Grammaticalization degrees in Catalan *anar* vs. *estar* + adjective in the 19th and 20th centuries: A language contact, corpusbased distributional approach

Marc Gandarillas

University of Florida m.gandarillascot@ufl.edu



Received: 17-07-20 Accepted: 06-10-20 Published: 24-11-20

How to cite: Gandarillas, Marc. 2020. Grammaticalization degrees in Catalan *anar* vs. *estar* + adjective in the 19th and 20th centuries: A language contact, corpus-based distributional approach. *Isogloss. Open Journal of Romance Linguistics* 6, 7. DOI: https://doi.org/10.5565/rev/isogloss.93

Abstract

The present study constitutes an exploratory analysis of the use of the Catalan constructions anar ('to go') + adjective and estar ('to be') + adjective in writing during the 19th and 20th centuries. In modern Catalan, anar and estar can be used before an adjective as copulas, with an identical or similar meaning (e.g., en Joan va begut, en Joan està begut 'John is drunk'). Both constructions show evidence of being highly grammaticalized and are apparently found in free variation rather than complementary distribution. However, a closer approach seems to reveal certain underlying linguistic factors that might be informing such contrast, specifically within the semantic-pragmatic sphere. This is here explored using the narrow theory of grammaticalization (Hopper & Traugott 2003), a multistage process by which lexical (i.e., contextual) items eventually become functional (i.e., grammatical) over time. A corpus-based search was conducted using the electronic version of the Corpus Textual Informatitzat de la Llengua Catalana (CTILC), which comprises literary and non-literary texts published between 1833 and 1988. Anar + adjective and estar + adjective tokens were obtained from a sample of

texts, then sorted into four different groups corresponding to century halves. The search results from the corpus were randomly sorted in terms of genre and date and, from these, the top 250 tokens on the list of results for each construction were selected and subsequently analyzed. Whereas few instances of both constructions are found overall in the first half of the 19th century, *estar* already prevails over *anar* when followed by an adjective. During the first half of the 20th century, however, this trend starts to be reversed. This suggests the possibility that *anar* and *estar* might have traditionally been found in free variation in a number of contexts, with 19th century texts showing a clear preference for *estar*.

Keywords: grammaticalization; corpus linguistics; Romance languages; historical linguistics; language change; pragmatics.

Table of Contents

- 1. Introduction
- 2. Literature review
- 3. Research questions and hypotheses
 - 4. Methodology

- 5. Results
- 6. Conclusion and discussion References

Appendix

1. Introduction

The present study constitutes an exploratory analysis of the use of the Catalan constructions anar ('to go') + adjective and estar ('to be') + adjective in writing during the 19th and 20th centuries. In modern Catalan, anar and estar can be used before an adjective as copulas, with an identical or at least very similar meaning (e.g., en Joan va begut, en Joan està begut 'John is drunk'). Both constructions show evidence of being highly grammaticalized and are apparently found in free variation rather than complementary distribution. However, a closer approach seems to reveal certain underlying linguistic factors that might be informing such contrast, specifically within the semantic/pragmatic sphere. This is here explored using the grammaticalization theory as the overarching common thread.

Specifically, I adopt the usage-based definition of grammaticalization provided by Hopper and Traugott (2003: Ch. 1), according to which this phenomenon could be defined as the multistage process by which a lexical (i.e., context) item eventually becomes functional (i.e., grammatical) over time. According to this definition, interaction is key in promoting language change. Consequently, Hopper and Traugott (2003) approach grammaticalization from a diachronic perspective, as they focus on every potential factor underlying the conversion of the semantic substance into a-more-grammatical meaning. In such process, all changes within the language (i.e., at the semantic, functional, morphosyntactic, or phonological levels) appear to be consistently interrelated to some degree, which leads to diverse, salient outcomes such as-most commonly—generalizations of meaning (and, relatedly, semantic bleaching) and/or contexts of use, phonological reductions (either substantial or temporary), syntactic fixation, specialization (e.g., Eng. will vs. shall, Fr. pas vs. point), or an increase in frequency. By adopting this conception of grammaticalization—in which grammar is epiphenomenal; i.e., a byproduct of the circumstances

surrounding the communication¹—I am therefore keeping my distance from some studies that admit a broader scope of the phenomenon (e.g., by embracing certain instances of what other researchers have more precisely referred to as "pragmaticalization" or "discoursivization").

1.1. Research goals

Even though—being a gradual process—grammaticalization intrinsically diachronic, it has occasionally been referred to as "panchronic" in the sense that it admittedly requires to be studied from a comprehensive—both diachronic and synchronic—perspective. Taking this into account, in this study I will mostly focus on the outcomes of the grammaticalization process rather than the process itself.

As mentioned earlier, I expect this grammaticalization-based approach to assist me in paving the way for establishing relevant comparisons between the Catalan constructions anar + adjective and estar + adjective, which might in turn shed light on the degree of grammaticalization and distribution (i.e., free, complementary, or otherwise) of these forms—both synchronically and diachronically. Even though I am not tracing here the evolution of these expressions from their earliest documentation, I still expect to encounter some enlightening changes over the course of the 19th and 20th centuries.

Consequently, my exploratory analysis pursues a threefold goal: (1) determine whether or not consistent variation can be found between the Catalan expressions anar + adjective and estar + adjective; (2) if so, tackle variation in terms of relative frequency and distributional patterns; (3) based on the previous goal, establish a series of linguistic (e.g., morphosyntactic, pragmatic) and/or sociolinguistic factors that appear to be playing a role in the aforementioned distributional patterns.

2. Literature review

Grammaticalization—both in Romance languages and otherwise—has received abundant and increasing scholarly attention over the course of the last two decades (Antolí 2012, Bisang 2017, Company Company 2004, Cuenca 2001, Hopper & Traugott 2003, Martines & Montserrat i Buendia 2014, Martínez 2018, Montserrat i Buendia 2004, Norde 2019, Sentí 2012). Some clues in the field of grammaticalization in Catalan may be found in literature as early as the 1950s (Rigau 1976; Roca i Pons 1958, 1978; Romà i Roura 1987). The study of this phenomenon has traditionally focused on the gradual development of a particular lexical item (e.g., noun or verb bearing a full semantic load) into a more functional, semantically deprived one (e.g., morph expressing futurity). The process in between can be explained based on a multistage cline whose delimitation is highly contingent on the researcher's own theoretical framework,

Even though this may be hard to prove, it might not be completely inconceivable that Pompeu Fabra's standardization of Catalan (Qüestions de gramàtica catalana, 1911) had contributed to favor anar over estar in certain contexts in which the use of the latter was—either rightly or wrongfully—viewed as a result of language contact with Spanish. This might have led to a number of hypercorrected instances.

the availability of supporting evidence, or the unique characteristics of the specific process under study (Heine 2002, Montserrat i Buendia 2004, Pérez-Saldanya 1998). As a general trend, pragmatic factors (cf. metonymy, reanalysis), psychosocial characteristics of those involved (cf. subjectification; *vid.* Traugott 1995), as well as constant readjustments to new sociolinguistic demands all appear to play a significant role in both the development and the outcomes of the grammaticalization process.

As mentioned earlier (vid. 1), for the scope of this study only Hopper and Traugott's (2003) strict, narrow definition of grammaticalization is considered. However, I was able to locate a number of publications that, though differing to an extent from Hopper and Traugott's (2003) definition, provide potentially interesting theoretical and methodological insights into the matter. Among these, Alturo and Chodorowska-Pilch (2009) take a corpus-based, quantitative stance on how the Catalan original protasis si us plau (lit. if it/that pleases you) eventually becomes a discourse marker, which commonly undergoes articulatory reduction to sisplau 'please' in informal speech (however, the term grammaticalization, instead of *pragmaticalization*, is the one used to define this process). Along the same lines, and resorting to similar terminology, Cuenca and Massip (2005) investigate the cline followed by certain conjunctive phrases (in Catalan, the authors refer to these as connectors; e.g., encara que 'although') prior to reaching their current status. This whole process is usually explained in terms of syntactic reanalysis based on a need for higher degrees of subjectification or "intersubjectification" through pragmatic processes.

The study of grammaticalization in Catalan appears to have received significantly less scholarly attention than in some of the surrounding Romance languages (e.g., Spanish, French, Italian). There appears to be, however, one outstanding exception to the rule—the case of *anar* + infinitive specializing in the expression of the past perfective, which has received significant attention both at the Catalan studies level (Alturo 2017, Pérez-Saldanya & Hualde 2003, Segura 2012, Vallduví 1989) and more globally (Juge 2005). In this respect, Catalan appears to follow the path opposite to that in other Romance languages, in which the present tense of 'to go' + infinitive has eventually specialized to convey an idea of near and/or intentional future (cf. Sp. *voy a comer*, Fr. *je vais manger* 'I am going to eat' vs. Cat. *vaig menjar* 'I ate').

In order to explain the grammaticalization process of *anar* + infinitive, Alturo (2017) qualitatively compares different hypotheses that have to date been proposed within the sociolinguistics field (e.g., reanalysis, syncretism, contact with Occitan). Though acknowledging that extensive pragmatic research has been conducted on the matter, Juge (2005) finds some gaps in research that affect the morphological sphere, as well as certain inconsistencies in terms of diachrony. The author proposes that the present/preterit syncretism found in ancient Catalan in the first-person plural, *anam*—still preserved in the Balearic varieties of the language—at some point allowed for a reanalysis of the construction. This view contrasts with previous proposals that support an explanation along the lines of the so-called "narrative present" (Vallduví 1989). Pérez-Saldanya and Hualde (2003) hypothesize that the contexts in which the *anar* + infinitive construction is first attested might be key to understanding such apparently "highly anomalous" development. Based on an ad hoc literary corpus comprising narrative texts, the

authors conclude that all these contexts are actually narrative contexts in which the infinitive corresponds to a verb of accomplishment and, for this reason, there is a subsequent action expressed in the past tense.

Even though no previous research has been found on the relationship between anar + adjective and estar + adjective, the aforementioned references are undoubtedly useful to guide my methodology, enlighten the interpretation of my own results, enrich subsequent discussions, and shed light onto potential further implications and next steps related to my study. This holds especially true regarding the development of the grammaticalization process and-more specifically—its underpinnings, ranging from formal language factors (i.e., morphology, syntax) to more pragmatic or discourse-related ones (e.g., sociolinguistic, psychosocial components).

3. Research questions and hypotheses

Based on the literature reviewed, I decided to establish the following research questions, which are followed by my initial assumptions:

1. Can significant variation be found in terms of grammaticalization degrees in the use of anar + adjective vs. estar + adjective in written texts from the 19th and 20th centuries?

Based on the grammaticalization model proposed by Hopper and Traugott (2003), not only do I expect these two highly grammaticalized expressions to differ synchronically in terms of grammaticalization—even when both show evidence of high grammaticalization—but I also expect these to show signs of change—both individually and relatively to each other—along the course of the two centuries.

2. If so, how can relative frequency and distributional patterns inform the above variation?

First and foremost, I anticipate that the expressions anar + adjective and estar + adjective will show a tendency to appear—at least to an extent—in complementary distribution. Additionally, I would initially expect to encounter inversely proportional frequencies between expressions.

3. Which linguistic (e.g., morphosyntactic, pragmatic) and sociolinguistic factors could be playing an active role in establishing the aforementioned distributional patterns?

Firstly, based on the reviewed literature, a tendency of these two expressions to appear in complementary distribution might be explained based on pragmatic factors (e.g., either verb might be showing a tendency to co-occur with adjectives that are semantically similar). Secondly, if found, inversely proportional frequencies between anar + adjective and estar + adjective might be helpful to shed some light onto the matter in terms of the sociolinguistic sphere—particularly in connection with the steadily increasing contact between Catalan and Spanish, which is welldocumented for the 20th century.

4. Methodology

4.1. Materials and procedure

A corpus-based search was conducted using the electronic version of the *Corpus Textual Informatitzat de la Llengua Catalana* (CTILC²), designed by the Institute for Catalan Studies³. This corpus comprises over 52 million words that were originally written in Catalan between 1833 and 1988. The corpus contains both literary (i.e., narrative, theater, poetry, essays) and non-literary texts (e.g., cross-disciplinary treatises/manuals, specialized and informative articles, legal texts, periodical publications, pamphlets).

Anar + adjective and estar + adjective tokens were obtained from a sample of texts⁴ and then sorted into four different groups, corresponding to the first (19a) and second (19b) halves of the 19th century, as well as the first (20a) and second (20b) halves of the 20th century. The search results from the corpus were randomly sorted in terms of genre and date and, from these, the top 250 tokens on the list of results for each construction were selected and subsequently analyzed.

Tokens in which the adjective was modified by a preceding adverb were included for further analysis:

- (1) Està ja revisat. to be-3sPI already to revise-PPms 'It is already revised.' (Pompeu Fabra, *Cartes a Joaquim Casas-Carbó*, 1910)
- (2) Anava molt més rerassagada. to go-3sII much more to lag behind-PPfs 'She was way behind.' (Josep Maria Poblet, *Aribau i abans i després*, 1963)

This equally applied to all adverb types, including the emphatic negative particle *pas*:

(3) No estarà pas enllestida a temps. not-AUX to be-3sF no way to get ready-PPfs in time 'There is no way she will be ready in time.' (Pompeu Fabra, *Cartes a Joaquim Casas-Carbó i a J. Massó i Torrents*, 1911)

Also, some (rare) instances were found in which a subject pronoun is inserted between the verb and the adjective:

_

The corpus can be accessed following the link https://ctlc.iec.cat/.

In Catalan, and officially, *Institut d'Estudis Catalans*, which is the official institution in charge of the standardization of the language.

For the specific texts from which the samples were obtained, please refer to the breakdown table included in Appendix 1.

(4) Estich jo malaltís. to be-1sPI I unhealthy-ms 'I am feeling unwell.' (Josep Yxart, Cartes a Joan Sardà, 1893)

Past participles preceded by anar or estar were considered functional adjectives and, consequently, also included for further analysis⁵:

- (5) [...] que vaigi acompanyat d' altres aliments necessaris. that to go-3sPS to accompany-PPms of other-mp food-mp necessary-mp '[...] that it is accompanied by other necessary foods.' (Andorra 7. Setmanari del Principat, 1978)
- (6) que estava anunciat oficialment and that to be-3sII to announce-PPms officially 'and that it was officially announced' (Josep Maria Llorens i Ventura, Cartes a Amadeu Bernardó, 1954)

On the other hand, exclusions mainly comprised instances in which the adjective appeared prior to the verb:

(7) donat 10 delicat que estic to give-PPms the-ns delicate-ms that to be-3Spi 'given how weak I am feeling' (Josep Yxart, Cartes a Joan Sardà, 1894)

Present participles were also excluded from the list of potential adjectives:

(8) teló lo arriantse magestuosament va the-ms curtain-ms to go-3sPI to lower-GER majestically 'the curtain is being lowered majestically' (Fèlix Socias i Urgellés, ¿Ja hi han entrat!, 1860)

4.2. Data management and analysis

Firstly, in order to comparatively track the evolution of anar + adjective and estar + adjective (and to be able to interpret this further in the light of pragmatic and sociolinguistic factors), raw frequencies for all tokens (N=500, 250 per construction) were obtained and sorted into one of the four time-period groups (i.e., 19a, 19b, 20a, 20b) along with their respective percentages. Secondly, every token for either construction was classified according to its particular degree of grammaticalization. Specifically, three degrees were established—nongrammaticalized (NG), semi-grammaticalized (SG)⁶, and grammaticalized (G).

In Catalan, as is the case in other world languages, it is commonplace to find instances of adjectives that originated in a past participle (e.g., estic esgotat 'I am exhausted'; la noia està cansada 'the girl is tired'; vas vestit tot elegant 'you are all dressed up').

The semi-grammaticalized label here is consistent with Heine's (2002) concept of "bridging context," since it appears to account for the transition from

In the case of *anar* ('to go'), tokens classified as non-grammaticalized were those in which verb internal arguments were other than the following adjectives:

(9) anava sola a abeurar sos ulls to go-3sII alone-fs to to soak-INF POSS-mp eye-mp 'she would go [somewhere] all by herself to soak her eyes' (Llorenç Riber i Campins, *Al sol alt*, 1931)

Semi-grammaticalized tokens were those in which, although the verb internal argument indicating the destination was not explicit (and, probably, also completely irrelevant), the idea of directionality and/or movement had been nonetheless preserved. Here, several different patterns were encountered, such as anar vestit/-ida, anar nu/-a, anar coix/-a, anar junts, anar descalç/-a (respectively and literally, 'to go dressed/naked/limping/together/barefoot'):

(10) Els pagesos, quan van descalços, la temen molt. the-mp peasant-mp when-CONJ to go-3pPI barefoot-mp ACC-3sf to be afraid-3pPI a lot 'When they go barefoot, peasants are very afraid of it.'
(Josep Calicó, *Apunts de la flora medicinal de Catalunya*, 1921)

Finally, tokens classified as grammaticalized were those in which the notion of movement originally conveyed by *anar* was no longer a possible interpretation. In fact, in multiple instances of *anar* + adjective the association between the verb and the adjective is strong (both in terms of frequency and semantic bleaching). Additionally, the semantic characteristics of the adjective appear to prevent *anar* from deploying its original notion of movement (e.g., *anar errat/-ada* 'to be (lit. go) wrong,' *anar ple/-ena* 'to be (lit. go) full'):

(11) Però, si no vaig molt errat, hom aconsella sovint [...] but if not-AUX to go-1sPI very wrong-ms someone to advise-3sPI often 'But, unless I am too wrong, it is oftentimes advised [...]' (Joan F. Mira i Casterà, *A la recerca de la història cultural*, 1974)

On the other hand, for *estar* + adjective no instances were found of non-grammaticalization (i.e., the original meaning of the verb, 'to stand,' from the Latin etymon STARE⁷). However, if we consider the original meaning of the verb,

grammaticalized contexts to non-grammaticalized ones (vid. Heine, Bernd 2002: 83-101).

In the first entry of the lemma *estar*, the *Diccionari català-valencià-balear* (a purely descriptive lexicographical work accounting for most varieties of the language) provides the following information: "*Trobar-se dempeus i aturat (per oposició a anar o caminar*)" ['To stand still on one's feet (as opposed to going or walking)']. The example provided dates back to the 13th century (Ramon Llull, *Cont.*, 315, 22: "S'esforsa... com pusca esser vertuós en son parlar e en son callar e en son anar e en son estar" ['He strives to be virtuous not only when he speaks, but also when he is silent, when he walks, and when he stands']). This could be suggesting that *estar* completed its own grammaticalization process well before *anar*, which would in turn

occurrences appear to be considerably split between semi-grammaticalized and non-grammaticalized instances. The former group includes multiple occurrences of estar + past participle that seem to approach the notion of the so-called middle voice (or even, in some cases, the passive voice itself), in which the 'verb + participle' bigram appears to have undergone significant semantic bleaching and increased its strength in terms of mutual association (most likely through increased frequency):

- (12)Les botigues de segells de Madrid estan clausurades policia. per la the-fp store-fp of stamp-mp of Madrid to be-3pPI to close down-PPfp by the-fs police-fs 'The stamp stores in Madrid have been closed down by police.' (Enric Puigferrat, Carta a Ramon d'Areny-Plandolit (16/IV/1956), 1956)
- (13)de venir Vaig estar temptat -VOS a veure. to go-1sPI to be-INF to tempt-PPms of to come-INF ACC-2p to see-INF 'I was tempted to come see you.' (Joan Sales, Carta a Joan Oller i Rabassa (27/XII/1957), 1957)

Other highly-frequent grammaticalized constructions from the corpus are estar ben vist ('to be well thought of'), estar content/-a de ('to be happy/glad to'), estar decidit/-ida a ('to be determined to'), estar pròxim/-a a ('to be about to'), all followed by an infinitive.⁸ Last but not least, distributional patterns were searched for every construction in order to support or reject the complementary distribution hypothesis.

5. Results

In connection with my first and second research questions, Table 1 below provides an overview of how constructions anar + adjective and estar + adjective quantitatively evolve over the course of the 19th and 20th centuries in written language. Whereas few instances of both constructions are found overall in the first half of the 19th century⁹, estar already prevails over anar when followed by an adjective. Such prevalence continues well into the second half of the 19th century—in which estar accounts for 70.18% of the tokens from both constructions. However, data from the first half of the 20th century shows that this

explain why no instances of non-grammaticalization were found in my own selection of the CTILC corpus for estar + adjective constructions during the 19th or 20th centuries.

In this respect, it might be relevant to emphasize that the earliest materials found in the CTILC corpus date back to 1833 (not 1800). Along the same lines, the most recent texts within the corpus belong to the year 1988 (not 1999).

However, it must be noted that the differentiation between grammaticalized and semi-grammaticalized is usually contingent, not on the 'estar + adjective' bigram, but on pragmatic and/or contextual factors. For instance, estar content/-a in the literal sense was considered as semi-grammaticalized, whereas estar content/-a de + infinitive was classified as grammaticalized (cf. Eng. I am happy (because I won the lottery) vs. I am happy to inform you that you have been selected).

trend starts to be reversed at this point. This becomes especially noticeable during the second half of the 20^{th} century, with estar + adjective constructions only accounting for 37.50% of the total number of tokens.

The above-presented data suggests the possibility that *anar* and *estar* might have traditionally been found in free variation (at least in certain contexts), with 19th century texts showing a clear preference for *estar*. The figures for *anar* + adjective and *estar* + adjective eventually become virtually even—still with a slight precedence of *estar*—during the first half of the 19th century. Finally, during the second half of the 20th century the initial trend is reversed, with *anar* + adjective bigrams accounting for 62.50% of the total number of tokens. With Pompeu Fabra—considered the father of Catalan standardization—having published his most influential works by mid-20th century (relevantly, the *Diccionari General de la Llengua Catalana*, 1932), the increase in the use of *anar* + adjective in written Catalan might—arguably—have presented some sort of connection with the clandestine standardization efforts in a time in which the language was being banned from the public sphere by the Francoist dictatorship¹⁰.

Table 1. Raw frequencies, relative frequencies, and percentages of *anar* + adjective and *estar* + adjective in the CTILC-based sample (1833-1988)

	19a (1833-1850)		19b (1851-1899)		20a (1900-1950)		20b (1951-1988)		
	anar	estar	anar	estar	anar	estar	anar	estar	TOTAL
raw/ rel. fr.	6 / 1.2	12 / 2.4	17 / 3.4	40 / 8	127 / 25.4	138 / 27.6	100 / 20	60 / 12	500
%	33.33	66.67	29.82	70.18	47.92	52.08	62.50	37.50	100
TOTAL raw/rel. fr.	18 / 3.6		57 / 11.4		265 / 53		160 / 32		
TOTAL %	3.60		11.40		53		32		

Source: Corpus Textual Informatitzat de la Llengua Catalana (CTILC)

On the other hand—as earlier indicated and exemplified in Section 4.2 ('Data management and analysis')—every token of anar + adjective and estar + adjective was annotated according to the degree of grammaticalization shown within its own context. This information was summarized in a table for further analysis and comparison. Table 2 below contains both the raw frequencies and percentages accounting for the tokens of anar/estar + adjective in each time period, considering their respective degrees of grammaticalization.

Without the possibility of readily-available access to materials published in Catalan (e.g., literary works from previous centuries), some purist currents at the time would hold the belief that the most correct choices in terms of linguistic structures or lexical items were those most differing from their Spanish counterparts. According to some sources, Pompeu Fabra earned the reputation of being a purist, even in the words of some of his colleagues at the Institute for Catalan Studies (Gracia Alonso, F. (2011), *Pere Bosch Gimpera: Universidad, política, exilio.* Madrid: Marcial Pons Historia, 146).

Table 2. Raw frequencies and percentages of *anar* + adjective and *estar* + adjective from the corpus-based sample (1833-1988) according to the three degrees of grammaticalization (i.e., grammaticalized (G), semi-grammaticalized (SG), and non-grammaticalized (NG)).

		19a (1833-1850)		19b (1851-1899)		20a (1900-1950)		20b (1951-1988)		
		anar	estar	anar	estar	anar	estar	anar	estar	TOTAL
7 h	f	6	7	15	14	85	48	77	30	282
G	%	2.13	2.48	5.32	4.96	30.14	17.02	27.30	10.64	100
כי	f	0	5	2	26	33	90	17	30	203
SG	%	0	2.46	0.99	12.81	16.26	44.33	8.37	14.78	100
_C	f	0	0	0	0	9	0	6	0	15
NG	%	0	0	0	0	60	0	40	0	100
TOT f	AL	6	12	17	40	127	138	100	60	500
TOT		33.33	66.67	29.83	70.18	47.93	52.08	62.5	37.5	

Source: Corpus Textual Informatitzat de la Llengua Catalana (CTILC)

Table 2 shows that, by the early 1800s, both *anar* + adjective and *estar* + adjective already appear to be highly grammaticalized. In fact, for the latter construction, no instances of non-grammaticalization are found, which indicates that the grammaticalization process had reached full completion well before the beginning of the 1800s¹¹. Non-grammaticalized tokens of *anar* (i.e., with the verb indicating movement toward an established destination) are actually widespread in Catalan. However, when non-grammaticalization is present, the *anar* + adjective structure is possible, yet not as common, since the internal argument of the verb is forced to appear after the adjective, not immediately following the verb (e.g., *en Joan va content a la feina* 'John goes happily (lit. *happy*) to work'). On the other hand, tokens of semi-grammaticalized *estar* + adjective—in which *estar* basically represents a transient condition or state—remain stable through all the time periods studied, by far outnumbering their *anar* counterparts.

Now focusing on the grammaticalized occurrences of *anar* vs. *estar* + adjective, these appear to be virtually and consistently even throughout the 19th century, whereas grammaticalization instances of *anar* + adjective considerably outnumber their *estar* + adjective counterparts throughout the 20th century. This might align well with the 1900s trend of switching to *anar* in constructions (usually fixed expressions) that might traditionally have resorted to *estar* (e.g., *anar cansat/-ada* 'to be (lit. *go*) tired'). As mentioned earlier, I believe that standardization efforts of the language within an adverse sociopolitical and sociolinguistic context should not be excluded as a possible factor. Additionally, in the instances analyzed of *anar* + adjective fewer examples of idiomatic expressions are found (e.g., *anar errat/-ada* 'to be (lit. *go*) wrong' (10 tokens)). Concurrently, these appear to be more frequent than in the case of *estar* +

Like in Spanish, the original use of *estar* 'to stand' (< Lat. STARE) has not been preserved in Catalan.

adjective, in which variety consistently overcomes frequency (e.g., estar cansat/-ada 'to be tired,' estar obligat/-ada 'to be obliged,' estar content/-a 'to be happy,' estar decidit/-ida 'to be determined,' estar segur/-a 'to be sure,' estar disposat/-ada 'to be willing'). This can be explained based on the fact that estar, unlike anar, shows evidence of having fully completed its grammaticalization process (or, at least, multiple uses appear to have been grammaticalized for considerably longer periods of time).

Finally, the answer to my third research question seems to involve greater complexity. On the one hand, given the fact that in many contexts the two constructions under study seem to appear in free variation (excluding idiomatic expressions such as the ones I mentioned earlier), there are necessarily certain factors—likely, language standardization efforts—that can explain the precedence of *anar* throughout the 1900s. This would also relate to the fact that *anar* emerges as being used in more grammaticalized contexts in the 1900s.

6. Conclusion and discussion

The results obtained are certainly interesting in the light of the theoretical framework proposed by Hopper and Traugott (2003), as they undoubtedly appear to point to certain directions for prospective research. Still, more research is needed to refine and validate my results. For instance, it could be enlightening to conduct a follow-up study based on the contrastive analysis of anar vs. estar + adjective across both literary and non-literary genres (also with an emphasis in oral speech; vid. Romà i Roura 1987). Probably, a step in the right direction would be to explore how this is reflected in letters, which—despite obvious reservations—probably constitute the genre closest to popular, spontaneous speech. Should language standardization efforts have actually played an important role in the expansion of anar + adjective contexts, data from the CTILC corpus could be compared to spontaneous speech samples prior to venturing hurried assumptions on the matter. Another possibility consists in delving into the semilinking (Cat. semicopulatiu) characteristics of anar and estar, as well as the divergences shown by these verbs to this respect (Bosque & Demonte 1999, RAE & ASALE 2009-2011, Solà 2002). A strict differentiation (whenever possible) of the nature of intervening factors (i.e., sociolinguistic vs. discourse-pragmatic factors) might also lead to supplementary insights into the matter. As mentioned earlier, a factor that is not to be completely dismissed lies in the influence of Pompeu Fabra's standardization of the language, which might have been a potential contributor to accelerating the grammaticalization process of anar in those instances in which it was consistently preferred over estar (based on hypercorrection or otherwise).

Another interesting clue is the frequency vs. variety contrast that is clearly observed between grammaticalized samples of anar + adjective and estar + adjective, with the former being more frequent in 20^{th} century writing, yet not presenting as many possible combinations as the latter. It would certainly be relevant to further explore these results in the light of the grammaticalization theory, since the fact that estar + adjective has been grammaticalized for longer seems to have allowed for a larger variety of combinations. In order to shed more

light onto the matter and help develop further perspectives, it is my belief that it could be really enlightening to trace the grammaticalization of *estar* and *anar* back to the earliest documentations of the language.

References

- Alturo, Núria. 2011. "L'auxiliar de passat 'anar' en català". *Apunts de pragmàtica del català. Teories de la gramaticalització: Cas 1* (p. 2-9). Barcelona: Dipòsit Digital de la Universitat de Barcelona.
- Alturo, Núria & Chodorowska-Pilch, Marianna. 2009. "La gramaticalització de *sisplau*". *Els Marges* 88: 15-38.
- Antolí, Jordi. 2012. "Canvi semàntic i gramaticalització en el sorgiment de marcadors evidencials: Evolució semàntica de PARERE i derivats en el llatí tardà i en el català antic (s. III-XVI)". *eHumanista/IVITRA* 2: 41-84.
- Bisang, Walter. 2017. *Grammaticalization*. Oxford Research Encyclopedia of Linguistics. https://doi.org/10.1093/acrefore/9780199384655.013.103
- Bosque, Ignacio & Demonte, Violeta. 1999. *Gramática descriptiva de la lengua española*. Madrid: Espasa-Calpe.
- Company Company, Concepción. 2004. "Gramaticalización por subjetivación como prescindibilidad de la sintaxis". *Nueva Revista de Filología Hispánica* 52(1): 1-27. https://doi.org/10.24201/nrfh.v52i1.2226
- Cuenca, Maria Josep. 2001. "La gramaticalització com a teoria de la variació morfosintàctica". In M. Lorente, N. Alturo, E. Boix, M.-R. Lloret & L. Payrató (eds.). *La gramàtica i la semàntica en l'estudi de la variació*, 151-185. Barcelona: PPU.
- Cuenca, Maria Josep & Massip, Àngels. 2005. "Connectors i processos de gramaticalització". *Caplletra* 38: 77-92.
- Heine, Bernd. 2002. "On the role of context in grammaticalization". In I. Wischer & G. Diewald (eds.), *New Reflections on Grammaticalization*, 696-707. Oxford: Oxford University Press. https://doi.org/10.1075/tsl.49.08hei
- Hopper, Paul & Traugott, Elizabeth Closs. 2003. *Grammaticalization*. Oxford: Oxford University Press. https://doi.org/10.1017/cbo9781139165525
- Institut d'Estudis Catalans [IEC]. 2016. *Nova gramàtica de la llengua catalana*. Barcelona: Institut d'Estudis Catalans.
- Solà i Cortassa, Joan (ed.). 2002. *Gramàtica del català contemporani*. Barcelona: Empúries.
- Juge, Matthew L. 2006. "Morphological factors in the grammaticalization of the Catalan 'go' past". *Diachronica*, 23(2): 313-339. https://doi.org/10.1075/dia.23.2.05jug
- Martines, Josep & Montserrat i Buendia, Sandra. 2014. *Subjectivació i interferència en l'evolució semàntica i en l'inici de la gramaticalització de "jaquir" (segles XI-XI)*. Barcelona: Publicacions de l'Abadia de Montserrat.
- Martínez, Caterina. 2018. "La gramaticalització de connectors de contrast: Estudi de corpus i aproximació segons la Teoria de la Inferència Invitada del Canvi Semàntic" (doctoral dissertation).

Montserrat i Buendia, Sandra. 2004. "Entre la dixi i la definitud: Els verbs de moviment resultatiu en català". *Caplletra* 39: 61-83.

- Norde, Muriel. 2019. "Grammaticalization in morphology". In Lieber, R. (ed.), *Oxford Research Encyclopedia of Linguistics* (vol. Morphology). Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.733
- Pérez-Saldanya, Manuel. 1998. *Del llatí al català: Morfosintaxi verbal històrica*. (Collecció Biblioteca Lingüística Catalana). Valencia: Universitat de València.
- Pérez-Saldanya, Manuel & Hualde, José Ignacio. 2003. "On the origin and the evolution of the Catalan periphrastic preterit". *Verbalperiphrasen in den* (*ibero*)romanischen Sprachen, 46-70. Hamburg: Buske.
- Real Academia Española [RAE] & Asociación de Academias de la Lengua Española [ASALE]. 2009-2011. *Nueva gramática de la lengua española. Morfología y sintaxis*. Madrid: Espasa.
- Rigau, Gemma. 1976. "Anem o venim?". Els Marges 8: 33-53.
- Roca i Pons, Josep. 1958. Verbs auxiliars afins a *estar* en català antic. *Estudis Romànics* 6: 165-168.
- Romà i Roura, Montserrat (1987). "L'argot dels joves gironins". Revista de Girona 120: 82-93.
- Segura, Carles. 2012. "El passat perifràstic en català antic: Una revisió a partir d'estudi de corpus". *eHumanista/IVITRA* 2: 118-147.
- Sentí, Andreu. 2012. "Gramaticalització i subjectivització del verb modal *haver* (a/de) en català antic: Un estudi de corpus segons la gramàtica cognitiva". eHumanista/VITRA 2: 85-117.
- Traugott, Elizabeth Closs. 1995. "Subjectification in grammaticalization". In D. Stein & S. Wright (eds.), *Subjectivity and Subjectivisation: Linguistic Perspectives*, 31-54. Cambridge: Cambridge University Press. https://doi.org/10.1515/9783110226102.1.29
- Vallduví, Enric. 1989. "Sobre el perfet perifràstic del català". *Actes del cinquè col·loqui d'estudis catalans a Nord-Amèrica*, 85-98. Barcelona: Publicacions de l'Abadia de Montserrat.

APPENDIX. Reference texts used from the Corpus Textual Informatitzat de la Llengua Catalana (CTILC)

time	year	author	work
frame		****	
		Alsina, Anton	Documents notarials, I. Barcelona
		Roca, F	Geografia
		Lagrifa i de Bono, Ramon	Documents notarials, I. Girona
		Bota, Josep	Documents notarials, I. Cassà de la Selva
19a		Gual, Esteve	Documents notarials, II. La Seu d'Urgell
		Tàpies, Ramon	Carta a Guillem d'Areny (11/X/1840)
		Alier i Romeu, Llorenç	Carta a Guillem de Plandolit i d'Areny
		Tomàs i Cortada, Ramon	Documents notarials, I. La Bisbal
		Gurri, Lluís	Documents notarials, I. Barcelona
		Prim, Josep	Documents notarials, I. Granadella
	1854	Josep, Andreu	Documents notarials, V. Barcelona
	1055	Cala Manià Issan	Documents notarials, III. Sant Feliu de
		Sala, Marià Josep	Guíxols ;Ja hi han entrat!
		Socias i Urgellés, Fèlix	
		Balaguer, Víctor Picó i Campamar, Ramon	Apuntaments y datos Cartes a Tomàs Forteza (1874)
		Pella i Forgas, Josep	Cartes a Josep Serra i Campdelacreu (1876)
	18/0	Associació Catalanista d'Excursions	Álbum pintoresch-monumental de
	1881	Científiques	Catalunya: Montserrat
		Sanpere i Miquel, Salvador	Aplicació del art á l'industria
		Costa i Llobera, Miquel	Carta a Francesc Matheu (15/VI/1882)
		Torres i Bages, Josep Maria	¿Qué es la masonería?
		Suñol, Esteve	Carta a Francesc Matheu (3/V/1884)
		Alcover, Antoni Maria	Carta a Marià Aguiló (26/VIII/1886)
19b		Soler, Frederic	Carta a Àngel Guimerà (8/III/1888)
170		Alcover, Antoni Maria	Carta a Marià Aguiló (24/VIII/1890)
		Bosch de la Trinxeria, Carles	Cartes a Josep Yxart (1890)
		Carreras i Candi, Francesc	Argentona històrica
	1892	, , ,	Carta a Joaquim Areny de Plandolit
		, ,	Carta a Joaquim Areny de Plandolit
	1892	Jordana, A.	(1/III/1892)
	1893	Yxart, Josep	Cartes a Joan Sardà (1893)
	1894	Yxart, Josep	Cartes a Joan Sardà (1894)
	1895	Yxart, Josep	Carta a Joan Sardà (31/I/1895)
	1895	Yxart, Josep	Carta a Joan Sardà
			Análisis fonològich-ortográfich de la llènga
		Nonell i Mas, Jaume	catalana antiga y modèrna
	1897	Roure, Conrad	;13!
	1899	Vayreda, Marià	Carta a Francesc Matheu
	1901	,	Cartes a l'arquebisbe de Tarragona (1901)
20a	1901	/ 3	Cartes a Joaquim Areny de Plandolit (1901)
	1901		¡Novicis!
	1902	,	Clichés
	1903	Aulet, Maria del Pilar	Carta a Francesc Matheu (1/VI/1903)
	1002		Aplech de sentencies y pensaments de
		Codinach i Espinalt, Joan	filosophs insignes
		Folch i Torres, Josep Maria	Antonia
		Cavallé, Pere	Aubada i posta
	1904	, 8	Cartes a Amadeu Vives (1904)
	1904	, i	Carta a Francesc Matheu (7/XII/1904)
	1905	Guimerà, Àngel	Andrónica

	Puig i Ferreter, Joan	Arrels mortes
	Maspons i Camarasa, Jaume	Agremiació agrícola
	Obrador Bennàssar, Mateu	Carta a Ernest Moliné i Brasés
	Romaní i Guerra, Amador	Cartes a Norbert Font i Sagué (1910)
	Fabra, Pompeu	Cartes a Joaquim Casas-Carbó (1910)
1910	Vinyes i Cluet, Ramon	Al florir dels pomers
1011	.	Cartes a Joaquim Casas-Carbó i a J. Massó
	Fabra, Pompeu	Torrents (1911)
1912	Barnils, Pere	Cartes a Antoni Maria Alcover (1912)
1010	T	Carta a Jaume Massó i Torrents
1912	Fabra, Pompeu	(20/IV/1912)
	Vogel, Eberhard	Carta a Francesc Matheu (3/XI/1912)
	Carner, Josep	Auques i ventalls
	Vidal i Lluch, Antoni	Carta a Francesc Matheu (21/II/1914)
	Barnils, Pere	Cartes a Antoni Maria Alcover (1915)
	Rubió i Lluch, Anton	Carta a Francesc Matheu (26/V/1915)
	Costa i Llobera, Miquel	Carta a Francesc Matheu (29/XII/1916)
	Vallès, Emili	Apologia de la llengua catalana
	Vallès, Emili	Aritmètica mercantil
	Oller, Narcís	Al llapis y a la ploma
	Riba, Carles	Cartes a Josep Obiols (1920)
	Alcover, Joan	Carta a Francesc Matheu (8/IV/1920)
	Riba, Carles	Cartes a Josep Obiols (1921)
	Riba, Carles	Cartes a Josep Obiols
	Collell, Jaume	Carta a Francesc Matheu (18/VIII/1921)
	Calicó, Josep	Apunts de la flora medicinal de Catalunya
	Carner, Josep	Cartes a Jaume Bofill i Mates (1922)
	Mir, Josep	Carta a l'Honorable Comú (13/II/1922)
1922	Bofill i Mates, Jaume	Carta a Josep Carner (28/X/1922)
	Federació Catalana de Natació Amateur	Anuari 1923
	Fontserè, Eduard	Atlas elemental de núvols
1926	Balanzó i Pons, Llorenç	Camí
1927	Martínez i Martínez, Francesc	Carta a Joan Givanel i Mas (27/XI/1927)
1928	Nadal i Mallol, Hipòlit	Articles de contraban: 1923-1927
1929	Riber i Campins, Llorenç	Any cristià, I
1930	Millàs-Raurell, Josep Maria	Anna Christie
1931	Riber i Campins, Llorenç	Al sol alt
1932	Albert Paradís, Caterina	Carta a Tomàs Roig i Llop (1/VI/1932)
1933	Ruyra, Joaquim	Assaig crític sobre el teatre de Molière
		Actuació de l'arquitecte als pobles i petites
1933	Martinell, Cèsar	viles
	Albert Paradís, Caterina	Carta a Tomàs Roig i Llop (15/I/1934)
1024	Dublicacione Deviddiesses	Alt Pallars. Quinzenal portantveu del Centr
	Publicacions Periòdiques	Republicà Català, \$\$18, Sort, 1934.
	Albert Paradís, Caterina	Cartes a Tomàs Roig i Llop (1935)
	Oliver, Joan	Allò que tal vegada s'esdevingué
	Llovet Mont-Ros, Josep	Apunts d'economia agrícola
	Pous i Pagès, Josep	Al marge de la revolució i de la guerra
	Albert Paradís, Caterina	Carta a Tomàs Roig i Llop (XII/1938)
	Thió de Gibert, Maria de	Carta a Núria Matheu (17/XII/1938)
	Amades, Joan	Apunts d'imatgeria
	Rodoreda, Mercè	Aloma
	· · · · · · · · · · · · · · · · · · ·	Carta a Montserrat i Núria Matheu
	Camarasa, Rafaela	(1/I/1939)
1939		Cartes a Joan Comorera (1946)
	Tarradelles, Josep	
1946	Tarradelles, Josep Albert Paradís, Caterina	
1946	Tarradelles, Josep Albert Paradís, Caterina	Carta a Tomàs Roig i Llop (10/III/1946)
1946 1946	Albert Paradís, Caterina	Carta a Tomàs Roig i Llop (10/III/1946) Carta a Ramon d'Areny-Plandolit
1946 1946 1946	Albert Paradís, Caterina Grifol i Foix, Andreu	Carta a Tomàs Roig i Llop (10/III/1946) Carta a Ramon d'Areny-Plandolit (16/11/1946)
1946 1946 1946 1946	Albert Paradís, Caterina Grifol i Foix, Andreu Sala i Jornet, Carles	Carta a Tomàs Roig i Llop (10/III/1946) Carta a Ramon d'Areny-Plandolit (16/11/1946) Acte de fe
1946 1946 1946 1946 1946	Albert Paradís, Caterina Grifol i Foix, Andreu	Carta a Tomàs Roig i Llop (10/III/1946) Carta a Ramon d'Areny-Plandolit (16/11/1946)

	1947	Amades, Joan	Apunts d'imatgeria		
		Massip, Evarist	Cartes a Pau Rovira (1948)		
		Hurtado i Miró, Amadeu	Carta a Antoni Rovira i Virgili (9/I/1948)		
			Carta a Antoni Rovira i Virgili		
		Hurtado i Miró, Amadeu	(26/VII/1948)		
		Hurtado i Miró, Amadeu	Carta a Antoni Rovira i Virgili (25/XI/1947)		
		Hurtado i Miró, Amadeu	Carta a Antoni Rovira i Virgili (18/III/1948)		
		Hurtado i Miró, Amadeu	Carta a Antoni Rovira i Virgili (1/XI/1948)		
		Rovira i Virgili, Antoni	Carta a Amadeu Hurtado (7/VI/1948)		
		Rovira i Virgili, Antoni Rovira i Virgili, Antoni	Carta a Amadeu Hurtado (3/V/1948) Carta a Amadeu Hurtado (2/II/1948)		
		Rovira i Virgili, Antoni	Carta a Amadeu Hurtado (2/11/1948) Carta a Amadeu Hurtado (11/VIII/1948)		
		Rovira i Virgili, Antoni	Carta a Amadeu Hurtado (11/ VIII/1948)		
		Hurtado i Miró, Amadeu	Carta a Antoni Rovira i Virgili (8/II/1949)		
	1747	Turudo i miro, rimadea	Carta a Antoni Rovira i Virgili		
	1949	Hurtado i Miró, Amadeu	(31/VIII/1949		
		Rovira i Virgili, Antoni	Carta a Amadeu Hurtado (26/III/1949)		
		Rovira i Virgili, Antoni	Carta a Amadeu Hurtado (2/IX/1949)		
		Granier-Barrera, Emili	Cartes (1950)		
		Grifol i Foix, Andreu	Carta a Ramon d'Areny-Plandolit (8/II/1950)		
			"Els orígens del coneixement" de Ramon		
	1950	Rosés i Lacoigne, Manuel	Turró		
	1951	Calvet, Agustí	Carta a Joan Oller i Rabassa (27/XII/1951)		
	1951	Albert Paradís, Caterina	Carta a Tomàs Roig i Llop (VI/1951)		
	1951	Grifol i Foix, Andreu	Cartes a Ramon d'Areny-Plandolit (1951)		
		Saltor, Octavi	Carta a Joan Oller i Rabassa (9/VII/1952)		
	1952	Grifol i Foix, Andreu	Cartes a Ramon d'Areny-Plandolit (1952)		
		Janer i Vinyes, Joan	Carta a Joan Oller i Rabassa (26/VI/1953)		
		Ferran de Pol, Lluís	M'he quedat a 'Abans de l'alba'		
		Llorens i Ventura, Josep Maria	Cartes a Amadeu Bernadó (1954)		
	1956	Soldevila, Ferran	Albert i Francina		
	1956		"Lo Verdader catalá". Primer òrgan		
		Casacuberta, Josep Maria de	periodístic de la Renaixença (1843)		
	1056		Carta a Ramon d'Areny-Plandolit		
		Puigferrat, Enric	(16/IV/1956)		
		Foix, Pere Sales, Joan	Apòstols i mercaders Carta a Joan Oller i Rabassa (27/XII/1957)		
		Camps i Reverter, Guiu Maria	Apocalipsi		
	1936	Camps i Keverter, Guiu Maria	Amics de muntanya: excursionisme i plantes		
	1961	Mascarell i Gosp, Josep	medicinals		
		Llorens i Ventura, Josep Maria	Cartes a Amadeu Bernadó (1962)		
		Poblet, Josep Maria	Aribau i abans i després		
20b		Fontana i Lázaro, Josep	Aribau i la indústria cotonera a Catalunya		
		Villalonga, Llorenç	Aquil·les o l'impossible		
		Berenguer Amenós, Jaume	Alexis Zorbàs		
		Llorens i Ventura, Josep Maria	Cartes a Amadeu Bernadó (1966)		
		Muñoz i Pujol, Josep Maria	Antígona		
	1969		Cartes a Amadeu Bernadó (1969)		
	1971	Poblet, Josep Maria	Aquell any 1917		
	1972	Llorens i Terol, J.	Antibiòtics en pediatria		
	1974	Mira i Casterà, Joan Francesc	A la recerca de la història cultural		
	1974	Riera Llorca, Vicenç	Cartes a Amadeu Bernadó (1974)		
			¿Has vist, algun cop, Jordi Bonet, Ca		
	1976	Bonet, Blai	N'Amat a l'ombra?		
			Arreu. Setmanari d'informació general de		
		Publicacions Periòdiques	Catalunya, \$\$7, Barcelona, 1976.		
	1978	Poblet, Josep Maria	1938, any de la davallada		
	1050	n., . n., .	Andorra 7. Setmanari del Principat, \$\$1,		
		Publicacions Periòdiques	Andorra, 1978.		
		Pedrolo, Manuel de	A las 20 haves field at		
		Carandell, Josep Maria	A les 20 hores, futbol		
	1980	Borrajo, Josep	Cartes a A. Planes (1980)		

		Alhora. Fulls de Gandesa, \$\$3, Gandesa,
1982	Publicacions Periòdiques	1982.
1982	Iglésies i Fort, Josep	Amb les cames i amb el cor
1983	Gaja i Molist, Esteve	Administració i gestió municipals
1983	Benguerel, Xavier	Apassionata
1983	Carner-Ribalta, Josep	Cartes a Emili Granier-Barrera (1983)
1985	Bastardas i Parera, Joan	Els camins del mar
	Xirau Vayreda, Maria; Vidal i Carou,	
1985	María del Carmen	Anàlisi química
1985	Arcarons, Melcior; Arqué, Maite	Aprofitament dels recursos marins
1985	Campillo, Maria; Gustà, Marina	"Mirall trencat" de Mercè Rodoreda