OPEN ACCESS

Ingressive verbal periphrases in Spanish and Portuguese[1]

Dana Kratochvílová — Jaroslava Jindrová (Prague)

ABSTRACT
The article analyses Spanish and Portuguese verbal periphrases that express the beginning of a process (ingressive manner of action). It analyses stylistically marked and stylistically unmarked ingressive constructions in order to determine similarities and differences between the two languages. The study is primarily based on the comparable web corpora *Aranea* which makes it possible to see the frequency of use of Spanish and Portuguese periphrases in contemporary language, compare them and determine semantic restrictions for auxiliated verbs in each construction. As a result of this corpus analysis, it is possible to trace the main systemic correspondences between Spanish and Portuguese ingressive verbal periphrases and see the relationship between the original meaning of a concrete semi-auxiliary verb and the group of infinitives that can appear in a periphrasis introduced by that verb.

KEYWORDS
verbal periphrasis, manner of action, aspect, ingressive, inchoative

1. INTRODUCTION

In this paper, we concentrate on Spanish and Portuguese verbal periphrases that express the beginning of a process. The aim is to show the similarities and differences that can be found between these two languages. The study is based on material obtained from language corpora wherein we analyse the most frequent Spanish ingressive constructions and, subsequently, we seek their closest Portuguese respondents. We take as a starting point the original meaning of a selected semi-auxiliary and, subsequently, analyse the frequency of use of concrete periphrases in both languages and their combinatorial possibilities (i.e. the characteristic of infinitives that appear in them). In this way we verify whether (or to which extent) the semantic characteristic of a semi-auxiliary verb in its original meaning has an influence on the meaning of the resulting verbal periphrasis and on the selection of auxiliated verbs. Since we work mainly with comparable corpora, our analysis also serves to analyse the respective frequencies of particular ingressive constructions in both languages.

2. INGRESSIVE VERBAL PERIPHRASES IN SPANISH AND PORTUGUESE

Both Spanish and Portuguese have a wide range of periphrastic constructions expressing the beginning of a process that are commonly referred to as inchoative or in-

1 This study was supported by the Charles University project Progres Q10, *Language in the shiftings of time, space, and culture*.

gressive verbal periphrases. The interpretation of the terms *inchoative* and *ingressive* is not always unanimous (see, for different approaches, Yllera, 1980; Gómez Torrego, 1999; Fernández de Castro, 1999; Pešková, 2006; Barroso, 1994; Castilho, 1966; Bechara, 2001). We will be using the terminology proposed by Zavadil and Čermák (2010) and the term ingressive manner of action which "characterises the process from the perspective of its beginning"[2] (316). Zavadil and Čermák (2010) make the following further distinctions:

1. imminent ingressiveness (process that is imminently expected)
2. dispositive ingressiveness (process that is imminently intended)
3. initial ingressiveness (process that is beginning)
4. inceptive ingressiveness (process that is initial in a series of other processes) (cf. 316–319).

Leaving aside subtypes 1, 2 and 4, our aim is to discuss the nature of initial ingressiveness, i.e. beginning of a process itself. For the sake of simplicity, the analysed constructions will be called ingressive periphrases.

In both languages, the principal construction for expressing initial ingressiveness is: (semi)auxiliary verb + preposition *a* + infinitive. In the case of Spanish, we can say that periphrastic constructions with the infinitive are typical of expressing the initial phase of a process, periphrases with the gerund are used for presenting a process from the perspective of its duration, the final phases are expressed by constructions with the infinitive or participle. This general assertion is based on the progressive character of the Spanish gerund which stands in opposition to the aspectually neutral infinitive and resultative participle.[3]

In Portuguese, the situation is less clear. Initial ingressiveness is expressed mainly through periphrases with the infinitive; nevertheless, in the Brazilian variety of Portuguese, some of them have synonymous constructions with the gerund (see Castilho, 1966).[4] This can be seen as a manifestation of the suppression of the durative features of the gerund and its approximation to the infinitive. Also, some Portuguese periphrastic constructions with the infinitive manifest clear features of durativity, for example, one of the most common ones, *estar a* + infinitive, that expresses a process which is taking place right now:

2 „charakterizuje pojmenovaný proces z hlediska jeho počátku". Translation by the authors.

3 Here, we can trace a relationship between the category of manner of action and the category of aspect that are both related to the study of verbal periphrases. For different approaches to these categories in Czech and Romance tradition see, for example, RAE, 2009; Zavadil-Čermák, 2010; Pešková, 2006; Hricsina, 2006; Jindrová, 2015.

4 These constructions will not be analysed in our study due to their diatopical restriction and the fact that they often cannot be sufficiently separated from constructions that express inceptive ingressivity, i.e. the beginning of an action that is first in a series of actions (for this interpretation of inceptivity, see Zavadil and Čermák, 2010, 319): *Começar falando = Começar a falar* (*To begin to talk*) or *Começar por falar* (*To begin talking about…*).

(1) **Estou a comer**
Be: 1st sg., pres., indicative; *to* (preposition); *eat*: infinitive
I am eating

The same meaning is expressed in Spanish by a construction with the gerund:

(2) **Estoy comiendo**
Be: 1st sg., pres., indicative; *eat*: gerund
I am eating

Similarly, a sentence like (3) is ambiguous in Portuguese:

(3) **Paula entrou a cantar**
Paula; *enter*: 3rd sg., past tense, indicative;
to (preposition); *sing*: infinitive

It could mean either *Paula entered the room, singing*, or *Paula started to sing* (*entrar* being here a semi-auxiliary in an ingressive periphrasis). Such an ambiguity would, however, be impossible in Spanish, where the infinitive cannot express a simultaneous process. That would have to be expressed by the gerund (5):

(4) **Paula entró a cantar**
Paula; *enter*: 3rd sg.; past tense, indicative;
to (preposition); *sing*: infinitive
Paula started to sing

(5) **Paula entró cantando**
Paula; *enter*: 3rd sg., past tense, indicative; *sing*: gerund
Paula entered the room, singing.

Apart from these differences, the lists of the most commonly used Spanish and Portuguese ingressive periphrases seem to be quite similar. In both languages, we can find neutral constructions with a verb that means *to begin*. They have no combinatorial restrictions and are used widely to express the beginning of any process that does not exclude separation into several phases. Secondly, we can find a wide range of constructions with semi-auxiliary verbs that have lost, partially or completely, their original meaning and, together with the preposition *a* and a lexical verb in the infinitive form, create periphrastic constructions that are stylistically marked and add a specific quality to the original notion of the beginning of a process (suddenness, abruptness, inappropriateness, previous retention, etc.). These stylistically marked constructions also present several combinatorial limitations and appear only with certain types of verbs. In the next section, we analyse both types of ingressive periphrases.

3. STYLISTICALLY UNMARKED CONSTRUCTIONS

3.1 COMENZAR / COMEÇAR / EMPEZAR / PRINCIPIAR / INICIAR A + INFINITIVE

Spanish has several verbs that mean to begin or to start. The most frequent ones are *comenzar* (which has a total of 84 543 occurrences in the corpus *CORPES XXI*[5]), and *empezar* (83 116 occurrences). Less frequent is the synonym *iniciar* (41 470 occurrences); rarely used is the verb *principiar* (334 occurrences). All these verbs can, theoretically, form an ingressive construction that marks the beginning of a process (start to do something). Nevertheless, despite their semantic closeness, these verbs do not behave in the same way regarding the possibilities to form an ingressive construction. Using the corpus *CORPES XXI*, we have analysed the frequency of their occurrence outside and inside ingressive constructions (verb + a + infinitive). For each one of them, we used analytical forms of queries, given below.

Outside a periphrastic construction (Figure 1):

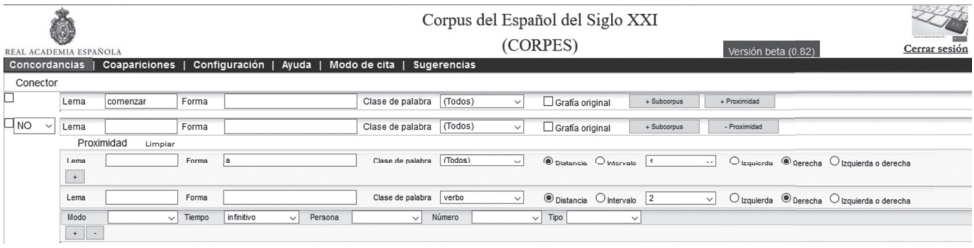

FIGURE 1. Corpus query (*CORPES XXI*)

As a part of an ingressive verbal periphrasis (Figure 2):

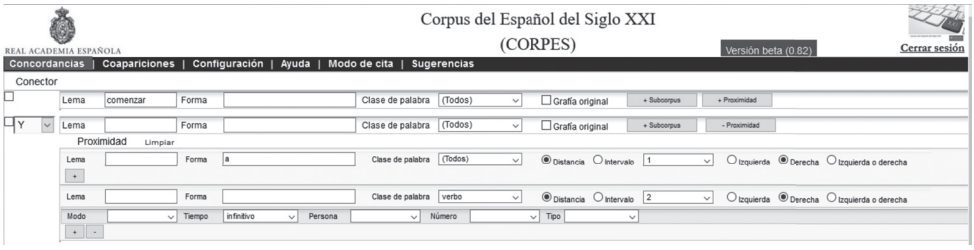

FIGURE 2. Corpus query (*CORPES XXI*)

The results are summarised in Table 1.

5 A reference corpus compiled by Real Academia Española, accessible from: http://web.frl.es/CORPES/org/publico/pages/consulta/entradaCompleja.view. In its version 0.82, it consists of 222 080 documents, the oldest from 2001, which altogether contain over 215 million tokens (cf. RAE, 2015a).

	Autonomous	Inside an ingressive periphrastic construction
Comenzar	38 969 (46,1 %)	45 574 (53,9 %)
Empezar	30 037 (36,1 %)	53 079 (63,9 %)
Iniciar[6]	41 439 (99,9 %)	31 (0,1 %)
Principiar	250 (74,9 %)	84 (25,1 %)

TABLE 1. Distribution of *comenzar*, *empezar*, *iniciar* and *principiar* outside and inside an ingressive periphrastic construction: the corpus ***CORPES XXI***

In Portuguese, there are only three verbs that generally express the beginning of an action: *começar*, *iniciar* and *principiar*. Unlike in Spanish, the verb *empeçar* is very rarely used. We have subjected them to a similar analysis in the corpus *CETEMPúblico*[7]. In order to determine the frequency of their use, we employed the following queries:

[lema="começar"]
[lema="iniciar"]
[lema="principiar"].

For their periphrastic uses, we adopted the following query which was analytical for the three of them:

[lema="começar"][word="a"][temcagr="INF"].

The results we obtained are similar to those in Spanish. The most frequently used verb is *começar* (19 781 occurrences), the verb *iniciar* is used less often (5 390 occurrences), *principiar* is seldom used (145 occurrences). Their use as fully lexical verbs and as semi-auxiliaries is summarised in Table 2.

	Autonomous	Inside an ingressive periphrastic construction
Começar	12 334 (62,6 %)	7 384 (37,4 %)
Iniciar[8]	5 390 (100 %)	0 (0 %)
Principiar[9]	134 (92,4 %)	11 (7,6 %)

TABLE 2. Distribution of *começar*, *iniciar* and *principiar* outside and inside an ingressive periphrastic construction: the corpus *CETEMPúblico*

6 When formulating the query with a lemma, the results also include the occurrences of the form *inicio* functioning as a noun; these were removed manually.

7 A large corpus that consists of 191 687 833 words. All texts come from the Portuguese newspaper *Público*; it reflects, therefore, mainly the European variety of Portuguese. The texts were written between 1991 and 1998 (cf. Linguateca, 2007; Santos — Costa, 2005).

8 We have manually removed the occurrences of the form *inicio* functioning as a noun (despite the absence of *acento ortográfico*) that appeared in the original concordance.

9 We have manually excluded the occurrences of the form *principio* functioning as a noun (despite the absence of *acento ortográfico*) that appeared in the original concordance.

In order to make an extended comparison between Spanish and Portuguese, we performed the same analysis in the corpora *Araneum Hispanicum Maius* and *Araneum Portugallicum Maius*, which form part of the family of comparable web corpora compiled by Vladimír Benko (see, Benko, 2014). Its Spanish part (accessible since 2013) consists of 1 200 000 609 tokens, the Portuguese part (in its April 2015 version) consists of 1 200 006 068 tokens. We used the following queries (analogical for every verb):

For Spanish:
[lemma="comenzar"&!word="[Cc]omenzad."]* and
[lemma="haber."][word="comenzado"]* for all uses of each verb and
[lemma="comenzar"][word="a"][tag="VLinf"&!word="partir"] for verbal periphrases.

For Portuguese:
[lemma="começar"&!word="[Cc]omeçad."]* and
[lemma="ter."][word="começado"]* for all uses of each verb and
[lemma="começar"][word="a"][tag="V."&!word="partir"]* for verbal periphrases.

Whenever necessary, the results were adjusted manually. We will mention only those adjustments that were more frequent or considered especially relevant.

Results are summarised in Tables 3 and 4.

	Autonomous	**Inside an ingressive periphrastic construction**
Empezar	183 316 (57,6%)	135 069 (42,4%)
Comenzar	212 461 (66,4%)	107 370 (33,6%)
Iniciar[10]	137 039 (99,8%)	301 (0,2%)
Principiar[11]	678 (85,6%)	114 (14,4%)

TABLE 3. Distribution of *empezar*, *comenzar*, *iniciar* and *principiar* outside and inside an ingressive periphrastic construction: the corpus *Araneum Hispanicum Maius*

In both languages we can observe large differences among the analysed verbs as to the possibility of their use as semi-auxiliaries. The Spanish verbs *comenzar* and *empezar* are largely used in ingressive constructions. The percentage of their use as semi-auxiliaries is different in the *CORPES XXI* and the *Araneum*, which might be attributed to different types of texts that constitute these corpora; nevertheless, we can say that their distribution between fully lexical verbs and semi-auxiliaries is relatively balanced.

10 We manually excluded 31 occurrences of *INICIO* (a noun, usually, referring to the *START* button).

11 When using the query *[lemma="principiar"]*, the corpus search found only the infinitive forms, therefore, we used a different type of query here: *[word="[Pp]rincip.*"][word="a"][tag="VLinf"&!word="partir"]* for the verbal periphrases and *[word="[Pp]rincipi.*"]* for all appearances of this verb and, consequently, we manually excluded all the occurrences that were not of interest.

	Autonomous	Inside an ingressive periphrastic construction
Começar	309 111 (57,6%)	227 651 (42,4%)
Iniciar[12]	138 444 (99,9%)	196 (0,1%)
Principiar[13]	841 (100%)	0 (0%)

TABLE 4. Distribution of *começar*, *iniciar* and *principiar* outside and inside an ingressive periphrastic construction: the corpus *Araneum Portugallicum Maius*

Similarly, in Portuguese the most frequent semi-auxiliary is the verb *começar* with similar distributions outside and inside an ingressive periphrastic construction (it is interesting to note that in the *Aranea* corpora the distributions of the Portuguese *começar* and the Spanish *empezar* are exactly the same).

The use of the verb *principiar* in an ingressive construction is not very frequent in either language (even though the percentage is still greater in Spanish). The fact that we have not found a single use of this construction in the *Araneum Portugallicum Maius* can be seen as proof that it is to a large extent limited to literary style.

The last verb from the set, *iniciar*, is perhaps the most interesting one. Even though it is frequently used in both languages as a fully lexical verb, its occurrence in an ingressive construction is very rare. For Portuguese, there is no mention of such a construction in Barroso (1994), Bechara (2001), Mateus et al. (2003) or Drzazgowska (2011), and we have not found a single case of an ingressive construction introduced by *iniciar* in the corpus *CETEMPúblico,* which includes only texts from the Portuguese newspaper *Público* (excluding, therefore, the Brazilian variety of Portuguese). On the other hand, we have been able to find 196 occurrences of *iniciar* + *a* + infinitive in the *Araneum Portugallicum Maius,* which includes texts from Portuguese, Brazilian and Angolan web pages (this, however, still constitutes a very low percentage, compared to 138 444 uses of this verb outside an ingressive construction). From this set, 176 occurrences were from Brazilian web pages, 2 from Angola, in 12 cases it was impossible to determine the nationality of the author (web pages that included posts from both Brazilian and Portuguese authors), only 6 occurrences were found on Portuguese websites.

Based on these results, we can assume that in Portuguese the ingressive periphrasis *iniciar a* + infinitive is strongly diatopically restricted and the use of *iniciar* as a semi-auxiliary is very limited.

The Spanish counterpart of this periphrasis is usually not mentioned in the literature either (it is not listed in the extensive dictionary of verbal periphrases by García Fernández et al., 2006, similarly, it is not mentioned by Zavadil and Čermák, 2010; RAE, 2009; Fernández de Castro, 1999; or Gómez Torrego, 1988). Nevertheless,

12 When using the query *[lemma="iniciar"]*, the results included also the forms *inicio*, *Inicio* and *INICIO* in the function of a noun (despite the absence of *acento ortográfico*), these were excluded manually.

13 When using the query *[lemma="principiar"]*, the results included also the forms *principio*, *Principio* in the function of a noun (despite the absence of *acento ortográfico*), these were excluded manually. We also excluded manually other interferences such as *principal-mente*.

we have been able to find 31 occurrences of this construction in the *CORPES XXI* (2 in European Spanish, 29 in Latin America). Consequently, we analysed this construction in the corpus *CREA*[14] (using the same form of query as for the *CORPES XXI*) and found only 6 cases, the oldest one from 1996, all from Latin America. The *Araneum Hispanicum Maius* yielded 301 occurrences of this construction, both from Latin-American and Spanish sites (although the Latin-American instances prevail). This might be seen as evidence of the increasing use of this construction in the last few years and of the preference for its use in Latin America (even though the diatopic restrictions are less strong than with its Portuguese equivalent).

Despite the occurrences of ingressive periphrases with *iniciar* presented so far, the use of *iniciar* as a semi-periphrastic verb is very limited in comparison with *empezar* (Sp.), *comenzar* (Sp.) and *começar* (Pt.). This leads us to the conclusion that the semantic and syntactic criteria for distinguishing between periphrastic and non-periphrastic constructions are not the only ones that should be taken into account when studying verbal periphrases. According to the criterion of semantic emptiness of the auxiliary word, the previously analysed constructions should not form part of the inventory of periphrastic constructions, for the verbs *empezar*, *começar* etc. do not lose their original meaning *to begin* (for an extended discussion of the semantic and syntactic criteria, see Pešková, 2006, 40–46). On the other hand, our analysis proves that both in Spanish and Portuguese there is a clear tendency to prefer concrete verbs of beginning as semi-auxiliaries, while the use of others, despite their synonymy, is strongly limited. The grammaticalization of a semi-auxiliary verb is, therefore, conditioned not only by the syntactic and semantic criteria, but also by the frequency of use and the level of stabilization of a concrete ingressive construction in the speaker's mind.

4. STYLISTICALLY MARKED CONSTRUCTIONS

Spanish and Portuguese offer a wide range of ingressive constructions that are, in some way, stylistically marked and bring additional semantic features to the original meaning of a process in its initial state. We will analyse the most frequent ones and compare the situation in Spanish and Portuguese. Since these constructions are not used as often as the previous ones and, therefore, there are fewer occurrences in the corpora, from now on, the analysis will concentrate on the comparable web corpora from family *Aranea*, which provide more examples both for Spanish and Portuguese and, due to their similar size and time of publication, they can very well reflect the possible differences in the frequency of use of a particular construction in Spanish and Portuguese.

The starting point will be the most frequent Spanish ingressive constructions with verbs that show higher levels of semantic emptiness: *ponerse a*, *romper a*, *echar a* and

14 A reference corpus created by RAE which includes over 160 million words. The texts are from the 1975–2004 period and come from all Spanish-speaking countries (cf. RAE 2015b). Its annotated version is accessible from: http://web.frl.es/CREA/.

echarse a. These constructions are frequently used and they are widely discussed in the relevant literature (see, for example, RAE, 2009; García Fernández, 2006; Fernández de Castro, 1999; Gómez Torrego, 1988; 1999). The semantic emptiness of these semi-auxiliaries will also help us to analyse with greater precision the results obtained from the corpus, since it is usually easy to separate an ingressive construction with *ponerse*, *romper* or *echar(se)* from other constructions "verb + *a* + infinitive" that are not ingressive verbal periphrases. With less frequent constructions such as those with verbs like *meterse* (*to get into something*), *lanzarse* (*to jump on something*), *entrar* (*to enter*), *largarse* (*to go away*), or *arrojarse* (*to plunge into something*), the boundaries between an ingressive construction and a construction in which these verbs retain their original full meaning are less clear.

The comparison between Spanish and Portuguese will be based on the semantic closeness of the semi-auxiliary verbs in their original meaning and, if possible, also on the combinatorial criteria. This approach allows us to analyse the impact of the original meaning of a semi-auxiliary on the meaning of a verbal periphrasis and on its combinatorial possibilities.

Unless indicated otherwise, the corpus queries are analogical for all the analysed verbs.

For Spanish:
[lemma="poner."&!word="[P]uest.*"][word="a"][tag="VLinf"]* and
[lemma="haber."][word="puesto"][word="a"][tag="VLinf"]* for verbal periphrases.
[lemma="poner."&!word="[Pp]uest.*"]* and
[lemma="haber."][word="puesto"]* for all uses of a concrete verb.

For Portuguese:
[lemma="pôr."&!word="[Pp]ost.*"][word="a"][tag="V.*"]* and
[lemma="ter."][word="posto"][word="a"][tag="V.*"]* for verbal periphrases.
[lemma="pôr."&!word="[Pp]ost.*"]* and *[lemma="ter.*"][word="posto"]* for all uses of a concrete verb.

The results we obtained sometimes had to be adjusted manually in order to exclude constructions that are not of interest to us. We shall not enumerate all the excluded corpus errors, for their list would be excessively long; again, we will mention only those excluded cases we consider to be frequent or somehow important.

4.1 *PONERSE* / *PÔR-SE A* + INFINITIVE

In Spanish, the most frequent stylistically marked ingressive periphrastic construction is formed with the verb *ponerse* (= *to put (on)*), its Portuguese counterpart being the verb *pôr-se*.

This periphrasis focalizes the beginning of a process with the notion of intentionality of the subject. García Fernández et al. (2006) also mention the combination with inanimate subjects, where this periphrasis points out a sudden beginning of a process (cf. 218). Nevertheless, the combination of this construction with verbs such as

llover (Sp.) / *chover* (Pt.) (*to rain*) is relatively limited in both languages. We have found only 24 occurrences of the Spanish construction *ponerse a llover* in the *CORPES XXI*, 112 in the *Araneum Hispanicum Maius* and just one case of *pôr-se a chover* in the *Araneum Portugallicum Maius* (no occurrences in the *CETEMPúblico*).

Despite the semantic closeness, this construction is distinctively more frequent in Spanish than in Portuguese, as shown in Table 5:

	Autonomous	**Inside an ingressive periphrastic construction**
Ponerse	679 856 (95,6%)	31 433[15] (4,4%)
Pôr-se	62 232 (98,3%)	1 098[16] (1,7%)

TABLE 5. Distribution of *ponerse* and *pôr-se* outside and inside an ingressive periphrastic construction: the corpora *Araneum Hispanicum Maius* and *Araneum Portugallicum Maius*

Even though the distributions of *ponerse* / *pôr-se* outside and inside a periphrastic construction are quite similar, the results from comparable corpora strongly indicate a more frequent use of the verb *poner(se)* in Spanish which, consequently, translates into a greater use of the periphrastic construction with *ponerse*. Despite the difference in the frequency of use, the periphrasis combines with similar infinitives in both languages. The ten most frequent ones were:

Ponerse a

pensar (*to think*): 2 638, *trabajar* (*to work*): 2 612, *hacer* (*to do*, *to make*): 1 368, *buscar* (*to search*): 1 123, *llorar* (*to cry*): 1 047, *escribir* (*to write*): 1 035, *hablar* (*to talk*): 878, *leer* (*to read*): 746, *jugar* (*to play*): 600, *bailar* (*to dance*): 496.

Pôr-se a

pensar (*to think*): 96, *trabalhar* (*to work*): 59, *correr* (*to run*): 36, *andar* (*to walk*): 36, *falar* (*to talk*): 33, *escrever* (*to write*): 32, *chorar* (*to cry*): 27, *fazer* (*to do*, *to make*): 25, *ler* (*to read*): 21, *refletir* (*to meditate about something*): 20.

4.2 *ROMPER* / *DESATAR A* + INFINITIVE

In Spanish, another relatively frequent ingressive construction is formed with the verb *romper* (*to break*). According to the original meaning of this verb, this periphra-

15 We manually excluded frequent constructions (usually related to cooking) that have causative meaning (*poner a hervir*, *poner a cocer*, *poner a calentar*, *poner a funcionar* etc.). We also excluded the frequent construction *poner a parir* (a colloquial idiom meaning *to irritate* or *to annoy*).

16 We manually excluded constructions *pôr a perder* (an idiom used especially in Brazilian Portuguese that means *to waste something*) and *pôr a funcionar* with the causative meaning *to make something work*. Causative constructions related to cooking (*pôr a cozer*, *pôr a ferver* etc.) were also excluded.

sis expresses an abrupt beginning of a process, often accompanied by the notion of a previous hold up:

(6) **María rompió a llorar**
María, break — 3rd sg., past tense, indicative;
to (preposition); *cry* — infinitive
María broke into tears

The list of verbs that can appear in this construction is limited. García Fernández et al. (2006) mention the combination of *romper a* + infinitive with verbs that express an emotional response (*to cry, to laugh, to shout*...), verbs of physical activity (*to dance, to run, to fly*...), verbs of interpretation (*to sing, to tell, to recite*...), verbs that express a change of state (*to cook, to boil, to grow*...) and verbs associated with meteorological phenomena (*to rain, to thunder*...) (cf. 231–232).

Portuguese has three main counterparts to the Spanish verb *romper*: *quebrar*, *partir* and *romper*. As a fully lexical verb, *quebrar* is frequently used (35 654 occurrences in the *Araneum Portugallicum Maius*); nevertheless, it does not form an ingressive periphrasis (we did not find a single occurrence in the *CETEMPúblico* or *Araneum Portugallicum Maius*).

The usage of the verb *romper* (Pt.) is more restricted (17 266 occurrences in the *Araneum Portugallicum Maius*). It can form an ingressive construction with a meaning similar to that of the Spanish *romper a* + infinitive. However, based on the corpus data, this periphrasis is not very frequent in contemporary Portuguese.

The verb *partir* has two meanings (*to leave* and *to break*), therefore the frequency of its use outside a periphrastic construction is not relevant to us. Based on the results from the *Araneum Portugallicum Maius*, its occurrence as a semi-auxiliary in an ingressive construction is theoretically possible, yet also very rare. We found only 8 occurrences of this construction, the auxiliated verb was always *rir* (*to laugh*).

The meaning of an abrupt beginning with the notion of a previous hold up is more often expressed in Portuguese by an ingressive construction with the verb *desatar* (*to untie*) which lacks a direct Spanish counterpart (although relatively highly frequent as a fully lexical verb, the Spanish verb *desatar* does not form an ingressive construction).

A concrete comparison is provided in Table 6.

	Autonomous	**Inside an ingressive periphrastic construction**
Romper (Sp.)	68 623 (98,8%)	863 (1,2%)
Romper (Pt.)	17 262 (>99,9%)	4 (<0,1%)
Partir (Pt.)	???	8
Desatar (Sp.)	10 092 (100%)	0 (0%)
Desatar (Pt.)	1 215 (85,1%)	212 (14,9%)

TABLE 6. Distribution of *romper*, *partir* (Pt.) and *desatar* outside and inside an ingressive periphrastic construction: the corpora *Araneum Hispanicum Maius* and *Araneum Portugallicum Maius*

The similarity between the Spanish *romper a* + infinitive and *desatar a* + infinitive is given by the semantic closeness between *romper* (to break something that was originally in one piece) and *desatar* (to untie something that was originally joined together). This original meaning of the semi-periphrastic verbs that has not disappeared completely from the ingressive construction also has a similar influence on the selection of auxiliated verbs (infinitives) that appear in this construction. The first five most common ones are:

Romper a (Sp.)[17]
llorar (*to cry*): 345, *hervir* (*to boil*): 329, *reír* (*to laugh*): 45, *cantar* (*to sing*): 22, *hablar* (*to speak*): 20.

Desatar a (Pt.)
chorar (*to cry*): 50, *falar* (*to speak*): 24, *rir* (*to laugh*): 22, *correr* (*to run*): 21, *fazer* (*to do*): 9.

In both cases, the verbs *llorar*/*chorar* (*to cry*) and *reír*/*rir* (*to laugh*) are among the three most frequent ones. The results also show the tendency of this construction to combine with verbs of speaking.

On the other hand, according to the corpus, the Portuguese construction does not show a clear tendency towards the combination with infinitives that express a change of state (especially one related to cooking). While the high frequency of *romper a hervir* in Spanish may be influenced by the type of texts that make up our corpus (web pages and blogs where recipes and cooking advice generally appear very often), the total absence of the Portuguese *ferver* (*to boil*) in *desatar a* + infinitive (Pt.) indicates an interesting discrepancy between the two periphrases.

Another difference is related to the combinatory potential of analysed periphrases with verbs of movement. While *desatar a* (Pt.) relatively often combines with *correr* (*to run*) and *fugir* (*to run away* — 8 occurrences), the Spanish construction with *romper a* combined with each of the verbs *correr* (*to run*) and *huir* (*to run away*) only once. This might suggest semantic closeness of the Portuguese construction with another Spanish periphrasis introduced by the verb *soltar* (*to let go*). This construction seems to be closer to the Portuguese one based on the original meaning of the semi-auxiliary; its frequency is, however, much smaller than in the case of *romper a* (Sp.), as shown in Table 7:

	Autonomous	**Inside an ingressive periphrastic construction**
Soltar	23 582 (99,6%)	95 (0,4%)

TABLE 7. Distribution of *soltar* (Sp.) outside and inside an ingressive periphrastic construction: the corpus *Araneum Hispanicum Maius*

17 *Romper a* + infinitive (Pt.) combined only with these four verbs: *aplaudir* (*to applaud*), *correr* (*to run*), *chorar* (*to cry*) and *montar* (*to mount*).

Although it is theoretically possible to form an ingressive construction with the Portuguese verb *soltar*, such a periphrasis is even less frequent than the Spanish one in present-day Portuguese: despite the relatively high total frequency of use of the verb *soltar* (Pt.) (18 447 occurrences), we found only one ingressive construction with this verb in the corpus *Araneum Portugallicum Maius*.

The frequency list of infinitives that appeared in the Spanish construction with the verb *soltar* is relatively close to that of the Portuguese periphrasis with *desatar*. The five most frequent verbs are:

> *llorar* (*to cry*): 22, *hablar* (*to speak*): 11, *andar* (*to walk*, *to go around*): 7, *caminar* (*to walk*): 7, *reír* (*to laugh*): 6.

This clearly reflects the semantic closeness of constructions with *desatar* (Pt.) and those with *soltar* (Sp.) which, however, is not manifested by similar distributions. The Portuguese periphrases with *desatar* are used more often and also seem to substitute, at least partially, for the frequent Spanish constructions with *romper a*.

4.3 *ECHAR(SE) A* / *DEITAR(-SE) A* + INFINITIVE

Ingressive constructions with verbs *echar* (*to throw*) and *echarse* (*to lie down*) are relatively frequent in Spanish. Both verbs can function as semi-auxiliaries in a periphrasis that expresses a sudden beginning of a process, the suddenness being here, however, less intense than in the case of constructions discussed in 4.2.

Periphrases with *echar* and *echarse a* + infinitive are sometimes treated separately (see, for example, García Fernández et al., 2006) because they combine with different infinitives. Gómez Torrego (1988) mentions only *andar* (*to walk*), *correr* (*to run*) and *volar* (*to fly*) for the construction *echar a* + infinitive and *reír* (*to laugh*), *llorar* (*to cry*), *correr* (*to run*), *volar* (*to fly*) and *temblar* (*to tremble*, *to be shaking*) for *echarse a* + infinitive (cf. 112).

In order to find these constructions in the *Araneum* corpus we used the following query: *[lemma="echar.*"][word="a"][tag="VLinf"&!word="perder"]* that directly excluded the construction *echar a perder* which is not an ingressive periphrasis and means *to waste* or *to spoil something*. However, we still had to manually exclude over 700 other occurrences that were not relevant, especially those cases where *echar a* + infinitive formed a causative construction (for example, *echar a volar la imaginación — let go your* imagination) or those where *echar* was not a semi-auxiliary (for example, *echarse a domir — to lie down and go to sleep*). The remaining cases where there is no doubt about the ingressive character of the construction, are listed in Table 8:

	Autonomous	**Inside an ingressive periphrastic construction**
Echar(se)	99 905 (96%)	4 208 (4%)

TABLE 8. Distribution of *echar(se)* outside and inside an ingressive periphrastic construction: the corpus *Araneum Hispanicum Maius*

Since it was impossible to formulate a query that would directly separate the pronominal and non-pronominal variants of this construction, we proceeded manually in order to separate constructions with *echar a* and *echarse a*. As a result, we obtained 2 430 occurrences of *echar a* + infinitive and 1 778 occurrences of *echarse a* + infinitive. The five most frequent infinitives that combined with each one of them are:

Echar a
andar (*to walk*): 1 157, *correr* (*to run*): 837, *volar* (*to fly*): 216, *rodar* (*to roll, to film*): 92, *llorar* (*to cry*): 22.

Echarse a
reír (*to laugh*): 594, *llorar* (*to cry*): 566, *temblar* (*to be shaking*): 315, *correr* (*to run*): 108, *andar* (*to walk*): 91.

The results presented so far confirm different preferences of both constructions regarding the fully semantic infinitive and, also, prove the widespread use of both constructions in present-day Spanish. Their high frequency notably contrasts with the situation in Portuguese.

Based on the criterion of semantic closeness between semi-auxiliaries in their original meaning, the closest Portuguese construction would be *deitar(-se) a* + infinitive. The meaning of the pronominal and that of the non-pronominal form of this verb are analogical to *echar* and *echarse*. The verb even forms the same idiom with the verb *perder* as the Spanish *echar* (*deitar a perder — to waste something*).

As for the use of *deitar(-se)* as a semi-auxiliary in an ingressive periphrasis, Barroso (1988) mentions only the non-pronominal variant *deitar a* + infinitive, which is, according to the author, strongly contextually limited and combines only with verbs *correr* (*to run*) and *fugir* (*to run away*) (cf. 69). The pronominal variant is not mentioned in the extensive list of Portuguese verbal periphrases by Drzazgowska (2011) either.

The preference for verbs of movement also seems to confirm the correspondence between ingressive constructions with *echar* and *deitar* in Spanish and Portuguese. Nevertheless, the corpus analysis shows very limited usage of *deitar(-se) a* + infinitive in contemporary Portuguese, see Table 9[18]:

	Autonomous	**Inside an ingressive periphrastic construction**
Deitar(-se)	11 574 (99,9%)	15 (0,1%)

TABLE 9. Distribution of *deitar(-se)* outside and inside an ingressive periphrastic construction: the corpus *Araneum Portugallicum Maius*

As shown in Table 9, we managed to collect only 15 ingressive periphrastic constructions with *deitar(-se) a*. As with the Spanish constructions with *echar(se)*, we manually analysed them in order to separate the pronominal and non-pronominal forms.

18 Analogically to the analysis of *echar(se) a* + infinitive, we manually excluded all constructions that were not ingressive verbal periphrases.

As a result, we obtained 12 periphrases with *deitar a* and 3 constructions with *deitar-se a*. The infinitives that appeared in those periphrases are listed as follows:

Deitar a
correr (*to run*): 5, *fugir* (*to run away*): 2, *falar* (*to talk*): 2; hapax legomena: *cantar* (*to sing*), *proclamar* (*to declare*), *andar* (*to walk*).

Deitar-se a
adivinhar (*to guess*): 2, *estudar* (*to study*): 1.

Although it is difficult to base any conclusion on such a small number of results, we can say that ingressive constructions with *deitar(-se)* are to a certain extent analogous to those with *echar(se)*: they permit both pronominal and non-pronominal variants (the second one being more frequent), the non-pronominal variant prefers verbs of movement in the position of auxiliated verbs. Once again, however, this semantic closeness does not manifest itself in similar tendencies of usage, for the differences in the frequency distributions of *echar(se) a* and *deitar(-se) a* are surprising.

5. CONCLUSION

The results we have collected so far lead to several conclusions. In the first place, we can say that the frequency of use of stylistically unmarked ingressive constructions is similar in both languages. The results regarding constructions with *empezar*, (*empeçar*), *comenzar*, *começar*, *iniciar* and *principiar* obtained from the comparable *Aranea* corpora are summed up in Table 10.

	Spanish: *Araneum Hispanicum Maius*		Portuguese: *Araneum Portugallicum Maius*	
	Outside	Inside	Outside	Inside
Empezar (Sp.) / *Empeçar* (Pt.)	183 316	135 069	7[19]	0
Comenzar (Sp.) / *Começar* (Pt.)	212 461	107 370	309 111	227 651
Iniciar	137 039	301	138 444	196
Principiar	678	114	841	0
Total	533 494 (68,7%)	242 854 (31,3%)	448 403 (66,3%)	227 847 (33,7%)

TABLE 10. Distribution of *empezar*, *empeçar*, *comenzar*, *começar*, *iniciar* and *principiar* outside and inside an ingressive periphrastic construction: the corpora *Araneum Hispanicum Maius* and *Araneum Portuallicum Maius*

19 It is possible to find a total of 102 occurrences using the query *[lemma="empeçar"]* in the *Araneum Portugallicum Maius*, however, most of these cases are included by mistake in the selection for they are in Galician or Spanish texts.

The great similarity between the two analysed languages was found in the total frequency of neutral ingressive constructions and in the distribution of *empezar*, *comenzar*, *começar*, *iniciar* and *principiar* as semi-auxiliaries / fully lexical verbs. Since the Portuguese *começar* usually stands for the Spanish *empezar* and *comenzar*, replacing both of them when they are fully lexical verbs, it also appears more often as a semi-auxiliary, in a construction corresponding to the Spanish ingressive constructions *empezar a* and *comenzar a* + infinitive. The total frequency of the Spanish ingressive constructions with *empezar* and *comenzar* is similar to the frequency of the Portuguese *começar a* + infinitive (242 439 : 227 651 occurrences or 202,03 : 189,70 in ipm).

Similarly, the verb *iniciar*, despite its relatively frequent use outside an ingressive construction, appears rarely as a semi-auxiliary in both languages.

Ingressive constructions with *principiar* were analysed together with stylistically unmarked ingressive periphrases, for they do not add any special feature like "abruptness" or "suddenness" to the expression of a beginning of a process and they do not display any combinatorial limitations regarding auxiliated verbs. On the other hand, the term stylistic unmarkedness might seem inappropriate here, since the verb *principiar*, in itself, is not very frequent in both languages and appears mainly in literary style. This characteristic seems to be reflected in the use of *principiar a* + infinitive too. According to Zavadil and Čermák (2010), the Spanish ingressive construction with *principiar a* "is completely literary and is virtually not used at all in the general spoken language"[20] (318). Similarly, the Portuguese *principiar a* is mentioned only by Baroso (1988), who based his research on excerpts from literature.

While this seems to be the reason why we did not find many occurrences of this ingressive construction in the *Aranea* corpora, there is a notable difference between Spanish and Portuguese in the frequency distribution of this construction, which leads to the conclusion that it is more common in Spanish. This fact and the stylistic limitation of the construction *principiar a* + infinitive place it among the stylistically marked constructions that were analysed in Section 4. The results we obtained regarding this group of ingressive constructions are summarised in Table 11:

	Spanish: *Araneum Hispanicum Maius*		Portuguese: *Araneum Portugallicum Maius*	
	Outside	Inside	Outside	Inside
Ponerse (Sp.) / *Pôr-se* (Pt.)	679 856	31 433	62 232	1 098
Romper	68 623	863	17 262	4
Desatar	10 092	0	1 215	212
Soltar	23 582	95	18 447	0
Echar(se) (Sp.) / *Deitar(-se)* (Pt.)	99 905	4 208	11 574	15
Total	882 058 (96%)	36 599 (4%)	110 730 (98,8%)	1 329 (1,2%)

TABLE 11. Distribution of *ponerse*, *pôr-se*, *romper*, *desatar*, *soltar*, *echar(se)*, *deitar(-se)* and *principiar* outside and inside an ingressive periphrastic construction: the corpora *Araneum Hispanicum Maius* and *Araneum Portugallicum Maius*

20 „zcela knižní a v běžně mluveném jazyce prakticky nepoužívaná". Translation by the authors.

While the previous set of results was similar for both Spanish and Portuguese (with the exception of constructions with *principiar*), Table 11 clearly shows distinctively higher frequency of stylistically marked ingressive constructions in Spanish. As in the case of *principiar a* + infinitive, we can observe that the lower frequency of an ingressive construction is only partly related to the frequency of its semi-auxiliary when used as a fully lexical verb. Frequent Portuguese verbs like *romper*, *quebrar*, *partir* or *soltar* rarely form an ingressive construction that would correspond to the Spanish *romper a* or *soltar a* + infinitive. On the other hand, we found 212 occurrences of the Portuguese periphrasis with *desatar a* that constitute almost 15% of all the uses of this verb and no occurrence of the *desatar a* + infinitive construction in the Spanish corpus.

In spite of the fact that especially the Portuguese set of results was often limited, we were usually able to identify clearly the semantic limitations for auxiliated verbs that appeared in a concrete construction. This leads us to the conclusion that the selection of auxiliated verbs in an ingressive construction is not primarily due to some kind of convention, but that there is a strong relationship between the original semantic characteristics of a semi-auxiliary and the auxiliated verb. This endorses our original assumption that an analysis of correspondences between Spanish and Portuguese ingressive constructions should also be based on the closeness of the original meanings of selected semi-auxiliaries.

It may be interesting to note that a similar conclusion can be drawn when analysing the Czech means for expressing the beginning of a process. As attested by an exhaustive analysis of a large sample of Czech translations of Spanish ingressive periphrases made by Kratochvílová (2015), the selected Czech construction is to a considerable extent also influenced by the semantics of the auxiliated verb that appears in it (cf. 160–173).

On the basis of the analogies between Spanish and Portuguese that can be found in the area of the original meaning of the semi-auxiliaries, in the semantic features associated with the resulting ingressive construction and in the characteristics of the infinitives that can appear in the periphrasis, it is possible to identify the main systemic correspondences between the most frequent ingressive constructions in Spanish and Portuguese. For the basic schema, see Figure 3.

<table>
<tr><th>Spanish</th><th>Portuguese</th></tr>
<tr><td>ponerse</td><td>pôr-se</td></tr>
<tr><td>romper</td><td>romper / partir</td></tr>
<tr><td>soltar(se)</td><td rowspan="2">desatar</td></tr>
<tr><td>echar(se)</td></tr>
<tr><td></td><td>deitar(-se)</td></tr>
</table>

FIGURE 3. Most frequent Spanish semi-auxiliaries in stylistically marked ingressive periphrases and their correspondences in Portuguese

However, as mentioned previously, Figure 3 displays only theoretical (systemic) correspondences and does not take into account the frequency of stylistically marked constructions. As we could see above, the difference between Spanish and Portuguese in this area is striking and opens up possibilities for further research.

REFERENCES

Barroso, H. F. (1988) *O aspecto verbal perifrástico em português contemprâneo* (Ph.D. thesis). Braga: Universidade do Minho.

Bechara, E. (2001) *Moderna gramática portuguesa*. Rio de Janeiro: Editora Lucerna.

Benko, V. (2014) Aranea: yet another family of (comparable) web corpora. In: Sojka, P., A. Horák, I. Kopeček, K. Pala (eds) *Text, Speech and Dialogue. 17th International Conference, TSD 2014, Brno, Czech Republic, September 8–12, 2014. Proceedings. LNCS 8655*, 257–264. Cham: Springer International Publishing.

Benko, V. (2015) *Srovnatelné webové korpusy Aranea* [Comparable Web Corpora Aranea]. Praha: Ústav Českého národního korpusu FF UK. Retrieved on 29 Feb 2016 from http://www.korpus.cz.

Castilho, A. T. de (1966) *Introdução ao Estudo do Aspecto Verbal na Língua Portuguesa*. Maria, FFCL.

Ceolin, R. (2003) Falsos amigos estruturais entre o português e o castelhano. *Ianua* 4, 39–48.

Drzazgowska, J. (2011) As perífrases verbais no português europeu. *Romanistica Cracoviensia* 11, 107–115.

Fernández de Castro, F. (1999) *Las perífrasis verbales en el español actual*. Madrid: Gredos.

García Fernández, L. et al. (2006) *Diccionario de perífrasis verbales*. Madrid: Gredos.

Gómez Torrego, L. (1988) *Perífrasis verbales*. Madrid: Arco Libros.

Gómez Torrego, L. (1999) Los verbos auxiliares. Las perífrasis verbales de infinitivo. In: Bosque, I. and V. Demonte (eds) *Gramática descriptiva de la lengua española*, 3323–3389. Madrid: Espasa Calpe.

Hricsina, J. (2006) *Vývoj modotemporálních paradigmat u portugalského verba finita z diachronního hlediska* [Diachronic Study of the Evolution of Modotemporal Paradigmes of the Portuguese "Verbum Finitum"] (Ph.D. thesis). Praha: FF UK.

Jindrová, J. (2015) Úvodem k ingresivním perifrázím [An Introduction to Ingressive Periphrases]. In: Čermák, P. and O. Nádvorníková (eds) *Románské jazyky a čeština ve světle paralelních korpusů* [Romance Languages and Czech in the Light of Parallel Corpora], 150–154. Praha: Karolinum.

Kratochvílová, D. (2015) Perifráze s významem iniciální ingresivnosti: španělština [Periphrasis with the meaning of initial ingression]. In: Čermák, P. and O. Nádvorníková (eds) *Románské jazyky a čeština ve světle paralelních korpusů* [Romance Languages and Czech in the Light of Parallel Corpora], 160–173. Praha: Karolinum.

Mateus M. H. M. et al. (2003) *Gramática da língua portuguesa*. Lisboa: Caminho.

Pešková, J. (2006) *Slovesné vazby s infinitivem a gerundiem jako prostředek k vyjádření povahy slovesného děje v současné španělštině* [Verbal constructions with infinitive and gerund as a mean of expressing the manner of action in contemporary Spanish] (Ph.D. thesis). Praha: FF UK.

RAE (2009) *Nueva gramática de la lengua española*. Madrid: Espasa.

Santos, D. — Costa, L. (2005) A Linguateca e o projecto "Processamento Computacional do português". *Terminómetro — Número especial nº 7 — A terminologia em Portugal e nos países de língua portuguesa em África*, 63–69.

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

Zavadil, B. and P. Čermák (2010) *Mluvnice současné španělštiny* [A Grammar of Contemporary Spanish]. Praha: Karolinum.

SOURCES

Linguateca (2007) *CETEMPúblico: a large corpus of Portuguese newspaper language*. Retrieved on 25 Feb 2016 from http://www.linguateca.pt/.

Linguateca *CETEMPúblico*. Retrieved on 25 Feb 2016 from http://www.linguateca.pt/cetempublico/.

RAE (2015a) *CORPES XXI*. Retrieved on 25 Feb 2016 from http://www.rae.es/recursos/banco-de-datos/corpes-xxi.

RAE (2015b) *CREA*. Retrieved on 3 March 2016 from http://www.rae.es/recursos/banco-de-datos/crea.

RAE Banco de datos (CORPES XXI, v. 0.82) [en línea]. *Corpus del Español del Siglo XXI (CORPES)*. Retrieved on 25 Feb 2016 from http://www.rae.es.

RAE Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual.*. Retrieved on 25 Feb 2016 from http://www.rae.es.

Dana Kratochvílová
Charles University, Faculty of Arts, Department of Romance Studies
nám. J. Palacha 2, 116 38 Praha 1, Czech Republic
dana.kratochvilova@ff.cuni.cz

Jaroslava Jindrová
Charles University, Faculty of Arts, Department of Romance Studies
nám. J. Palacha 2, 116 38 Praha 1, Czech Republic
jaroslava.jindrova@ff.cuni.cz