The contribution of aspectual auxiliary verbs to the factual value of verb periphrases in Spanish: An empirical study

Ana Fernández Montraveta[a] – Glòria Vázquez[b] – Hortènsia Curell[a]
Universitat Autònoma de Barcelona[a] / Spain
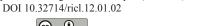University of Lleida[b] / Spain

Abstract – This paper presents the results of a corpus-based study on the contribution of Spanish aspectual auxiliary verbs to the factual interpretation of the predicates they modify and with which they constitute the verb complex (a verb periphrasis, VP). The study was carried out as part of a project that, based on linguistic knowledge, aimed at automatizing the annotation of the factual status of texts in this language. We analyzed 674 sentences in European Spanish extracted from four corpora, where 28 VPs were represented, considering only indicative tenses. Interestingly, the results show that, although one of the most important aspects that affects the factual interpretation in sentences with VPs is the verb tense of the auxiliary, in a few VPs, the type of auxiliary changes the factual value expected for the VP taking into account the tense used. Finally, based on the data under analysis, the study concludes that it is feasible to state general rules to automate the annotation of factuality for most of the aspectual VPs studied.

Keywords – aspectual verb periphrases; factuality; tense; annotation of corpora; Spanish

1. INTRODUCTION

Each speaker narrates events from their own perspective. In any written sentence, the author's commitment to the truth of what is being said is stated. Event factuality is then defined as the way in which an event is presented in relation to certainty (Narita *et al.* 2013). That is, factuality is not directly connected to the truth value of a fact with respect to the world but to the attitude of the speaker (their stance) towards its truth value (Saurí 2008; Wonsever *et al.* 2016). In this way, factuality relates to epistemic modality (Barrios 2018).

In the field of corpus annotation there has been a growing interest in the labeling of the factual status of narrated events (Rudinger *et al.* 2018; Ross and Pavlick 2019, among others), given that fact extraction is at the base of applications such as fact-checking or

fake news detection. Undoubtedly, the referent and annotation model for many proposals in this area is *FactBank* (Saurí and Pustejovsky 2009), a corpus developed at Brandeis University which is manually annotated for factual information.

Manual corpus annotation is a time-consuming task; hence, being able to automatize the annotation of factual information is, without doubt, a pressing need. TAGFACT (Alonso *et al.* 2018) is a project whose aim is to develop an automated annotation tool to determine the factual status of events for Spanish, a language for which very little work has been done so far in this field. The present study is part of this project.

The aim of this paper is to determine the role (if any) played by aspectual verb auxiliaries in the factual value of verb periphrases of European Spanish. The role of modal auxiliaries in VPs is clear but, since aspectual information relates to the internal temporal distribution of the event, the contribution of this type is not so straightforward. Specifically, three hypotheses are put forward:

1. In Spanish, the addition of an aspectual auxiliary to the predicate may change its factuality.
2. A by-default factual value can be associated with each subclass of aspectual VP.
3. The tense of the auxiliary also plays a role in the factual value. It should be borne in mind, however, that the present study does not aim at studying the frequency of use of verb periphrasis, but rather the contribution of meaning to factual interpretation.

This is a corpus-based study. As regards methodology, VPs were first classified according to their semantics and a factual value was proposed for each class and subclass, taking into account tenses (cf. Section 3). Then, the proposed values were contrasted with the values attested in real sentences retrieved from corpora. We analyzed 674 sentences exemplifying 28 aspectual VPs. As shall be noted presently, interestingly, the results indicate that, although tenses are one of the most important factors regarding the factual value in sentences with aspectual VPs, a few of them change the factual value of the predicate. Therefore, the main contribution of the present study will be the formulation of rules to be used in the automation of the factual annotation for these special aspectual VPs that have a non-standard behavior in this matter. The corpora used in the analysis were: i) *Corpus*

del Español: News on the Web (NOW),[1] ii) *Corpus del Español: Genre/Historical*,[2] iii) *Corpus del Español del Siglo XXI* (CORPES)[3] and the *Spanish SenSem Corpus*.[4]

The paper is structured as follows. Section 2 briefly presents the term VP, various classifications of Spanish aspectual VPs and how they have been dealt with in several projects related to factuality. Section 3 addresses the methodology followed in the study and Section 4 presents and discusses the main results. Finally, Section 5 concludes the study.

2. ASPECTUAL VERB PERIPHRASES AND FACTUALITY

A VP is a combination of two verbs (auxiliary or V1 and lexical or V2, as in *estoy comiendo* 'I'm eating', *empezamos a leer* 'we started reading', and *viene envuelto* 'it is wrapped') where typically only the auxiliary can be finite. As Topor (2011: 93) states, a VP generally shows the following characteristics:

1. The auxiliary expresses grammatical rather than lexical meaning, such as tense, aspect, modality or voice, as in *seguir* GER 'keep GER' (*siguieron molestando* 'they kept bothering'). Here the situation is presented as occurring before and after the moment of speech.

2. There is loss of categorical meaning in the auxiliary, that is, the selection restrictions observed in the sentence, the commutation properties and the diathesis alternations will depend on the lexical verb and not on the auxiliary, as in *ponerse a* INF 'begin/start INF/GER' (*se pusieron a partir las nueces* 'they started cracking walnuts'; **se pusieron (a) las nueces*; *partieron las nueces* '*they started walnuts; 'they cracked the walnuts').

3. The two verbs form a solid structure, making up a complex structure with internal cohesion, since, for example, the insertion of elements between V1 and V2 is usually not allowed, as in *ir a* INF 'be going to INF' (*no vas a hablar* 'you are not going to talk'; **vas a no hablar* '*you are going not to talk').

Studies dealing with Spanish VPs are quite numerous and diverse (cf. Olbertz 1998; Fernández de Castro 1999; García Fernández 2006; Topor 2011; Fábregas 2019, among others). Olbertz (1998) and Fernández de Castro (1999) offer comprehensive studies on Spanish

---

[1] https://www.corpusdelespanol.org/now/
[2] https://www.corpusdelespanol.org/hist-gen/
[3] https://www.rae.es/banco-de-datos/corpes-xxi
[4] http://grial.edu.es/sensem/corpus

VPs from a functionalist perspective, including a definition of the term VP and criteria of inclusion, syntactic behavior and meaning. Both García Fernández (2006) and Topor (2011) systematically deal with the meaning of VPs and propose a semantic classification. García Fernández (2006) lists VPs in alphabetical order, and each entry in the dictionary includes the meaning, a structural and syntactic description of the VP, as well as a discussion and references, profusely illustrated with examples. In turn, Topor (2011) is a contrastive study of VPs in Spanish and Romanian. After establishing the criteria to consider a sequence of two verbs as a VP and dealing with aspect and modality, the study provides information about the form, class, subclass, definition, examples, paraphrase, synonyms, translation into Romanian and the total number of examples retrieved from her corpus. This is followed by a discussion about the meaning and the degree of grammaticalization, determined by the number of VP criteria fulfilled.

All the above-mentioned studies agree that aspectual VPs may be divided into four main classes —phase, imperfective, perfect and telic (see Table 1 below)— each containing various subclasses, except for telics. Phase VPs refer to the point at which an event is located within its temporal development, such as the beginning or the end. In imperfective VPs, the event is seen "from within" (Comrie 1976: 24). Perfect VPs focus on "the time period that follows the termination or culmination of the eventuality" (Fábregas 2019: 69). Finally, telic VPs present the event as completed (from beginning to end). García Fernández (2006) proposes three subclasses for phase (inchoative, progressive and terminative), four for imperfective (continuous, habitual, inchoative and progressive) and two for perfect (continuative and resultative) VPs. In the present study, we will follow the classification shown in Table 1 below, which is mainly based on García Fernández's (2006) classification, with three differences. The first difference is that García Fernández (2006) classifies as one single class two subclasses which, in the present study (cf. Table 1) are treated as different, namely egressive VPs (*cesar/dejar/parar de* INF 'cease INF/GER; stop GER; stop GER') and terminative VPs (*acabar de* INF 1 'finish INF' and *terminar de* INF 'finish INF'). This is so because there is a crucial difference between the two subclasses (cf. Havu 1997: 197) that may influence the factual status of the VPs: in egressive VPs, the end of the situation is not natural, and so this is presented as interrupted, whereas in terminative VPs, the end is natural and the situation is presented as culminated. The second difference is that García Fernández (2006) groups together in the inchoative phase subclass what Havu (1997) and Topor (2011) consider the inceptive phase (*comenzar a* INF 'begin/start INF/GER', *echar a*

INF 'begin/start INF/GER', *empezar a* INF, *ponerse a* INF 'begin/start INF/GER') and imperfective inchoative VPs (*quedarse* GER 'VERB in progressive tense form'). The distinction between them is that, although both indicate the beginning of a situation, in inchoative VPs the focus is on the fact that the situation is durative. Finally, the third difference is that García Fernández (2006) considers *pasar a* INF 'move on INF' a discursive marker of continuity, while Topor (2011) regards it as phasal, more specifically an ingressive auxiliary, presenting the end of a situation and the beginning of a new one.

Class	Subclass	Meaning and examples
Phase	Egressive	The focus is on the end of the situation, which is presented as interrupted. *Cesar de* INF*, dejar de* INF, *parar de* INF.
	Inceptive	The focus is on the beginning of the situation. *Comenzar a* INF, *echar a* INF, *empezar a* INF, *ponerse a* INF.
	Ingressive	The end of a situation and the beginning of a new one is presented. *Pasar a* INF.
	Prospective	The focus is on a moment prior to the beginning of the situation. *Estar a punto* de INF, *estar por* INF, *ir a* INF, *tardar en* INF.
	Terminative	The focus is on the natural end of the situation. *Acabar de* INF 1, *terminar de* INF.
Imperfective	Continuous	Situation presented as occurring before and after the moment of speech. *Continuar* GER, *ir* GER, *seguir* GER.
	Habitual	The situation is presented as a habit. *Acostumbrar (a)* INF, *soler* INF.
	Inchoative	The beginning of a durative situation is presented. *Quedarse* GER
	Progressive	The focus is on a point in the development of the situation. *Andar* GER, *estar* GER.
Perfect	Continuative	The focus is on a situation from the beginning up to a central point- *Llevar* GER, *venir* GER.
	Resultative	The focus is on the result after the finished situation. *Acabar de* INF 2, *venir* PART.
Telic		The situation is presented as totally completed. Coger y VERB, ir y VERB.

Table 1: Classification of aspectual VPs (based on García Fernández 2006)[5]

Regarding the periphrases selected for the study, we follow Topor (2011), whose work has been the foundation of a trilingual dictionary of periphrases.[6] We have however excluded the following auxiliaries: i) *llevar* PART 'VERB in progressive tense form' and *tener* PART

---

[5] A translation of the VPs can be found in Appendix 1.
[6] http://grial.edu.es/sensem/perifrasis/main?idioma=es

'have Direct Object PART', because they are highly defective; ii) *volver a* INF 'VERB again', because we consider it discursive and not aspectual; and iii) *meterse a* INF 'begin/start INF/ GER', because it is not used in European Spanish. We have completed the list with three additional aspectual VPs taken from García Fernández (2006): *echar* a INF 'begin/start INF/GER', *venir* PART 'VERB in passive voice' and *ir y* VERB 'up and VERB'.

To our knowledge, no studies on the influence of aspectual auxiliaries on the interpretation of the factual status of predicates have been carried out. In contrast, modal auxiliaries have been largely studied since they are considered to express epistemic knowledge, which belongs to the branch of modality directly related to factuality (Portner 2009). Similarly, in the field of corpus annotation, more specifically in the annotation of factuality, the information provided by modal verbs is included in every annotation guide, since modal verbs play a relevant role in the tagging process. However, no particular attention has been paid to aspectual verbs. Exceptionally, in some annotated corpus, like *FactBank* (Saurí and Pustejovsky 2009) or SIBILA (Wonsever *et al.* 2008), a project about the annotation of factuality in Spanish based on *FactBank* (see Wonsever *et al.* 2016), the information provided by verbs that are considered to express aspectual information lexically is reported, but just for a few cases.

In both *FactBank* and SIBILA, this type of verb is identified as an autonomous event and, consequently, the auxiliary is tagged independently from the main verb. For example, in SIBILA, in the expressions *empezar a correr* 'start running' or *parar de correr* 'stop running', the aspectual verbs are tagged with the label ASPECT and *correr* 'run' is tagged as OCCURENCE. However, the two verbs are related to one another, e.g., in *FactBank* the verb *start* and the predicate that follows it are linked by the label INITIATES.

Nonetheless, as explained above, in Spanish, the combinations of these two verbs are considered VPs, and aspectual verbs are considered verb auxiliaries; in other words, they are not considered autonomous or main verbs. The reason for this is that when aspectual verbs modify other predicates, they do not behave as regular main verbs, but add aspectual meaning (i.e., grammatical) to the main verb, which is the one contributing the lexical meaning. In our project, TAGFACT, we have decided to annotate the two verbs together, and deal with them as a single unit while associating just one factual tag to the verb group. Consequently, if the aspectual auxiliary conveys information related to factuality, something that we want to explore in this paper, it should be considered in the process of annotation.

3. Methodology

To carry out the study, we followed five steps. The first was to select the aspectual auxiliaries to analyze, together with their classification, which yielded 28 VPs grouped in four classes and 12 subclasses, as was shown in Table 1 (cf. Section 2).[7] As shall be shown in Section 4 (cf. Tables 4–6), within these 12 subclasses, VPs are further subdivided according to their meaning, so that new groups are created. Some of these groups consist of just one item while others assemble several auxiliaries together. Overall, we have 18 groups; for example, within the class of inceptive VPs, two groups are defined: a group for *echar/ponerse a* INF 'begin/start INF/GER' and another one for *comenzar/empezar a* INF 'begin/start INF/GER'.

The second step was to establish the tagset to annotate event factuality. We used the TAGFACT's proposal (Vázquez and Fernández-Montraveta 2020), which follows Saurí and Pustejovsky's (2009) and Diab *et al*.'s (2009) projects. Thus, events were annotated regarding four categories: commitment, polarity, time and type of event (dynamicity). Each of these categories was assigned different values and each predicate (whenever factual annotation was applicable) was tagged with a combination of them, e.g., commitment + positive + past + event, which is the most typical combination. To allow the comparison of our results with those in other corpora, the researchers of the TAGFACT project simplified the results and translated each possible combination into just one label (Vázquez and Fernández-Montraveta 2020). The simplified list of tags which are used is shown below.

A) FACT (F), which includes events —as in (1)— and states referring to the present or past presented by the narrator as having happened. This label also includes absolute truths and actions presented as habitual, as in (2) and (3) respectively.

(1) *Se fijó en las aceitunas machacadas y <u>cogió y se comió</u> dos de golpe*. (CORPES)
     'He/she noticed the crushed olives up and ate two simultaneously'.

(2) *El hombre se <u>ha solido erigir</u> como la autoridad religiosa para determinar lo lícito o lo ilícito*. (NOW)
     'Man <u>has typically considered</u> himself the religious authority to determine what is licit and what is not'.

(3) *Pekín <u>sigue calificando</u> a Taiwán de provincia rebelde*. (SenSem)
     'Beijing <u>keeps qualifying</u> Taiwan as a rebel province'.

---

[7] Telic auxiliaries are considered both a class and a subclass since they are not further subdivided.

Here it is worth noticing that truths (cf. 2) and habitual actions (cf. 3) do not refer to any particular event that happens at a specific place and time. Absolute truths are situations that are true for a community while habitual actions describe an iteration of events. In this sense, some authors understand that in habitual actions more than one action is predicated (Mendikoetxea 1999; Fernández-Montraveta and Vázquez 2017). We decided to tag examples such as (2) and (3) as FACT since they refer to real situations that took place, at least once, independently of future repetitions.

B) COUNTERFACT (CF), which refers to all the events and situations that are presented as not having occurred:

(4) *Pues Carlos <u>iba a hacer</u> un coulant pero le salió un bizcocho seco*.[8] (NOW)
'Carlos <u>was going to make</u> a coulant but it turned out to a dry sponge cake'.

C) UNDERSPECIFIED (U), which we use to tag events and states referred to future events or to a present or past situation described by the narrator as possible or probable to a greater or a lesser degree, as shown in (5). In our proposal, future events are categorized as uncertain because this temporal dimension is intrinsically related to doubt to a larger or smaller extent. Thus, like in most studies dealing with this type of semantic annotation (Saurí and Pustejovsky 2009; Soni *et al.* 2014; Minard *et al*. 2016, among others), future events are placed together with uncertain present or past events.

(5) *Los autores ya <u>andarán buscando </u>otros lugares donde resarcirse*. (CORPES)
'The authors <u>must already be looking for</u> other places to make up for'.

D) NON-APPLICABLE (NA), which is used in wishes, hypotheses, orders, questions, suggestions, and all situations which are part of an imaginary world and are not relevant for factuality.

(6) *En el peor escenario los ciudadanos <u>empezarían a vender </u>viviendas provocando una gran caída de precios*. (NOW)
'In the worst-case scenario, people <u>would start selling </u>homes causing prices to fall sharply'.

The third step was to establish default factuality values for each VP, based on semantics and tense, as shown in Table 2. Present and past tenses are associated to FACT, future tenses to UNDERSPECIFIED and conditional to unreal situations (NON-APPLICABLE). Only some egressive and prospective VPs have been assigned a different value for present and past.

---

[8] This example has been slightly modified to simplify its translation.

Class	Subclass	Hypotheses Present	Past	Future	Conditional
Phase	Egressive	CF	CF	U	NA
	Inceptive	F	F	U	NA
	Ingressive	F	F	U	NA
	Prospective	U	CF	U	NA
	Terminative	F	F	U	NA
Imperfective	Continuous	F	F	U	NA
	Habitual	F	F	U	NA
	Inchoative	F	F	U	NA
	Progressive	F	F	U	NA
Perfect	Continuative	F	F	U	NA
	Resultative	F	F	U	NA
Telic		F	F	U	NA

Table 2: Hypothesized behavior of aspectual auxiliaries

As can be seen in Table 2, future and conditional tenses are predicted to have a uniform value, independently of the semantics of each VP. All events in the future are expected to be UNDERSPECIFIED, given that there is not enough information to determine whether they have taken place. As for conditional tenses, they are typically used to refer to wishes and hypothetical situations, hence the NON-APPLICABLE label is postulated (Real Academia Española 2009: 1778).

As for present and past tenses, phase auxiliaries behave differently depending on their meaning. First, egressive VPs are complex —that is, they indicate that a situation stops, which means that they were true (FACT) prior to their end, but they are no longer true at the time of reference (COUNTERFACT). Thus, we have predicted a COUNTERFACT value since these VPs focus on the cessation of the action. In contrast, terminative VPs also denote that the situation is finished, but the focus is on the final phase of that situation, and hence we have hypothesized a FACT value. Second, prospective VPs in the present were labeled UNDERSPECIFIED because they refer to future situations, whereas the VPs in past tenses tend to express a failed attempt, and so they have been assigned the value COUNTERFACT. The remaining phase VPs, as well as imperfective, perfect and telic VPs, are expected to describe facts in the present and in the past.

The fourth step consisted of the retrieval of examples of the aspectual VPs under study from the NOW corpus, the *Corpus del Español: Genre/Historical*, CORPES and *SenSem*

in their subsection of texts in Spanish (cf. Section 1). Only affirmative indicative clauses were collected, and the journalistic register was the most frequent register analyzed (even though there are some literary examples as well). As for subordinate clauses, only relatives and adversatives were considered since it has been shown that the factuality of the main clause does not affect their own factuality (Saurí 2008). When insufficient examples were attested in the corpora, these were taken from other online Spanish newspapers.

For each of the 18 groups of VPs, five examples of indicative tenses were randomly collected: simple present, present perfect, past imperfective, past perfect, past perfective, past anterior, simple future, perfect future, simple conditional and perfect conditional. These five examples were distributed as evenly as possible between the different auxiliaries in each group. Obviously, some auxiliaries were more frequent and easier to retrieve while some combinations were not found. However, the analysis of the frequency of the various auxiliaries is not part of the objectives of the present study.

All in all, we analyzed 674 sentences, but it was not possible to find examples for certain tenses, especially for the past anterior, as this combines with terminative VPs only. The distribution by tenses is presented in Table 3. All the other tenses have a similar distribution (from 71 to 90 sentences), except for the future perfect and conditional perfect (51 and 56 examples, respectively). The simple present (90 sentences), the past imperfective (90 sentences) and the past perfective (84 sentences) are the most frequent tenses.

Tense	Number of sentences	Tense	Number of sentences
Simple present	90	Past anterior	5
Present perfect	74	Simple future	74
Past imperfective	90	Future perfect	51
Past perfect	71	Simple conditional	79
Past perfective	84	Conditional perfect	56

Table 3: Distribution of tenses analyzed in the corpus

The fifth and final step was the annotation of the sentences with respect to their factual value. First, for each sentence, the VP was assigned a factuality label by three annotators. The assignment was based on the whole context in which the sentence occurred, the meaning of the auxiliary plus the meaning of the lexical verb, the tense and any other relevant aspects. A consensus was reached in controversial examples. As discussed in Section 2, unlike in other proposals, our study considers the main verb together with the auxiliary

and, therefore, only one factual value is assigned to the whole structure, understanding that there is only one event, rather than two, modified by the auxiliary (Topor 2011). Likewise, whenever necessary, each sentence was rewritten and reannotated without the aspectual auxiliary, to analyze how 'transparent' VPs are in relation to factuality (see our first hypothesis in Section 1).

4. Results and discussion

This section explores to what extent aspectual auxiliaries modify the factual status of predicates. It provides the results of comparing the values predicted for the 12 subclasses of aspectual VPs in each of the ten tenses of the indicative mood with the tagging of the examples analyzed and how these predictions vary, if they do, when the auxiliary is deleted.

Sections 4.1 (phase), 4.2 (imperfective) and 4.3 (perfect and telic) discuss the results for each class. Section 4.4 offers general remarks about the homogeneity of the behavior observed with respect to factual values. In other words, we analyze to what extent the factual values of the different groups in each tense are constant or if, by contrast, there is too much variability to draw any definite conclusion regarding the formalization of rules.

A total of 50 sentences per group was expected: five examples for each tense and group, and ten tenses. However, it was not always possible to retrieve examples for each combination (see Section 3). The class that shows the lowest frequency is the telic class (19 examples), whereas the class with the highest frequency is the egressive one (45 examples). Out of the 18 groups under analysis, 11 groups are exemplified by at least 40 sentences. A total of 674 sentences were analyzed, distributed as follows: 365 for phase VPs (five subclasses, nine groups and 15 VPs), 189 for imperfective (four subclasses, five groups and eight VPs), 101 for perfect (two subclasses, three groups and four VPs) and 19 for telic (two VPs).

As for the tenses with the lowest number of examples, the past anterior was only attested with the auxiliary *acabar de* INF 1 'finish INF' and *terminar de* INF 'finish INF' (terminative VPs), and the conditional and future perfect, which were not found with *estar a punto de* INF 'be about to INF', *ir a* INF 'be going to INF', *coger y* VERB 'up and VERB', *ir y* VERB 'up and VERB', *andar* GER 'VERB in progressive tense form + always' and *quedarse* GER 'VERB in progressive tense form'. This low frequency is in line with the incidence of these tenses in general language (Troya Déniz 2007: 593).

Most of our predictions regarding factual values (Table 1) were confirmed. However, our association of conditional tenses (both simple and perfect) with the value NON-APPLI-CABLE for all VPs, regardless of the subclass, does not hold in all cases. We assumed that conditional tenses mostly describe desired and hypothetical situations (NON-APPLICABLE). However, conditional tenses can also be used to express future actions (UNDERSPECIFIED) or to present future events narrated in the present or the past (FACT, COUNTERFACT). In fact, the examples showed that it is not possible to establish a predominant use, that is, a default value.

In what follows, results are considered for each group of VPs independently, and only those cases in which the prediction is not met —or some special behavior is observed— will be discussed.

4.1 Phase VPs

Table 4 shows the data for sentences with a phase VP. From a quantitative perspective, and in relation to the use of VPs in different tenses, some special cases concerning perfect tenses in combination with the prospective VP *ir a* INF 'be going to INF' are observed. According to García Fernández (2006: 179), this VP is incompatible with all perfect tenses, including the past perfective. Nevertheless, examples of the periphrastic use of *ir* a INF are attested in the past perfective in our corpus, as shown in (7):

(7) *<u>Fue a echarse</u> entonces <u>mano</u> al móvil, pero se frenó: no; había cosas que se decían sólo en persona aunque tuvieran que esperar*. (CORPES)
'Then he <u>was about to reach</u> for his cell phone, but stopped himself: no, there were things that could only be said in person, even if they had to wait'.

Phase	Egressive	Inceptive	Ingressive		Prospective				Terminative	Total
VPs	Cesar/dejar/ parar de INF	Echar/ ponerse a INF	Comenzar/ empezar a INF	Pasar a INF	Estar a punto de INF	Ir a INF	Estar por INF	Tardar en INF	Acabar de INF 1/ terminar de INF	Total
Simple present	5CF	4F/1U	4F/1U	5F	5U	5U	5U	4F/1CF	5F	**45**
Present perfect	5CF	5F	5F	5F	5CF	0	2F/ 2CF	5F	4F/1CF	**39**
Past imperfective	5CF	5F	5F	5F	5CF	2F/ 1CF/2U	4CF/1U	2F/3CF	5F	**45**
Past perfect	5CF	5F	5F	5F	5CF	0	0	5F	5F	**35**
Past perfective	5CF	5F	5F	5F	5CF	5CF	5CF	5F	5F	**45**
Past anterior	0	0	0	0	0	0	0	0	5F	**5**
Future	5U	5U	5U	5U	5U	0	4U	5U	1CF/4U	**39**
Future perfect	5U	4U/1F	5U	5U	5U	0	0	3U	5U	**33**
Simple conditional	1F/ 4NA	3U/ 2NA	1F/2U/ 2NA	2F/2U/ 1NA	5U	5U	3F/1CF/1NA	4F/1U	4F/1U	**45**
Conditional perfect	1CF/2U/2NA	4U	5F	5U	5U	0	0	5CF	1CF/4U	**34**
Total	**45**	**44**	**45**	**45**	**45**	**20**	**28**	**43**	**50**	**365**

Table 4: Factuality values for Phase VPs in the corpus

There are other special cases in relation to phase VPs. There are two examples with one sentence annotated as UNDERSPECIFIED, when they were expected to depict FACT. This is the case in all inceptive VPs *echar/ponerse a* INF 'begin/start INF/GER' (8a) and *comenzar/ empezar a* INF 'begin/start INF/GER' (9a). As the examples show, both sentences refer to a future planned event (which is one of the potential uses of the simple present) and this would also be true without the periphrastic auxiliary (as in examples (8b) and (9b)).

(8a) *El Director de FERCAM terminó comentando que a partir de mañana <u>se pone a trabajar</u> en la 57 Edición.*[9]
'The Director of FERCAM ended by remarking that as of tomorrow <u>he will start working</u> on the 57th Edition'.

(8b) *El Director de FERCAM terminó comentando que a partir de mañana <u>trabajará</u> en la 57 Edición.*
'The Director of FERCAM ended by remarking that from tomorrow <u>he will work</u> on the 57th Edition'.

(9a) *Esta es una de las consideraciones del borrador de contrato-programa que hoy <u>comienza a analizar</u> el consejo de administración del ente.* (SenSem)
'This is one of the considerations of the program contract draft that the entity's board of directors <u>begins to analyze</u> today'.

(9b) *Esta es una de las consideraciones del borrador de contrato-programa que hoy <u>analiza</u> el consejo de administración del ente.*
'This is one of the considerations of the program contract draft that the entity's board of directors <u>is analyzing</u> today'.

In addition, the prospective VP *tardar en* INF 'take time INF', as a member of this subclass of phase VPs, was expected to involve UNDERSPECIFIED as the interpretation value for this present tense. However, the results show that this was never the case. To start with, four out of the five collected examples were tagged as FACT. These four examples do not really describe a standard fact but a habitual action (10), also considered FACT in the present analysis (see Section 3).

(10a) *Las catedrales <u>tardan en construirse</u> siglos, tardan en reconstruirse años, décadas.* (NOW)
'Cathedrals <u>take </u>centuries <u>to build</u>, years, decades to rebuild'.

(10b) *Las catedrales <u>se construyen </u>en siglos, se reconstruyen en años, décadas.*
'Cathedrals <u>are built</u> in centuries, rebuilt in years, in decades'.

Secondly, the other examples were tagged as Counterfact (11). The prospective value of the auxiliary *tardar en* 'take time INF' differs slightly from the other VPs in this group as, rather than presenting a future event, it focuses on the period before, where it is not accomplished yet. If the auxiliary would not have been there, the sentence would have been completely different, so it can be stated that it is not transparent in the present when the interpretation is not habitual, since with the auxiliary the predicate is Counterfact (11) and, without it, it would be FACT.

(11) *Ahora sujetan fuerte entre esos dedos el billete que acaban de regalarles. Los 20 minutos que <u>tarda en arrancar</u> este tren se hacen eternos.* (NOW)
'Now they hold tightly in their fingers the ticket they have just been given. The 20 minutes this train <u>takes to start </u>last forever'.

Some of our initial hypotheses were not met for some examples in the past imperfective, past perfective and past perfect of prospective VPs. For example, we expected sentences in the past imperfective to merely express Counterfact, but different values were rather attested: for instance, examples (12a), (13a) and (14) illustrate the use of *ir a* INF 'be going to INF' with, FACT, UNDERSPECIFIED and COUNTERFACT values, respectively. In (12a) it is used to refer to a past event, in (13a) to a future event and in (14) to a situation that did not happen.

(12a) *Tarde o temprano <u>iba a pasar</u>, Violeta y Julen tenían que verse las caras después de que ella decidiera romper con la relación.* (NOW)
'Sooner or later, <u>it was going to happen</u>, Violeta and Julen had to meet face to face after she decided to break off the relationship'.

(12b) *Tarde o temprano ? <u>pasaba</u>, Violeta y Julen tenían que verse las caras (...).*
'Sooner or later ? <u>it happened</u>, Violeta and Julen had to face each other (...)'.

(13a) *... en contra de la ley del matrimonio gay que <u>iba a ser aprobada</u> por el Congreso.* (NOW)
'... against the gay marriage law that <u>was to be approved</u> by Congress'.

(13b) *(...) en contra de la ley del matrimonio gay que ? <u>era aprobada</u> por el Congreso.*
'(...) against the gay marriage law that ? <u>was approved</u> by the Congress'.

(14) *Nicolás <u>fue a hablar</u>, pero Teresa le hizo el gesto del silencio.* (CORPES)
'Nicolás <u>was about to speak</u>, but Teresa signaled him to keep quiet'.

The examples below illustrate the auxiliary *estar por* INF 'be about to INF' in the past imperfective. Here, COUNTERFACT —the expected value— is the most frequent value identified (see example 15a), except for example (16a), where the value UNDERSPECIFIED is attested.

(15a) *Estaba por continuar el recuento de mis peripecias cuando el doctor Sol-devila se asomó a la puerta del despacho con aspecto cansado y resoplando.* (CORPES)
‘I was about to go on the account of my adventures when Dr. Soldevila appeared at the office door looking tired and snorting’.

(15b) *Continuaba el recuento de mis peripecias cuando el doctor Soldevila se asomó a la puerta del despacho con aspecto cansado y resoplando.*
‘I went on with the account of my adventures when Dr. Soldevila appeared at the office door looking tired and snorting’.

(16a) *Entonces llamé al tipo. Yo estaba por escribir una nota romántica sobre cómo eran los pueblos indígenas.* (NOW)
‘Then I called the guy. I was about to write a romantic note about what indigenous people were like’.

(16b) *Entonces llamé al tipo. Yo? escribía una nota romántica sobre cómo eran los pueblos indígenas*.
‘Then I called the guy. I was writing a romantic note about how the indigenous people were’.

According to García Fernández (2006: 157–158), *estar por* INF ‘be about to INF’ tends to occur with imperfective tenses but our data show that it may also occur with the present perfect and the past perfective. However, in contrast to García Fernández’s claim that the combination of this VP with present perfect and past perfective expresses an attempt (*conatus*), that is, UNDERSPECIFIED, in most of our examples with these two tenses the value is COUNTERFACT.

These last two VPs, *ir a* INF ‘be going to INF’ and *estar por* INF ‘be about to INF’, together with other elements in the context, determine the factual value of the predicate and, thus, are not transparent, as can be seen in (12b) and (13b) above, a rephrasing of (12a) and (13a), without the auxiliary. In the case of *ir a* INF, deleting the auxiliary makes the sentences ungrammatical in (12b) and (13b). As for *estar por* INF ‘be about to INF’, the factuality of the sentence changes and the UNDERSPECIFIED interpretation becomes FACT in both (15b) and (16b).

In the case of *tardar en* INF ‘take time INF’, the VP also presents a different factual behavior in relation to the rest of VPs in this subclass in the past. It was expected to describe COUNTERFACT when referring to past situations, but this is only attested in three cases in the past imperfective (17), while all other sentences in the past were assigned the FACT value (18).

(17)   *Pero el comunicado <u>tardaba en salir de</u>l horno.* (NOW)
       'But the communication <u>took a long time to come out </u>of the oven'.

(18)   *La defensa alemana, dormida, <u>tardó en reaccionar</u>.* (NOW)
       'The sleepy German defense <u>was slow to react</u>'.

Regarding future tenses, *ir a* INF 'be going to INF' (prospective) in the simple future, we have not found affirmative sentences for this VP in this tense, that is, the only sentences attested are both negative and interrogative, as shown in example (19) below. In fact, these kinds of sentences are rhetorical and do not pose real questions, as can be seen in (20), where no question marks are used. Furthermore, in all cases, the meaning of this VP is actually different from the periphrastic use of *ir a* INF. When used in interrogative (and negative) sentences, it adds an element of disbelief or emphasis to the basic prospective meaning. In this context, it could be considered an idiomatic expression with the meaning of 'I expect you not to do something'.

(19)   *¿No me <u>irás a fallar</u> ahora?* (CORPES)
       'You<u> won't fail</u> me now, right?'.

(20)   *No <u>irás a hacerte </u>la víctima. No lo soporto.* (CORPES)
       '(I hope that) <u>you're not going to play</u> the victim. I can't stand it'.

For the future tense, in general, our prediction was that both the simple future and the future perfect would behave in the same way and would be associated with the UNDERSPECIFIED factual value. However, our sentences for the VPs *echar/ponerse a* INF 'begin/start INF/ GER' (a group of inceptive VPs) show a different behavior with the future perfect, where we found four sentences exemplifying UNDERSPECIFIED (22) —the expected value— and one with a FACT meaning (21a).

(21a) *¡Qué sé yo las veces que Fernanda <u>se habrá puesto a tomarme el pelo</u>, lla-*
       mándome la devora libros! (CORPES)
       'I don't know how many times Fernanda <u>must have started to tease me</u>, calling me the book devourer!'

(21b) *¡Qué sé yo las veces que Fernanda <u>me habrá tomado el pelo</u>, llamándome la*
       devora libros!
       'I don't know how many times Fernanda <u>must have teased me</u>, calling me the book devourer!'

(22)   *Habrá gente que se <u>habrá puesto a pitar</u> y todo en medio del atasco.* (NOW)
       'Some people <u>will have </u>even <u>started to honk</u> in the middle of the traffic jam'.

Example (21a), annotated as FACT, has a habitual reading (focalized in the past) because of the expression *Qué se yo las veces que…!* 'I don't know how many times'. This expression is responsible for the interpretation of the event, which is introduced as something that actually occurred. It is clear that the event did happen in the past on a repeated basis, and the uncertainty typically associated with the future perfect is lost. What the reader does not know is the exact number of times that it took place. Actually, the annotation would be the same without the periphrastic auxiliary, as can be seen in (21b), since it is a feature of the habitual construction. In fact, this would have also been the case if the same expression had been used in (22).

Finally, as regards conditional tenses, they behave in different ways (cf. Table 3). This is exemplified in sentences with the egressive subclass, where the rich casuistry is present both in the simple conditional and in the perfect conditional. In (23a), we find an example of an unreal situation (NON-APPLICABLE) in the simple conditional, i.e., an event that depends on a condition. As for (24a), also in the simple conditional, it exemplifies a FACT: an affirmative future situation from a past perspective, so we know it has actually happened. In (25a), there is a negative situation narrated from the past using conditional perfect (COUNTERFACT). Finally, instance (26a), also in the conditional perfect, exemplifies an UNDERSPECIFIED event whose factual status is unclear since it expresses a possibility. In all these cases, the trigger for these interpretations is the verb tense. Therefore, the periphrastic auxiliary does not play a role in establishing the factual status of the sentences, that is, it is transparent, as can be seen in (23b–26b).

(23a) *Si el cuerpo no estuviese animado por el alma, <u>cesaría de existir</u>.* (NOW)
  'If the body were not animated by the soul, it <u>would cease to exist</u>'.

(23b) *Si el cuerpo no estuviese animado por el alma, no <u>existiría</u>.*
  'If the body were not animated by the soul, it <u>would not exist</u>'.

(24a) *Toñi <u>dejaría de presentar</u> Viva La Vida para dejar paso a Emma García.* (NOW)
  'Toñi <u>would stop presenting</u> Viva La Vida to make way for Emma García'.

(24b) *Toñi <u>ya no presentaría</u> Viva La Vida para dejar paso a Emma García.*
  'Toñi <u>would no longer present</u> Viva La Vida to make way for Emma García'.

(25a) *La penúltima causa estudia la venta de parcelas municipales, por la que el ayuntamiento andaluz <u>habría dejado de ingresar</u> 6,4 millones.* (CORPES)
  'The penultimate case studies the sale of municipal plots, for which the Andalusian city council <u>would have stopped receiving</u> 6.4 million'.

(25b) *La penúltima causa estudia la venta de parcelas municipales, por la que el ayuntamiento andaluz <u>no habría ingresado</u> 6,4 millones.*
'The penultimate case studies the sale of municipal plots, for which the Andalusian city council <u>would not have received</u> 6.4 million'.

(26a) *La mitad de ellas <u>habría dejado de usarla</u> en 2002 al conocer sus potenciales efectos secundarios*. (CORPES)
'Half of them <u>would have stopped using</u> it in 2002 when they learned of its potential side effects'.

(26b) *La mitad de ellas <u>ya no la usaría</u> en 2002 al conocer sus potenciales efectos secundarios.*
'Half of them <u>would no longer use</u> it in 2002 when they learned of its potential side effects'.

All cases were tagged as UNDERSPECIFIED for *estar a punto de* INF 'be about to INF', in both conditional tenses (27), and for *ir a* INF 'be going to INF' only in the simple conditional (28).

(27) Un avión italiano <u>habría estado a punto de estrellarse</u> con un ovni en 1991.[10]
'An Italian plane <u>was reportedly on the verge of</u> crashing into a UFO in 1991'.

(28) *A continuación me expuso su plan: (...) Alicia <u>se iría a vivir</u> con su hijo a la calle de las Carolinas*. (CORPES)
'He then told me his plan: (...) Alicia <u>would go to live</u> with her son on Carolina street'.

4.2 Imperfective VPs

Table 5 shows the distribution and behavior of imperfective VPs in our corpus. As for the distribution of tenses, no examples in the past anterior were attested. This was also the case in the future and conditional perfect with the VP *quedarse* GER 'VERB in progressive tense form' and in the simple conditional for *andar* GER 'VERB in progressive tense form + always'. As for the simple future in these last two VPs, we could not find all five sentences. Finally, regarding *andar* GER, the same was found for present and past perfect.

---

[10] Taken from https://www.publico.es/internacional/avion-estuvo-punto-chocar-ovni.html

Imperfective	Continuous	Habitual	Inchoative	Progressive		Total
Verb phrase	*Continuar/ir/ seguir* GER	*Acostumbrar (a)/ soler* INF	*Quedarse* GER	*Estar* GER	*Andar* GER	
Simple present	5F	5F	5F	5F	5F	**25**
Present perfect	5F	5F	5F	5F	3F	**23**
Past imperfective	5F	5F	5F	5F	5F	**25**
Past perfect	5F	5F	5F	5F	1F	**21**
Past perfective	5F	5F	5F	5F	5F	**25**
Past anterior	0	0	0	0	0	**0**
Simple future	5U	5U	1U	5U	4U	**20**
Future perfect	5U	5U	0	5U	0	**15**
Simple conditional	2F/3NA	2F/3NA	0	5U	5U	**20**
Conditional perfect	2CF/3U	2CF/3U	0	5U	0	**15**
Total	**45**	**45**	**26**	**45**	**28**	**189**

Table 5: Imperfective VPs

When it comes to the factual values for imperfective VPs, they were all expected to show FACT in present and past tenses. Future tenses, as mentioned before, are predicted to present situations as UNDERSPECIFIED (29). All these expectations were fulfilled for all subclasses and all VPs, as can be seen in Table 5.

    (29a) *Hoy no iré a comer, o <u>me quedaré trabajando</u> hasta muy tarde*. (CORPES)
          'I won't go to lunch today, or I'<u>ll be working</u> late'.

    (29b) *Hoy no iré a comer, o <u>trabajaré</u> hasta muy tarde.*
          'Today I will not go to eat, or I <u>will work</u> until very late'.

As mentioned above, conditional tenses are used in different contexts in Spanish. In the case of imperfective VPs, we found three possible uses: i) presenting unreal situations (desires or conditions), tagged as NON-APPLICABLE (30a), ii) depicting present situations narrated from the past, when they can express either FACT (31a) or COUNTERFACT and iii) when the factual status is unclear (32a), tagged as UNDERSPECIFIED.

As for transparency, we can say that the role of the auxiliaries of this class is not crucial in any tense since the factual value remains the same, as shown in (28b)–(32b).

(30a) *Rayo McQueen es adorado por millones de niños, que si tuvieran que escoger entre perder su coche de juguete o a su abuelita, <u>se quedarían pensando</u> un rato.* (CORPES)
'Lightning McQueen is adored by millions of children who, if they had to choose between losing their toy car or their grandmother, <u>would think</u> it over'.

(30b) *Rayo McQueen es adorado por millones de niños, que si tuvieran que escoger entre perder su coche de juguete o a su abuelita, <u>pensarían</u> un rato.*
'Lightning McQueen is adored by millions of children, who if they had to choose between losing their toy car or their grandmother, they <u>would think</u> for a while'.

(31a) *El contenido y duración <u>iría variando</u> a lo largo del siglo XIX.* (NOW)
'The content and duration <u>would vary</u> throughout the 19th century'.

(31b) *El contenido y duración <u>variaría</u> a lo largo del siglo XIX.*
'The content and duration <u>would vary</u> throughout the 19th century'.

(32a) *Si lo llego a saber, se <u>habría quedado vendiendo</u> máquina de batidos.* (NOW)
'If I had known, she <u>would have been selling</u> smoothie makers'.

(32b) *Hunj Li, de 44 años, <u>habría conversado</u> con un socio y con otro amigo de la misma nacionalidad, quienes fueron los primeros arrestados por la Policía.*
'Hunj Li, 44, <u>was reportedly talking</u> with an associate and another friend of the same nationality, who were the first to be arrested by the police'.

4.3 Perfect and telic VPs

Table 6 shows the data regarding perfect and telic VPs. It can be noticed that, as was the case with other VPs discussed above, no examples in the past anterior were attested, and examples in the future perfect were not frequent either. In perfect VPs there is a higher degree of defectivity than in phase and imperfective VPs: with the exception of the simple present and the past imperfective, it was not possible to retrieve five examples for other tenses. The group of telic VPs is particularly noticeable in this respect, and we could only retrieve full representation of these VPs for three tenses out of ten. This is probably because they are typically used in oral Spanish production (Topor 2011: 226).

Perfect	Continuative	Resultative		Total perfect	Telic
Verb phrase	*Llevar/venir* GER	*Venir* PART	*Acabar de* INF 2		*Andar* GER
Simple present	5F	5F	5F	15	5F
Present perfect	5F	5F	1F	11	1F
Past imperfective	5F	5F	5F	15	5F
Past perfect	5F	5F	5F	15	0
Past perfective	1F	5F	3F	9	5F
Past anterior	0	0	0	0	0
Simple future	3U	5U	4U	12	3U
Future perfect	1U	0	2U	3	0
Simple conditional	5U	2F/3U	4U	14	0
Conditional perfect	1U	5U	1U	7	0
Total	**31**	**40**	**30**	**101**	**19**

Table 6: Perfect and telic VPs

Overall, the data fulfill the expected results for perfect and telic VPs: present (33a) and past (34a) express FACT and future tenses (35a) UNDERSPECIFIED, and these expectations are met, as can be seen in Table 6. In all these cases, the auxiliary is transparent, that is, the factual value is related to the tense, as can be seen in the (33b–35b).

(33a) *La carta de cese fulminante <u>viene firmada</u> por el nuevo director de Radio Nacional, Raúl Heitzmann*. (NOW)
'The termination letter <u>is signed</u> by the new director of Radio Nacional, Raúl Heitzmann'.

(33b) *La carta de cese fulminante <u>está firmada</u> por el nuevo director de Radio Nacional, Raúl Heitzmann*.[11]
'The letter of termination <u>is signed</u> by the new director of Radio Nacional, Raúl Heitzmann'.

---

[11] The use of *venir* PART 'VERB in passive voice' (in this case, *firmada* 'signed') implies a resultative interpretation. Thus, the present of the verb *firmar* 'sign' cannot be used to check transparency. Instead, copulative *estar* 'be' has been used, since this verb with a participle also gives a resultative reading.

(34a) *El mercado de la vivienda <u>ha acabado de aterrizar</u> y a partir de ahora se espera una subida de precios*. (NOW)
'The housing market <u>has just</u> <u>landed</u> and from now on a price increase is expected'.

(34b) *El mercado de la vivienda <u>ha aterrizado</u> y a partir de ahora se espera una subida de precios.*
'The housing market <u>has landed</u> and from now on prices are expected to rise'.

(35a) *Presidente, te <u>estará llegando</u> el visitante, que algo vendrá pidiendo…* (CORPES)
'President, the visitor <u>will be coming to</u> you; he will be asking for something…'.

(35b) *Presidente, te <u>llegará</u> el visitante, que algo pedirá…*
'President, the visitor <u>will come</u> to you; he will be asking for something…'.

As for conditional tenses, they present fewer labels, since we find FACT in the resultative VP *venir* PART 'VERB in passive voice' (36a) and all other groups present an UNDERSPECIFIED interpretation (37a, 38a). Furthermore, the perfect conditional always keeps the UNDERSPECIFIED value.

These auxiliaries behave again with transparency in conditional tenses, as can be seen in examples (36b–38b), where the same factual values are kept. The translation into English would in fact be the same and the meaning of approximation contributed by the auxiliary would be lost.

(36a) *La primera etapa <u>vendría marcada</u> por una época de esplendor, cuando obtuvo el rango colegial durante el reinado de Sancho IV (...)* (CORPES)
'The first stage <u>would be marked</u> by a period of splendor, when it obtained the collegiate rank during the reign of Sancho IV (...)'.

(36b) *La primera etapa <u>estaría marcada</u> por una época de esplendor, cuando obtuvo el rango colegial durante el reinado de Sancho IV…*
'The first stage <u>would be marked</u> by a period of splendor, when it obtained the collegiate rank during the reign of Sancho IV…'.

(37a) *Estas enfermedades <u>vendrían derivadas</u> de los fuertes cambios climáticos que experimentaba el planeta*. (NOW)
'These illnesses <u>were probably derived</u> from the strong climatic changes that the planet was experiencing'.

(37b) *Estas enfermedades <u>se derivarían</u> de los fuertes cambios climáticos que experimentaba el planeta.*
'These illnesses <u>were probably due to</u> the strong climatic changes that the planet was experiencing'.

(38a) *Un censo único europeo <u>acabaría de resolverlo</u>*. (NOW)
'A single European census <u>would solve the problem</u>'.

(38b) *Un censo único europeo <u>lo resolvería</u>.*
　　'A single European census <u>would solve the problem</u>'.

4.4 Homogeneity of factual values

Here we discuss to what extent there is variability in the factual values of the different groups of VPs and tenses. If there is variability, the prediction of a factual interpretation would be difficult and, consequently, the formalization of rules would be challenging.

Broadly speaking, the array of factual interpretations is quite homogeneous in most cases in our data. In particular, nine out of the 18 VP groups under study show a unique factual interpretation in all the examples collected in our corpus. Three groups show several interpretations for at least in one tense and the group that shows the highest diversity displays it in three tenses. The class of phase auxiliaries is clearly the most complex class, with more variability from a factual point of view. In the other two classes of VPs, the behavior is homogeneous except for conditional tenses, especially the simple conditional.

As for phase VPs, the only stable VP is *estar a punto de* INF 'be about to INF', with no variety of annotations whatsoever. Secondly, *ir a* INF 'be going to INF' and *pasar a* INF 'move on INF' only show two possible interpretations in the past imperfective, the former, and in simple conditional, the latter. Thirdly, *acabar de* INF 1 'finish INF' and *terminar de* INF 'finish INF' show two possible tags in two tenses, namely the present perfect and the simple conditional. All other phase auxiliaries show multiple interpretations in three tenses (normally conditional tenses, past imperfective or present). The behavior of imperfective VPs is more homogenous: there are deviations from the expected behavior in two out of five groups (continuous and habitual VPs), but in both cases these differences are attested in conditional tenses only. Regarding perfect and telic VPs, they constitute the most homogeneous classes. There is only one case with more than one interpretation, namely *venir* PART 'VERB in passive voice', in the simple conditional. In the remaining VPs and tenses of these classes, the behavior is completely homogeneous.

The second aspect that was analyzed is tenses. All of them show more than one interpretation, except for the simple future, which is always labeled with just one tag. The simple present has several interpretations and we have identified all values for it except for NON-APPLICABLE. The same holds for the past imperfective, even though the tag UNDER-SPECIFIED is rare. The present perfect, the past perfect and the past perfective only present

two possibilities: FACT (the most common) or COUNTERFACT. As for the future perfect, also two tags were identified: UNDERSPECIFIED (the most common) and FACT. The label NON-APPLICABLE is restricted to conditional tenses only, and the simple conditional is the one with the highest number of possible interpretations (all four tags).

5. CONCLUSIONS

The present study is part of the TAGFACT project, whose aim is to create a tool for automatically annotating the factuality of predicates. Our main objective has been to prove the relevance of aspectual auxiliaries for the factuality of the sentences in which they appear. We grouped aspectual VPs in four classes, 12 subclasses and 18 groups and analyzed their behavior in 674 sentences. Whenever possible, the examples were equally distributed between the ten indicative tenses included in the study.

Regarding our first hypothesis that, in Spanish, the addition of an aspectual auxiliary to the predicate may change its factuality, the study has shown that this is only partially true. From a quantitative point of view, the relevance is low, since out of the 28 VPs in our study, only six auxiliaries actually modify factuality.

As for the second hypothesis, we expected the assignment of a default value for each subclass of VPs to be possible. The corpus analysis proved that this was the case in most sentences, taking into account that tense plays a part in the pre-assignment of factual values. A total of 48 predictions were formulated (Table 1). Only 11 of the labels that were predicted (22.92%) fail to meet the hypothetical values. In fact, 10 out of these 11 unexpected labels correspond to the two conditional tenses, which have proven to be the most complex ones as regards the prediction of a value. That is, the NON-APPLICABLE value that we proposed for these two tenses was clearly an oversimplification. Even in some VPs (terminative, progressive, continuative or resultative) none of the sentences with these tenses in the study was tagged as NON-APPLICABLE. It can therefore be concluded that no by-default factual value can be assigned to these two tenses.

Another non-predicted label is that assigned to the past of prospective VPs, where the expected COUNTERFACT only fits 51.56 percent of the examples analyzed. Almost half of the examples show a different value. A more in-depth analysis shows that in *tardar en* INF 'take time INF' none of the 20 sentences in the past tense were tagged as COUNTERFACT. Furthermore, in the present tense, this subclass was predicted to express an UNDERSPECIFIED

value, which was never the case, as shown by the data. Our proposal is to set apart the case of *tardar en* INF from the rest of prospective VPs, since it behaves differently from a factual perspective.

Our third hypothesis, namely, that tense plays a more prominent role in determining the factual value, is confirmed in the data analyzed. For example, in present and past tenses, the factual value of the auxiliary is only relevant for two phase VPs (egressive and prospective). For the other phase subclasses (inceptive, ingressive and terminative), tense determines factuality. As for the future tense, only four out of 126 sentences have not been tagged as UNDERSPECIFIED, and they correspond to inceptive and prospective VPs. As regards the conditional, the nature of this tense allows different options with respect to factuality in all VPs, except for inchoative.

On the basis of the corpus data, we can conclude that it is feasible to create general rules to automate the annotation for the majority of the aspectual VPs studied. In present and past tenses, the most common factual value is FACT for affirmative sentences (if polarity was negative the factual value would be COUNTERFACT), whereas in future tenses the most common factual value is UNDERSPECIFIED. Specific rules can be suggested for the egressive VPs *cesar de* INF 'cease INF/GER', *dejar de* INF 'stop GER', and the prospective VPs *estar a punto de* INF 'be about to INF', *ir a* INF 'be going to INF' and *estar por* INF 'be about to INF' in the present and the past. In the case of *tardar en* INF 'take time INF', this is also true for the simple present and the past imperfective, but only when it is not a habitual interpretation.

As regards verb tenses, very few instances do not follow the expected behavior. This is the case of inceptive VPs that, occasionally, show a different value in the present (other than FACT) and future (other than UNDERSPECIFIED). Also, some cases of prospective VPs in the present and in some past tenses display a different value (other than the expected UNDERSPECIFIED and COUNTERFACT, respectively).

In summary, it can be concluded that the combination of the factual status of aspectual VPs and verb tenses allows the prediction of verbal behavior and the implementation of rules based on this information. Nevertheless, it should also be acknowledged that the present study was carried with a limited number of examples for each tense, so it may be advisable to expand the study with the analysis of more examples of use of aspectual periphrases.

REFERENCES

Alonso, Laura, Irene Castellón, Hortènsia Curell, Ana Fernández-Montraveta, Sonia Oliver and Glòria Vázquez. 2018. Proyecto TAGFACT: Del texto al conocimiento: Factualidad y grados de certeza en español. *Procesamiento del Lenguaje Natural* 61: 151–154.

Barrios, Leyre. 2018. *La Factualidad en las Oraciones Adversativas, Concesivas y Condicionales en Español: El Papel de los Tiempos Verbales en la Anotación Automática de Corpus*. Lleida: University of Lleida Dissertation.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Diab, Mona T., Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran and Weiwei Guo. 2009. Committed belief annotation and tagging. In Manfred Stede, Chu-Ren Huang, Nancy Ide and Adam Meyers eds. *Proceedings of the Third Linguistic Annotation Workshop*. Singapore: Suntec, 68–73.

Fábregas, Antonio. 2019. Periphrases in Spanish: Properties, diagnostics and research questions. *Borealis: An International Journal of Hispanic Linguistics* 8/2: 1–82.

Fernández de Castro, Félix. 1999. *Las Perífrasis Verbales en el Español Actual*. Madrid: Editorial Gredos.

Fernández-Montraveta, Ana and Glòria Vázquez. 2017. *Las Construcciones con 'Se' en Español*. Madrid: Arco Libros.

García Fernández, Luis. 2006. *Diccionario de Perífrasis Verbales*. Madrid: Editorial Gredos.

Havu, Jukka. 1997. *La Constitución Temporal del Sintagma Verbal en el Español Moderno*. Helsinki: Academia Scientiarum Fennica.

Mendikoetxea, Amaya. 1999. Construcciones con 'se': Medias, pasivas e impersonales. In Ignacio Bosque and Violeta Demonte eds. *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa-Calpe, 1631–1721.

Minard, Anne-Lyse, Manuela Speranza, Rubén Urizar, Begoña Altuna, Marike van Erp, Anneleen Schoen and Chantal van Son. 2016. MEANTIME, the *NewsReader Multilingual Event and Time Corpus*. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odijk and Stelios Piperidis eds. *Proceedings of the 10th Language Resources and Evaluation Conference*. Portorož: European Language Resources Association, 4417–4422.

Narita, Kazuya, Junta Mizuno and Kentaro Inui. 2013. A lexicon-based investigation of research issues in Japanese factuality analysis. In Ruslan Mitkov and Jong C. Park eds. *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Nagoya: Asian Federation of Natural Language Processing, 587–595.

Olbertz, Hella. 1998. *Verbal Periphrases in a Functional Grammar of Spanish*. Berlin: Mouton de Gruyter.

Portner, Paul. 2009. *Modality*. Oxford: Oxford University Press.

Real Academia Española. 2009. *Nueva Gramática de la Lengua Española*. Madrid: Espasa Calpe.

Ross, Alexis and Ellie Pavlick. 2019. How well do NLI models capture verb veridicality? In Kentaro Inui, Jing Jiang, Vincent Ng and Xiajun Wan eds. *Proceedings of the 9th International Joint Conference on Natural Language Processing*. Hong Kong: Association for Computational Linguistics, 2230–2240.

Rudinger, Rachel, Aaron Steven White and Benjamin Van Durme. 2018. Neural models of factuality. In Marylin Walker, Heng Ji and Amanda Stent eds. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational*

Linguistics: Human Language Technologies. New Orleans: Association for Computational Linguistic, 731–744.

Saurí, Roser. 2008. *A Factuality Profiler for Eventualities in Text*. Waltham: Brandeis University dissertation.

Saurí, Roser and James Pustejovsky. 2009. *FactBank*: A corpus annotated with event factuality. *Language Resources and Evaluation* 43/3: 227–268.

Soni, Sandeep, Tanushree Mitra, Eric Gilbert and Jacob Eisenstein. 2014. Modeling factuality judgments in social media text. In Kristina Toutanova and Hua Wu eds. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Baltimore: Association for Computational Linguistics, 415–420.

Topor, Mihaela. 2011. *Perífrasis Verbales del Español y Rumano: Un Estudio Contrastivo*. Lleida: University of Lleida dissertation.

Troya Déniz, Magnolia. 2007. Frecuencia de los tiempos verbales de indicativo y subjuntivo en la norma culta de España y América. *Revista de Filología de la Universidad de La Laguna* 25: 589–602.

Vázquez, Glòria and Ana Fernández-Montraveta. 2020. Annotating factuality in the TAGFACT corpus. In Miguel Fuster-Márquez, Carmen Gregori-Signes and José Santaemilia Ruiz eds. *Multiperspectives in Analysis and Corpus Design.* Granada: Comares, 115–125.

Wonsever, Dina, Marisa Malcuori and Aila Rosá. 2008. SIBILA: Esquema de Anotación de Eventos. Reporte técnico RT 08-11. Instituto de Computación. Universidad de la República Montevideo. https://www.colibri.udelar.edu.uy/jspui/bitstream/20.500.12008/3419/1/TR0811.pdf

Wonsever, Dina, Aiala Rosá and Marisa Malcuori. 2016. Factuality annotation and learning in Spanish texts. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odijk and Stelios Piperidis eds. *Proceedings of the Tenth International Conference on Language Resources and Evaluation*. Portorož: European Language Resources Association, 2076–2080.

Corresponding author
Glòria Vázquez
University of Lleida
Department of Foreign Languages and Literatures
Plaza Víctor Siurana 1
25003
Lleida
Spain
Email: gloria.vazquez@udl.cat

APPENDIX 1: TRANSLATION OF SPANISH VPS INTO ENGLISH[12]

Spanish VP	Subclass	Translation of the verb phrase into English
Acabar de INF 1	Terminative	'Finish INF'
Acabar de INF 2	Resultative	'VERB in perfect tense + just'
Acostumbrar (*a*) INF	Habitual	'Usually VERB /used to'
Andar GER	Progressive	'VERB in progressive tense form + always'
Cesar de INF	Egressive	'Cease INF/GER'
Coger y VERB	Telic	'Up and VERB'
Comenzar a INF	Inceptive	'Begin/start INF/GER'
Continuar GER	Continuous	'Keep GER'
Dejar de INF	Egressive	'Stop GER'
Echar a INF	Inceptive	'Begin/start INF/GER'
Empezar a INF	Inceptive	'Begin/start INF/GER'
Estar GER	Progressive	'VERB in progressive tense form'
Estar a punto de INF	Prospective	'Be about to INF / be on the point of GER'
Estar por INF	Prospective	'Be about to INF / be on the point of GER'
Ir GER	Continuous	'VERB in progressive tense form'
Ir a INF	Prospective	'Be going to INF'
Ir y VERB	Telic	'Up and VERB'
Llevar GER	Continuative	'VERB in perfect progressive tense form'
Parar de INF	Egressive	'Stop GER'
Pasar a INF	Ingressive	'Move on INF'
Ponerse a INF	Inceptive	'Begin/start INF/GER'
Quedarse GER	Inchoative	'VERB in progressive tense form'
Seguir GER	Continuous	'Keep GER'
Soler INF	Habitual	'Usually VERB / used to'
Tardar en INF	Prospective	'Take time INF'
Terminar de INF	Terminative	'Finish INF'
Venir GER	Continuative	'VERB in perfect progressive tense form'
Venir PART	Resultative	'VERB in passive voice'

---

[12] These translations are actually glosses. When contextualized in an example, a more idiomatic expression might have been chosen.