, they appear to show higher degrees of grammaticalization as they progress further into the grammaticalization process. In accordance with previous findings (Pérez-Saldanya 1998, Segura 2002), the pragmatic component appears to have been relevant in shaping the Catalan outcome of *anar* (‘to go’) + infinitive, which contrasts with other Romance languages. Relatedly, our results show that it is indeed possible and useful to establish Segura’s (2012) threefold cline ‘movement > directionality/purpose (Note 19) > emphasis.’ Specifically, motion-implying infinitives (e.g., *va comprar* ‘s/he went shopping’) appear to constitute bridging contexts (Heine 2002), in the sense that they can be alternatively interpreted as either grammaticalized or non-grammaticalized. In the final stage of the cline (i.e., “emphasis”), the grammaticalization process shows evidence for having reached full completion (clearly from the 18<sup>th</sup> century; likely before).

**6. Conclusion**

Our results appear to be consistent with the literature reviewed (e.g., Pérez-Saldanya & Hualde 2003, Segura 2002, Pérez-Saldanya 1998) and, along these lines, they certainly add another tool to the toolbox for corpus-based exploration of the grammaticalization process of *anar* + infinitive. According to this, it appears to be important to reconcile theoretical frameworks with quantitative, corpus-based data. However, challenges remain, especially concerning the difficult classification of certain critical items whose contexts are mostly opaque. As mentioned earlier, in historical linguistics, corpus-based methodologies definitely pose a number of constraints, with written samples usually not reflecting spontaneous language use—which appears to be crucial to fully comprehend how grammaticalization operates as a strongly communication-based process. That being said, in the absence of spoken samples of medieval Catalan, corpus data is essential to complement theoretical frameworks and provide valuable, consistent information in terms of general development trends. Additionally, our results appear to be relevant also in connection with the theoretical foundations laid by Hopper and Traugott (2003) and, concurrently, they appear to point to several directions toward future research. Still, further analysis is required to validate and refine these

results. For instance, it might be interesting to verify if our results are consistent with those obtained for the first and/or second persons singular and plural (along the lines of Juge's (2005) proposal, for whom the traditional first person plural, *anam*, provides a potential morphological bridging context). Also, we believe that it might be relevant to observe how these figures vary across genres in the corpus.

Finally, an important takeaway appears to lie in the fact that the *anar* + infinitive periphrasis coexists with the traditional synthetic tense, the so-called *perfet simple*, for the expression of perfective past. In most spoken varieties of modern Catalan this tense has been completely relegated by the construction that we have studied. Therefore, in prospective research, it might be potentially interesting to explore (also from a diachronic corpus-based approach) to what extent these alternatives could appear in complementary distribution in the medieval language, as well as how the aforementioned overlap eventually succeeded in spontaneous speech. For this, our results appear to emphasize the relevance of the pragmatic factors mentioned in some of the literature reviewed (Hopper and Traugott 2003: Ch. 4, Traugott 1995, Heine 2002).

## References

- Alturo, N. (2017). L'auxiliar de passat *anar* en català. *Apunts de pragmàtica del català. Teories de la gramaticalització: Cas 1*. Barcelona: Dipòsit Digital de la Universitat de Barcelona, 2-9.
- Alturo, N., & Chodorowska-Pilch, M. (2009). La gramaticalització de *sisplau*. *Els Marges*, 88, 15-38. SSN 0210-0452
- Bisang, W. (2017). *Grammaticalization*. Oxford Research Encyclopedia of Linguistics. <https://doi.org/10.1093/acrefore/9780199384655.013.103>
- Cuenca, M. J. (2001). La gramaticalització com a teoria de la variació morfosintàctica. *La gramàtica i la semàntica en l'estudi de la variació*, 151-185. Barcelona: PPU. DL B-4.289-2016
- Cuenca, M. J., & Massip, À. (2005). Connectors i processos de gramaticalització. *Caplletra*, 38, 77-92.
- Fagard, B., & Mardale, A. (2012). The pace of grammaticalization and the evolution of prepositional systems: Data from Romance. *Folia Linguistica*, 46(2), 303-340. <https://doi.org/10.1515/flin.2012.011>
- Heine, B. (2002). On the role of context in grammaticalization. *New Reflections on Grammaticalization*, 696-707. Oxford: Oxford University Press. <https://doi.org/10.1075/tsl.49.08hei>
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Oxford: Oxford University Press. <https://doi.org/10.1017/CBO9781139165525>
- Iacobini, C. (2012). Grammaticalization and innovation in the encoding of motion events. *Folia Linguistica*, 46(2), 359-385. <https://doi.org/10.1515/flin.2012.013>
- Juge, M. L. (2005). Morphological factors in the grammaticalization of the Catalan *go* → past. *Diachronica*, 23(2), 313-339. ISSN 0176-4225 / E-ISSN 1 569-9714. <https://doi.org/10.1075/dia.23.2.05jug>
- Montserrat i Buendia, S. (2004). Entre la dixi i la definitud: Els verbs de moviment resultatiu en català. *Caplletra*, 39, 61-83.
- Norde, M. (2019). Grammaticalization in morphology. *Oxford Research Encyclopedia of Linguistics*, vol. Morphology. <https://doi.org/10.1093/acrefore/9780199384655.013.733>
- Pérez-Saldanya, M. (1998). *Del llatí al català: Morfosintaxi verbal històrica*. Col·lecció Biblioteca Lingüística Catalana. València: Universitat de València.
- Pérez-Saldanya, M., & Hualde, J. I. (2003). On the origin and evolution of the Catalan periphrastic preterit. *Verbalperiphrasen in den (ibero)romanischen Sprachen*. Hamburg: Buske. 46-70. ISSN 0947-0573
- Segura-Llopes, C. (2012). El passat perifràstic en català antic. Una revisió a partir d'estudi de corpus. *eHumanista/IVITRA*, 2, 118-147. ISSN 1540 5877
- Traugott, E. C. (1995). Subjectification in grammaticalisation. *Subjectivity and Subjectivization: Linguistic Perspectives*, 31-54. Cambridge: Cambridge University Press. ISBN-13 978-0-521-47039-1
- Vallduví, E. (1989). Sobre el *perfet* perifràstic del català. *Actes del cinquè col·loqui d'estudis catalans a nord-Amèrica*. Barcelona: Publicacions de l'Abadia de Montserrat. 85-98. <https://doi.org/10.1017/CBO9780511554469.003>

## Notes

Note 1. To provide a relevant example, the conditional protasis *si us plau* ‘[lit.] if it pleases you’ eventually becomes a discourse marker meaning ‘please,’ thus leaving the grammar sphere behind (Alturo & Chodorowska-Pilch 2009).

Note 2. This is, a past in which the action conveyed by the verb is clearly expressed as complete (e.g., *va menjar* ‘s/he ate’).

Note 3. See Alturo 2017, Pérez-Saldanya 1998, Pérez-Saldanya & Hualde 2003, Segura 2012, Vallduví 1989.

Note 4. Along these lines, the weight of the pragmatic component within grammaticalization processes has received considerable scholarly attention (Traugott 1995, Heine 2002, Alturo & Chodorowska-Pilch 2009). Still, most research is limited to tracing the morphophonological evolution of the forms under study, which appears to be more easily retrievable and less subject to speculation, yet far from self-sufficient to account for every change within the cline (Norde 2019, Cuenca & Massip 2005, Pérez-Saldanya 1998, Pérez-Saldanya & Hualde 2003).

Note 5. Pragmaticalization occurs when the evolving expression gradually becomes less grammaticalized over time, to the extent that it is eventually transferred from the language internal structure to the communicative domain.

Note 6. In fact, the movement here has proceeded from more to less grammaticalized. *Si us plau*, which originally agrees with *vós* (traditional courtesy pronoun in Catalan), is also used with *tu* (informal treatment). In other words, it appears to be fossilized in this form, regardless of the surrounding grammatical context.

Note 7. In popular Catalan, *si us plau* ‘please’ is commonly articulatorily shortened as *sisplau*.

Note 8. This article specifically focuses on the change that occurred based on the early association of preterit and present auxiliaries, which resulted in the exclusive use of the present, as well as the parallel development of an alternative series of auxiliaries. Specifically, the author proposes that the present-preterit syncretism found in old Catalan first person plural *anam*—to date preserved in the Balearic varieties—at some point allowed for a reanalysis of the construction. This analysis contrasts with previous proposals, which support an explanation along the lines of so-called narrative or atemporal present.

Note 9. Pérez-Saldanya and Hualde (2003) also notice that a subsequent action in the past tense is usually present within the immediate context.

Note 10. While the role of metonym as a “problem-solving device” (Hopper & Traugott 2003: Ch. 4) is embraced by some authors, metaphor appears to be more controversial, since it leads to more abstract explanations within the sphere of philosophy of language. However, in recent years, both terms tend to be avoided while terms such as semantic bleaching, analogy, or reanalysis are more commonly accepted by the community of researchers.

Note 11. The contrast between abstraction and concretion is mentioned as a relevant one in terms of directionality.

Note 12. Notoriously, when the (optional) emphatic negative particle *pas* is used (cf. Fr. *pas*), this can be inserted between the auxiliary form of *anar* and the infinitive (e.g., *en Joan no va pas arribar ahir a les tres* ‘John did absolutely not arrive yesterday at three’—also possible: *en Joan no va arribar pas ahir a les tres* [same meaning]), which shows that the expression, though being fully grammaticalized, is still not fully lexicalized (i.e., treated as a sole lexical item).

Note 13. The CICA corpus can be accessed electronically by following the link [www.cica.cat](http://www.cica.cat).

Note 14. In Catalan this institution is known as the Institut d’Estudis Catalans, which is officially in charge of the standardization of the language.

Note 15. Namely, the members that participated in creating the corpus were project manager Dr. Joan Torruella (ICREA-UAB), Dr. Manuel Pérez-Saldanya (UV-IEC), and Dr. Josep Martines (UA-IEC).

Note 16. Since most occurrences pertain to the 3<sup>rd</sup> person (both singular and plural), we determined that obtaining these forms would be the best choice for our exploratory analysis. Also, evidence from the literature review conducted suggests that this grammaticalization process is closely linked with narrative contexts, in which the 3<sup>rd</sup> person prevails (Pérez-Saldanya & Hualde 2003).

Note 17. Catalan was not standardized in terms of writing or formal speech until the first half of the 20<sup>th</sup> century. Original pre-standardized writing found in the CICA corpus has been observed in this paper.

Note 18. For further reference, a quantitative breakdown from the chart has been included in Appendix A.

Note 19. Transfer between directionality and purpose is likely explainable in terms of metonymy.

**Appendix A**

**Raw frequencies and percentages of *anar* + infinitive in the CICA corpus according to the three degrees of grammaticalization (i.e., grammaticalized (G), bridging context (BC), and non-grammaticalized (NG))**

|                | c. | 11 | 12 | 13    | 14    | 15    | 16    | 17    | 18   | TOTAL |
|----------------|----|----|----|-------|-------|-------|-------|-------|------|-------|
| G              | n  | 0  | 0  | 0     | 23    | 14    | 7     | 43    | 5    | 92    |
|                | %  | 0  | 0  | 0     | 14.84 | 46.67 | 38.89 | 37.07 | 62.5 | 26.59 |
| BC             | n  | 0  | 0  | 8     | 81    | 10    | 9     | 54    | 3    | 165   |
|                | %  | 0  | 0  | 42.11 | 52.26 | 33.33 | 50    | 46.55 | 37.5 | 47.69 |
| NG             | n  | 0  | 0  | 11    | 51    | 6     | 2     | 19    | 0    | 89    |
|                | %  | 0  | 0  | 57.89 | 32.90 | 20    | 11.11 | 16.38 | 0    | 25.72 |
| <b>TOTAL n</b> |    | 0  | 0  | 19    | 155   | 30    | 18    | 116   | 8    | 346   |
| <b>TOTAL %</b> |    | 0  | 0  | 5.49  | 44.80 | 8.67  | 5.20  | 33.53 | 2.31 | 100   |

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).