

FREQUENCY AND LOCAL CONTEXTS IN LEXICALIZATION AND INCIPIENT GRAMMATICALIZATION

JOSÉ PINTO DE LIMA
(Universidade de Lisboa)

ABSTRACT: Work by Joan Bybee (2007; 2010) and others under the heading of ‘usage based’ perspectives on language has shown the importance of considering frequency in studies of language change: new meanings and constructions seem to emerge from frequent, conventionalized, semantically close items with high token frequency, and expand from such local contexts to wide distribution. The present paper checks the validity of these views by studying the syntactic and semantic changes undergone by the European Portuguese verb arrancar. Starting from three senses in the 19th century, arrancar acquires two new inchoative senses in the 20th century, which have developed in ways compatible with Bybee’s usage based conceptions. In its last stages, this development seems to be one of incipient grammaticalization, which may lead to the formation of a new semi-auxiliary in the future.

1. Importance of frequency and local contexts in lexicalization and grammaticalization.

The aim of my paper is to look anew at processes of lexicalization and grammaticalization, mainly in their incipient stages, in order to deepen our understanding of them. I will mainly rely on ideas from the work of Joan Bybee (2007; 2010) and Bybee and Eddington (2006), which go by the name of ‘usage based conceptions of grammar’.

Bybee has called our attention to the importance of considering the frequency of occurrence of linguistic units when studying processes of change such as lexicalization and grammaticalization. In both processes, two levels have to be taken into account: the level of **categories** – morphological, syntactic, etc. – and the level of their **members**. Changes do not proceed ‘directly’, so to speak, from category to category, but instead innovation begins by affecting a particular member of a category, or a limited number of members. Eventually, **analogy** and **similarity** to these primordial members will

motivate the expansion of the change to other members, so that in the end we are left with a modified, or even a new, category. In grammaticalization, a construction gives rise to another (more grammatical) construction, and constructions are made up of categories: notwithstanding that, the process of change does not go from category to category, but instead from particular members to categories, so that new categories, or new configurations of categories, arise: indeed, this is what change from construction A to construction B is all about. It is a process where attention should be paid both to the **category** and to the **membership** levels. (Notice that constructions can be thought of as categories consisting of combinations of other categories, so that explanations can be framed simply in terms of category and membership.)

Another important aspect of change processes that should be taken into account is this: the primordial, innovative, members of a category will tend to occur very **frequently**, so that, at the beginning, we will have just a **few** members of the future category, but with a **high frequency of tokens** of that member (the member is taken here as being a **type**). Still another feature of these early changes is that the first members will tend to belong to well defined **semantic groupings**, and thus they will have a restricted distribution and will occur mainly in **local contexts**.

I would like to illustrate what I have been saying with an example from Bybee/ Eddington (2006) (presented succinctly in Bybee, 2010, 62-64): the Spanish inchoative construction formed by a verb of becoming followed by adjective or past participle, which we will label the *V'become' + ADJ/PP* construction. Examples of members of this construction are the following combinations:

<i>ponerse nervioso</i>	to get nervous
<i>quedarse solo</i>	to end up alone
<i>quedarse sorprendido</i>	to be surprised
<i>volverse loco</i>	to go crazy

From work with corpora data it could be ascertained that new or innovative combinations of the construction were semantically close to the more frequent and more conventionalized combinations. For instance, the corpora evince several combinations which are semantically similar to *quedarse solo*, a very frequent combination (with 28 occurrences):

<i>quedarse solo:</i>	
<i>quedarse a solas</i>	to end up alone
<i>quedarse soltera</i>	to end up unmarried (fem.) = an old maid
<i>quedarse aislado</i>	to become isolated
<i>quedarse sin novia</i>	to end up without a girlfriend

There is thus a semantic grouping or cluster around the idea of 'be(come) alone'. A further grouping focuses around *quedarse sorprendido*

(‘be(come) surprised’): this combination occurs several times and semantically builds a set with other, not so frequent, combinations:

<i>quedarse sorprendido</i> :	
<i>quedarse deslumbrado</i>	to become dazzled
<i>quedarse asombrado</i>	to become amazed
<i>quedarse pasmado</i>	to become stunned, astonished
<i>quedarse asustado</i>	to become afraid

Besides *quedarse*, the verb *ponerse* also enters combinations with adjectives, of which *ponerse nervioso* is the most frequent (17 occurrences), thus being at the center of a group of semantically related combinations:

<i>ponerse nervioso</i> :	
<i>ponerse pálido</i>	to become pale
<i>ponerse histérico</i>	to become hysterical
<i>ponerse furioso</i>	to become furious
<i>ponerse coloradoto</i>	turn red/become flushed

In what follows, I will try to apply these tenets of Bybee’s usage based theory to the development of the Portuguese verb *arrancar* (‘to pull out’) from the 19th century until today. I will try to show Bybee’s ideas to be illuminating about the importance of type and/or token high frequency: chains of changes seem to exist indeed which begin with the emergence of a small number of highly frequent new members of a category in local contexts, followed by an expansion of membership, which in turn leads to the creation of a new category or categories.

2.1. Constructions with the verb *arrancar* in the 19th century.

The verb *arrancar* – whose first meaning is ‘pull out, tear away’ – had basically three senses in the 19th century:¹

A) The transitive ‘pull out’ sense:

- (1) Ele *arrancou* o prego *da* parede²
‘He *pulled* the nail *out of* the wall’
- (2) Gonçalo *arrancou* (...) folhas *dum* ramo de lúcia-lima
‘Gonçalo *picked* some leaves *from* a branch of verbena’ (CdP, 19th c.)

¹ Although the analysis of *arrancar* that follows is structured in terms of senses, it should not be forgotten that a full grasp of the senses in semantic and syntactic terms presupposes that they can be made to correlate with *constructions*, a concept we will also use throughout (on constructions and their importance for explaining sentence meaning and language change, see Goldberg, 1995, and Bergs / Diewald, 2008).

² In this sentence and throughout the article, all italics and underlinings are mine, unless otherwise indicated.

SUBJ _{hum} agent	<i>arrancar</i> 'pull out'	OBJ patient	P COMP (<i>de</i>) source
------------------------------	-------------------------------	----------------	--------------------------------

We have here a three argument construction, with a human subject, a direct object, which is typically a physical object, and a prepositional complement introduced by the preposition *de*, which indicates the source of the movement of pulling out.³

B) The 'draw' sense

(3) Ele *arrancou* da espada
'He *drew* his sword'

(4) queria esperar Francisco de Mendonca, *arrancar da espada* contra ele
'[he] wanted to wait for Francisco de Mendonça, to *draw* his sword against him' (CdP, 19th c.)

SUBJ _{hum} agent	<i>arrancar</i> 'draw'	P COMP (<i>de</i>) patient
------------------------------	---------------------------	---------------------------------

Here a human subject pulls out an object suddenly, energetically. Note that the prepositional complement *da espada* (Prep. *de* + *a espada*) does **not** signal a source, but simply the **object** that is being pulled out, and is therefore given the patient role. Semantically, *da espada* is indeed the **object** of *arrancar*. It could be classified as a 'prepositioned' object⁴, but the presence of the preposition *de* is still in need of an explanation.

C) The 'start' sense

(5) Ele *arrancou* em direção à porta
'He *started* for the door'

(6) Gonçalo *arrancou* do banco, erguendo os braços
'Gonçalo *started* from the bench, raising his arms' (CdP, 19th c.)

³ Besides being used literally, the verb is also frequently used metaphorically, to express a great variety of more abstract actions or processes, as exemplified by: *arrancar um autógrafa a alguém* ('to *snatch* an autograph from someone'); *arrancar uma palavra de alguém* ('to *get a word out of* somebody'). (These examples were taken from the CRPC and the CdP respectively.)

⁴ 'Prepositioned' objects are rare in Portuguese, but the possibility of the sentence *Ele viu a Deus* ('He saw God'), where the object of the verb *ver* is introduced by the preposition *a* is a case in point. The preposition *a* is also used with objects when these are emphasized by double marking, such as in *Eu queria ajudá-lo a ele* (literally: 'I wanted to help him to him'): here *a ele* is an emphatic repetition of the object *lo*. In Spanish, objects introduced by *a* are frequent, indeed obligatory, when these objects are human: *yo conozco a Juan* ('I know Juan').

SUBJ _{hum} agent	<i>arrancar</i> 'start' (movement)
------------------------------	---------------------------------------

This is an intransitive use with a human subject, signifying a sudden movement away from a place. There can be an indication of direction by means of an adjunct: either the goal (*em direcção à porta*, 'for the door') or the source (*do banco*, 'from the bench') can be indicated. The emergence of this sense may be found in those instances of the transitive 'pull out' sense (A) where the human subject and the object are co-referential, or partly so: i.e., sentences where the subject 'pulls' him/herself, or his/her body, out of a place, so to speak. Consider the following example:

- (7) Titó (...) *arrancou* do canapé o corpanzil imenso
'Titó (...) *pulled out* from the couch his huge body' (CdP, 19th c.)

In sentence (7), the object designates (the body of) the human which is the subject of the sentence. This co-reference between subject and object causes the expression of the object to be felt to be redundant, so that the object ceases to be verbalized and in the end disappears, giving rise to the intransitive 'start' construction exemplified by (5) and (6).

There is a variant of this construction with a non-human subject, as exemplified by (8):

- (8) O barco *arrancou* (do porto) (às dez)
'The boat *started* (= *broke away*) (from the harbor) (at ten)

SUBJ _{n-hum} theme	<i>arrancar</i> 'start' (movement)
--------------------------------	---------------------------------------

In these variants, the subject may be human or not, but what is common to both is that the verb expresses a sudden movement, involving detachment from a position.

This 'start' sense is related to the 'pull out' sense by **valence reduction**: the object of the 'pull out' sense corresponds semantically to the subject of the movement sense. What is retained from the 'pull out' sense is the idea of a **sudden movement of an object**.⁵

To understand better the relative importance of these senses in the 19th century, we have looked up frequency data in the CdP corpus. We only considered the three most frequent grammatical forms of the verb, which were

⁵ The genesis of this sense can also be explained in cognitive-semantic terms by appealing to the concept of a **scene**. The transitive 'pull out' sense calls up a whole scene: there is an agent that causes an object to suddenly move or to detach itself from a place. On the contrary, the intransitive 'start' sense only highlights a part of this scene, namely the object suddenly moving or detaching itself from the place.

the Infinitive (Inf.), the 3sg Preterit (Pret.) and the 3sg Present Indicative (Pres. Ind.) – *arrancar*, *arrancou* and *arranca* –, and found there to be 122 European Portuguese (E.P.) occurrences.⁶

Looking at Table 1, we observe that the most frequently used sense in the 19th century was by far the **transitive ‘pull out’ sense**, with 115 occurrences, representing more than 94% of the total. The other senses – i.e., the movement ‘start’ sense and the ‘draw’ sense – have only about 3% of the occurrences each.⁷

Senses of <i>arrancar</i> :	Nr occurrences	Percentage
Transitive ‘pull’ sense	115	94,3%
Movement ‘start’ sense	4	3,3%
‘Draw’ sense	3	2,4%
Total	122	100%

Table 1 – Occurrences in the CdP of the E.P.
verb forms *arrancar*, *arrancou* and *arranca*, 19th century.

2.2. Constructions with *arrancar* in the 20th century: emergence of the inchoative ‘begin’ senses.

Coming now to the 20th century, the most relevant phenomenon concerning *arrancar* is the emergence of **two new senses** of the verb, a process which we will go on to describe next.

It seems plausible to admit that, at the origin of this change, is the **movement ‘start’ sense** of *arrancar*, as exemplified by sentence (8), repeated here simplified as (9):

- (9) O barco *arrancou* (às dez)
‘The boat *started* (= *broke away*) (at ten)’

By expressing the **sudden beginning of a movement**, *arrancar* also comes to signal the **beginning** of the process. Indeed, by saying (9), the speaker *means* (9a), but allows the *invited inference* (9b):

⁶ For the 19th century, the corpus registers a total of 1 607 occurrences of the verb *arrancar*, both from Brazilian Portuguese (B.P.) and European Portuguese (E.P.). The three most frequent grammatical forms - *arrancar*, *arranca* and *arrancou* - account for 594 occurrences, i.e. 37% of the total. As our work concerns only E.P., and the corpus does not allow automatic extraction of the E.P. occurrences, we had to proceed manually to extract the E. P. occurrences of these forms, which were found to be 122.

⁷ These two senses, judging from their small number of occurrences, might be either falling into disuse or, on the contrary, might be emerging as new senses; but the number of occurrences does not allow any definitive conclusions.

- (9a) The boat *broke away* (at ten) (= > semantic meaning)
 (9b) The boat *began its voyage* (at ten) (= > invited inference)

In (9), we thus have a *critical context or bridging context* in the sense of Diewald (2002) and Heine (2002), where **both** the ideas of **movement** and of **beginning** are present, the second being only an invited inference. This invited inference later becomes an **implicature**⁸ (Grice, 1989) and, later still, it **conventionalizes**, i.e. the implicature ‘begin (a voyage), start (on a path)’ turns into a **new, inchoative, meaning** of the verb. There is thus an extension from **spatial** to **temporal** uses of the verb – a very well known development in grammaticalization.

We have seen that, in the movement ‘start’ sense, the subject may be human, as in (5-6), or non-human, as in (9). As in this sense *arrancar* meant a movement, the non-human subjects often designated vehicles (through nouns such as *barco*, ‘boat’, or *carro*, ‘car’).

But when the subject was human, mention of the vehicle was possible by adding an adjunct in instrumental function, mostly in the form of a prepositional phrase introduced by *com* (‘with’). For instance, besides sentences like (9), it became possible to have sentences like (10), where the cause of the movement is given as shared between a human subject and an instrumental adjunct that designates the vehicle:

- (9) O barco *arrancou* (às dez)
 ‘The boat *started* (at ten)’
 (10) O capitão *arrancou com o barco* (às dez)
 ‘The captain *started with the boat* (at ten)’

This possibility is well attested in the 20th century:

- (11) quando Lima tentara *arrancar com o veículo*.
 ‘when Lima had tried *to start with the vehicle*.’ (CRPC, 20th c.)

However, it was not just instrumental adjuncts that could occur with the human subjects of *arrancar*. Adjuncts could be of various kinds, such as: company (*ele arrancou para Paris com o irmão*, ‘he started for Paris with his brother’), manner (*ele arrancou com pressa*, ‘he started with haste’), etc. When *arrancar* acquired the implicational-conversational meaning of ‘begin’, it came to be used with **adjuncts that were incompatible with the movement sense**, as is the case in (12):

⁸ Implicatures are intended by the speaker, while invited inferences are inferences of the hearer, which may or may not have been intended by the speaker.

- (12) Ele *arrancou* com o trabalho às dez.
 ‘He began with the work at ten.’

In (12), the prepositional phrase will not make sense if *arrancar* is taken as a verb of movement (‘he *moved* with the work’??), but it will make perfect sense if it is taken to mean ‘begin’. Contexts such as (12) are *switch* or *isolating contexts* in the sense of Diewald (2002) and Heine (2002), i.e. contexts where, in the process of a change from an old sense to a new one, the interpretation of an expression according to the older sense is no longer possible, and only the interpretation according to the new one is. In the present example, an interpretation according to the movement ‘start’ sense is no longer possible and only an interpretation according to the inchoative ‘begin’ sense is. It thus emerges what I will call the *inchoative agentive sense*. It should be noted that, in (12) and in the other instances of this sense, the prepositional phrase is no longer an adjunct but a complement, as its presence is necessary for the grammaticality of the sentence.

D) The inchoative agentive sense

- (13) No dia seguinte à autorização [nós] poderemos *arrancar* com as obras.
 ‘The day after the authorization [we] will be able *to start* with the works’
 (CRPC, 20th c.)

SUBJ _{hum} agent	<i>arrancar</i> ‘begin’	P COMP (<i>com</i>) theme
------------------------------	----------------------------	--------------------------------

Besides this *inchoative agentive construction*, another construction comes into existence in the 20th century, which is exemplified by sentence (14).

E) The inchoative non-agentive sense

- (14) As obras devem *arrancar* no próximo Verão.
 ‘The works shall *begin* next Summer.’

SUBJ _{n-hum} theme	<i>arrancar</i> ‘begin’
--------------------------------	----------------------------

In this construction, the human agent responsible for the beginning of the activity is no longer mentioned and the subject position is occupied by the name of the activity that was expressed by the prepositional complement (P COMP) in the inchoative agentive construction. I will call this sense of *arrancar* the *inchoative non-agentive sense*.

As for the question of how the inchoative non-agentive sense might have originated, we believe three possibilities should be considered:

i) It might have originated from the **temporal interpretation** of the movement ‘start’ sense C (see above interpretation (9b) of sentence (9)):

O barco *arranca* às dez => As obras *arrancam* às dez
The boat *begins its voyage* at ten => The works *begin* at ten

According to this hypothesis, there would have been an extension in the kinds of subjects accepted by the movement ‘start’ sense in its temporal reading: from humans or vehicles to other entities related to humans or vehicles, such as activities. The relation is by **metonymy**: humans or vehicles are contiguous with *their* activities (on changes of a metonymical nature, see Heine *et al.*, 1991, and Hopper/ Traugott, 2003, among others).

ii) It might have emerged from inchoative agentive sense D, by something like an **anticausative operation of valence reduction**, by which the prepositional complement of the inchoative agentive sense would become the subject of the new construction, and the subject of the inchoative agentive sense would cease to be expressed:

Nós *arrancamos* com as obras (às dez) => As obras *arrancam* (às dez)
We *begin* with the works (at ten) => The works *begin* (at ten)

SUBJ _{hum} agent	<i>arrancar</i> ‘begin’	PCOMP _(x) (<i>com</i>) theme		SUBJ _{n.} - _{nam(x)} theme	<i>arrancar</i> ‘begin’
------------------------------	----------------------------	---	--	--	----------------------------

(Arguments marked ‘(x)’ are co-referential)

(iii) Finally, it might have had its origin in *both* types of changes (i) and (ii), because these changes are not incompatible and may have had taken place simultaneously; indeed, I believe this to be the most plausible hypothesis

Having seen how two new senses of *arrancar* meaning ‘begin’ have emerged in the 20th century, the questions we should now approach concern the **importance** of these senses and their **kind of development**: i.e., we should study these new senses under the perspectives opened up by the work of Bybee (2007; 2010), investigate their **frequency** in comparison to the other senses and ascertain how this frequency relates to their development in terms of indicators such as **type diversity, token frequency and productivity of the verb arguments**.

When we go from the 19th to the 20th century, the most remarkable feature of change is that, while in the 19th century the **transitive ‘pull out’ sense** was overwhelmingly dominant, with 94% of the occurrences, the data for the 20th century show a quite different picture: although in this century the **transitive ‘pull out’ sense** continues to be dominant, it is only slightly above 50% of the occurrences. The reason for this **decrease** in frequency is

the rise of the two **new inchoative senses**, which have grown in the 20th century to about **38% of the occurrences**, as reflected in **Table 2**.

Senses of <i>arrancar</i> :	Nr occurrences	Percentage
Transitive ‘pull’ sense	166	52,2%
Inchoative non-agentive sense*	93	29,3%
Inchoative agentive sense*	27	8,5%
‘Start’ sense	29	9,1%
‘Draw’ sense	3	0,9%
*Total of the inchoative senses	120	37,8%
Total of all senses	318	100%

Table 2 – Occurrences in the CdP of the E.P.
verb forms *arrancar*, *arrancou* and *arranca*, 20th c.

In order to explain and characterize the remarkable **increase in frequency** of the inchoative senses in the 20th century, we had recourse to the *CRPC – Corpus de Referência do Português Contemporâneo*.⁹

We have collected the first 100 random occurrences of the infinitive *arrancar*, and have eliminated the occurrences that: (i) belonged to the other senses, (ii) were not dated, and (iii) were incomprehensible (one occurrence). We came up with 42 dated occurrences of *arrancar*, where *arrancar* is either the non-finite verb of a clause or part of verbal periphrases made up of finite verb + infinitive *arrancar*, such as *vai arrancar*, *deve arrancar* (‘will begin, shall begin’), etc.¹⁰

The first thing that strikes us is that the emergence of the inchoative senses is quite a recent development. Occurrences before 1990 must have been infrequent and in our sample we found none. However, this does not mean that there were no occurrences before that date, and that the emergence of the inchoative senses had not begun sooner. It certainly had begun sooner, but the first occurrences must have been *ambiguous* between the temporal reading of the movement ‘start’ sense and the new inchoative meaning.

3. The development of the ‘inchoative senses of *arrancar*. Local contexts.

The next task we turned to was to characterize the kind of development that these new senses have undergone since their emergence. According to

⁹ The consulted version was the *CRPC - Portugal only, v2.1*. This version contains only occurrences of European Portuguese and consists of about 290 million words, mainly 20th century, but including a very small amount of 19th century occurrences.

¹⁰ In most cases these verbal periphrases are made up of an auxiliary, modal, or semi-modal verb, followed by the infinitive *arrancar*.

usage-based conceptions such as Bybee's (2010), it is to be expected that new senses and constructions will develop starting **from local similarities of meaning, i.e. from quite well delimited, restricted semantic contexts of use**, beginning **with a small number of types, some of which will evince however a high rate of tokens per type**, so that these tokens may eventually be taken as the basis for the expansion of the category by similarity (Bybee, 2010, 62-64).

And indeed this seems to be what happened in the present case, as we will try to show. Let's consider the **inchoative agentive construction** first, and look at the token and type frequency of its subjects and the prepositional complements. When we consider the chosen occurrences, it becomes obvious that they can be grouped in **semantic clusters**, as mirrored in Table 3 below.

As for the **subjects**, we notice that:

- (i) The most important semantic cluster centers around the concepts of **municipality or town hall**. The noun *câmara* ('town hall') occurs twice, and so does the noun *autarquia* ('local government').
- (ii) A second important cluster is made up of proper names, more specifically names of **companies** (such as *Carrefour*) or of **humans** representing **companies**.

In what concerns the **prepositional complements**, it can be seen that:

- (i) The most important cluster is built around the concept of **public works and building constructions** as given by the word *obras* (which means precisely 'constructions', 'works'). **All** prepositional complements have the noun *obras* in them, so that there occurs **only one type** and **all tokens** are of that one type. The prepositional complements thus often assume the form *com as obras* ('with the construction(s)/works'). The word *obras* is therefore a **high-frequency exemplar** of its category and is likely to be interpreted as a **central member** that will attract other new members to the category (Bybee, 2010, 79-80).
- (ii) The next cluster of prepositional complements groups names of **products** of the activity of construction: for instance, names of **buildings** that were built or were expected to be built.

The data allow us to conclude confidently that the inchoative agentive sense of *arrancar* is concentrated on a **high frequency of exemplars** that might have a sentence like (14) as their prototype:

- (14) A Câmara *vai arrancar* com as obras
 'The Town Hall *will begin* with the construction(s)'

Notwithstanding the preponderance of the semantic clusters related to the concept of building construction, expansion to other semantic domains can be observed to be well under way, as shown by (15): in this occurrence, *arrancar* has as its agentive subject the name of a pop singer (*Gonzo*) and as

its prepositional complement the name of the song that he began singing (*'Lugares'*):

- (15) Gonzo subiu ao palco (...) para *arrancar* com "Lugares"
 'Gonzo went on stage (...) to *begin* with 'Lugares'" (CRPC, 20th c.)

We will now move on to our sample of occurrences of the **inchoative non-agentive sense** and check if the semantic cluster around the concept of building construction keeps on being predominant. As we have said, the subjects of the sentences in the non-agentive sense correspond to the prepositional complements of the agentive sense. It is therefore to be expected that, just as these prepositional complements mostly designated works or constructions, the subjects of the non-agentive sense also designate **works or constructions**. And indeed, this is exactly what happens, as shown in **Table 4**. If we analyze the data in this Table in terms of semantic clusters, we observe that:

- (i) The strongest semantic cluster in the class of subjects is precisely the one around **construction/ works**, with 8 occurrences. Of these, 6 are tokens of the type *obras* and the other 2 are near-synonyms - *construção* and *reconstrução*.
- (ii) Close to this cluster, there is a second one that specifies **construction activities**, such as the building of a road connection or the recuperation of a beach.
- (iii) The third semantic cluster has to do with **projects or plans**, namely those related to **constructions**: from the 4 occurrences in this cluster, 2 refer to **projects involving constructions**.

Just as with the inchoative agentive sense, the non-agentive sense also revolves around the cognitive domain of works and construction. However, it has already expanded significantly to other domains, as is the case of the semantic cluster around **public events** (festivals, sports events, etc.), with 5 occurrences, and the one around **public services**, with 4 occurrences.

SUBJ _{hum} : semantic clusters:	SUBJ _{hum} : occurrences	Nr of occ. (main types/ tokens)	P COMP: semantic clusters.	P COMP: occurrences	Nr of occ. (main types/ tokens)
municipalities, public institutions	4- <i>a autarquia vilandense</i> 'the Town Hall of Vila do Conde' 5- <i>a Câmara</i> 'The Town Hall' 39- <i>a autarquia vilandense</i> 40- <i>a Câmara de Odivelas</i> 41- <i>Feira</i> (town name) 42- [<i>a administração</i>] [the administration]	6 - main types: autarquia: 2 tokens Câmara: 2 tokens	works/ constructions	4- <i>com as obras de remodelação da marginal</i> 'with the works of remodeling the coastal road' 23- <i>com as obras</i> 'with the works' 25- <i>com as obras</i> 38- <i>com as obras de terraplanagem</i> 'with the works of ground leveling' 39- <i>com as obras de remodelação da marginal</i> 40- <i>com cinco obras de vulto</i> 'with five important works'	6 - only type: obras: 6 tokens (rate: 100%)
companies	25- <i>Carrefour</i> 26- <i>A MEDI Telecom</i> 27- <i>O mercado</i> 'The market' 23- [<i>nós</i>] [we] 24- [<i>Human, non-specified</i>]	6	public buildings	5- <i>com o albergue municipal</i> 'with the municipal hostel' 24- <i>com a Casa-Museu</i> 'with the Museum House'	2

human entity, impersonal or definite	38- [Human, non-specified] 6- [nós] [we] 22- [Gonzo] (proper name) 28- [a docente da Faculdade de Psicologia] [the teacher of the Faculty of Psychology]	3	public services	41- <i>com novo serviço de transportes</i> ‘with the new transport service’ 42- <i>com um serviço de consulta e tratamento</i> ‘with a service of consultation and treatment’ 6- <i>com o processo</i> ‘with the process’ 22- <i>com "Lugares"</i> ("Lugares": name of a song)	2 - only type: serviço : 2 tokens
Total occ.		15*	projects/ processes song Total occ.	1 1	12*

*The disparity between 15 occurrences of subjects and only 12 occurrences of prepositional complements has to do with the fact that the prepositional complements were not always expressed.

Table 3 – Subjects and prepositional complements of the occurrences of the inchoative agentive construction, classified by **semantic clusters**, in the first 100 occurrences of a random list of E.P. *arrançar* (Inf.) of the CRPC, 20th century. All the occurrences have been numbered: the numbers of the examples indicate their order in the random list.

SUBJ _{hum} : semantic clusters:	SUBJ _{n-hum} : occurrences:	Nr of occ. (main types/ tokens)
works/ construction	11- <i>a reconstrução do muro</i> ‘the reconstruction of the wall’ 13- <i>As obras</i> ‘the works’ 14- <i>a sua construção</i> [dos matadouros] ‘their construction [of the slaughterhouses]’ 15- <i>uma obra</i> 16- <i>as obras</i> 19- <i>As obras</i> 20- <i>As obras de construção do porto de abrigo</i> ‘the construction works of the harbor’ 34- <i>as obras</i>	8 main type: <i>obra(s)</i> 6 tokens (rate: 75%) other types: <i>(re)construção</i> 2 tokens
activities involving works/ construction	2- <i>a [ligação] da Marateca a Grândola</i> ‘the connection Marateca-Grândola’ 31- <i>Recuperação da praia da Apúlia</i> ‘Recuperation of the beach of Apúlia’ 33- <i>Dragagens na foz do Minho</i> ‘Dredging at the mouth of the Minho River’	3
projects/ plans	7- <i>16 planos de ordenamento das áreas protegidas</i> ‘16 plans for the redesign of the protected areas’ 12- <i>uma fase deste projecto</i> ‘a phase of this project’ 18- <i>O protocolo</i> ‘The protocol’ 30- <i>cerca de dez dos projectos apresentados</i> ‘about ten of the projects presented’	4

public events	<p>1- <i>a 12.ª edição do Festival dos Capuchos</i> 'the 12th edition of the Capuchos Festival'</p> <p>3- <i>A campanha</i> 'the campaign'</p> <p>32- <i>o evento desportivo</i> 'the sports event'</p> <p>37- <i>Época balnear</i> 'Bathing season'</p> <p>35- <i>a Bagdad Expo</i></p>	5
public services or activities	<p>8- <i>O COMÉRCIO electrónico</i> 'Electronic commerce'</p> <p>21- <i>a telemedicina</i> 'TV-medicine'</p> <p>29- <i>Centro agro-alimentar de Castelo Branco</i> 'Agriculture and Food Center of Castelo Branco'</p> <p>36- <i>Ações de segurança</i> 'Security actions'</p>	4
companies	<p>10- <i>Terceiro operador [de telecomunicações]</i> 'Third [telecommunications] operator'</p> <p>17- <i>[a] nova Lisnave</i> '[the] new Lisnave (=name of company)'</p>	2
processes	<p>9- <i>um processo deste tipo</i> 'a process of this kind'</p>	1
Total		27

Table 4 – Subjects of the inchoative non-agentive construction, classified by **semantic clusters**, in the first 100 occurrences of the random list of E.P. *arrancar* (Inf.), CRPC, 20th c.

These data are in accordance with the hypotheses put forward by Bybee and associates on category formation and expansion. They show that new senses tend to emerge around a few, limited number of **types**, which are close **semantically** or at least evince a semantic-pragmatic relatedness (for an example of semantic-pragmatic relatedness, take the second cluster of Table 4: the subject *recuperação da praia* ('recuperation of the beach'), is not semantically related to works or constructions in any direct manner, but our knowledge of the world tells us that this kind of recuperation generally involves works and constructions.) Also, some of the types tend to show a remarkably **high token frequency**, as is the case with the first semantic cluster of Table 4. From this situation, expansion to **other types** follow, accompanied by an increase in **overall token frequency**.¹ It is this last stage of the change that *arrancar* has probably not fully reached yet. The question deserves some comments, to which we will proceed in the next section.

4. From local contexts to wide distribution and increased frequency

As we said above, the **inchoative non-agentive construction** is expanding to other types: its subjects do not all belong to the semantic domain of constructions, but can be found in other domains, such as public events and public services, as shown in Table 4. The **inchoative agentive construction** is also expanding the range of the semantic clusters of its prepositional complements from the realm of works and constructions to public services, to projects, and even to quite different domains such as singing, as evinced by Table 3.

In order to have an interconnected view of the processes by which these senses have emerged and developed, let's present a brief summary of the changes. We will consider the inchoative agentive construction first. As for its subject, we see that:

- i) At first, its subject is human (sentence 16, below).
- ii) Then there is an expansion from human subjects to organizations and collective entities such as municipalities, companies, public institutions, associations, clubs, etc. (from sentence 16 to 17).

As for the prepositional complement:

¹ Bybee's theory about the role of *high frequency* in language change would benefit from a systematic comparison with Rudi Keller's idea of the importance of *cumulation* of individual (speech) acts in the emergence (or death) of linguistic conventions (see Keller, 1994). Bybee's *usage theory* and Keller's *invisible-hand theory* are compatible and somehow complementary, the first one being more empirical and the latter more conceptual. Both recognize the importance of *individuals' actions* in language change, a recognition that is not always present in theories of language change that prefer to frame change in purely mentalistic terms (on this question, see Luraghi, 2010).

- i) At first we have general names of human activities, such as *obras* ('works') or *construção* ('construction') (sentences 16 and 17).
 - ii) Then an extension takes place from the name of the activity to the name of the product of the activity: for instance, from 'construction' to 'school' (from sentence 17 to 19); an intermediate stage in which both the activity and its product are named is also possible (sentence 18):
- (16) O presidente da Câmara *arranca* com a construção (em junho)
'The mayor begins with the construction (in June)'
 - (17) A Câmara arranca com a construção (em junho)
'The Town Hall begins with the construction (in June)'
 - (18) A Câmara arranca com a construção da escola (em junho)
'The Town Hall begins with the construction of the school (in June)'
 - (19) A Câmara arranca com a escola (em junho)
'The Town Hall begins with the school (in June)'²

We can see that this change is metonymical in nature, as it goes from the activity to the product of the activity; the subject of the construction could also be said to have undergone metonymical extension from human individuals to human collectives: for instance, from 'mayor' to 'town hall' (again, on metonymy see Heine et al., 1991, and Hopper/ Traugott, 2003).

Turning now to the non-agentive construction, its development parallels the development of the agentive construction:

- (20) A construção arranca (em junho)
'The construction begins (in June)'
- (21) A construção da escola arranca (em junho)
'The construction of the school begins (in June)'
- (22) A escola arranca (em junho)
'The school begins (in June)'³

Other expansions have gone from activities to events, services, projects, plans, programs, etc. All of them are linked metonymically, such as the following one, which goes from an *activity* to the *program* according to which the activity should be carried on:

- (23) A vacinação arranca em junho
'Vaccination begins in June'
- (24) O programa de vacinação arranca em junho
'The vaccination program begins in June'

Expansion to other types will render possible the use of the inchoative non-agentive construction with virtually *any* subject that might be related

² Meaning: '... begins with the construction of the school...'

³ Meaning: 'the construction of the school begins...'

with the notion of activity: e.g., the following sentences are already acceptable in present-day E.P.:

- (25) O concerto *arrancou* às 10 horas.
'The concert *began* at 10 o'clock'
- (26) O jantar *arrancou* com uma excelente sopa de legumes
'Dinner *began* with an excellent vegetable soup'
- (27) O primeiro semestre *arranca* a 15 de setembro
'The first semester *begins* on September 15'

Later on, expansion goes even beyond the concept of activity: as activities take place in **time**, any noun designating a fragment or quantity of time will be allowed as the subject of the inchoative non-agentive construction:

- (28) O ano *arranca* a 1 de janeiro
'The year *begins* on January 1'

In the end, inchoative *arrancar* will be possible in an enormous variety of contexts and its frequency will increase considerably. This way it will get very close, in distribution and meaning, to the verb that is, in present-day E.P., the inchoative verb *par excellence*, i.e. the verb *começar*, and it is possible that it will begin competing with it. In the long run, *arrancar* may even come to take the place of *começar*. As of today, although the wording of a sentence to the effect that the year begins on the 1st of January will normally use *começar*, as in (29), we must acknowledge that sentence (28), with *arrancar*, is no longer felt to be deviant by many E.P. speakers.

- (29) O ano *começa* a 1 de janeiro
'The year *begins* on January 1'

5. Concluding remarks: on the expectable further grammaticalization of *arrancar*.

The recent development taken by inchoative *arrancar*, namely its extension to new contexts in a way that causes its distribution to be close to that of the verb *começar*, justifies the question whether *arrancar* is entering a path of grammaticalization. Semantically, the development of *arrancar* since the 19th century shows that it began as a verb with a concrete meaning of manipulation and movement (see sense A) and came to signal the beginning of an action, event or period, i.e. it came to express *inchoativity*, which is a concept that belongs to the category of *aspect* and is expressed in many languages by *grammatical means*. This development towards the expression of inchoativity can undoubtedly be equated with a change from the less to the more grammatical. It cannot however be said that the verb *arrancar* has already acquired the functions of an aspectual (semi-)auxiliary verb: for instance, it cannot enter into a periphrastic construction with a main verb; on the other hand, the verb *começar* is allowed in this type of construction,

where it functions as an aspectual semi-auxiliary verb and is followed by the preposition *a* and by the non-finite form of a main verb: **começar + a + *V_{inf}***: (equivalent to the Eng. *begin + V-ing*). The following examples show that this construction is possible with *começar* but not with *arrancar*:

(30) Ele *começou* a estudar inglês

(30') *Ele *arrancou* a estudar inglês
'He *began* studying English'

(31) Ele *começou* a treinar

(31') *Ele *arrancou* a treinar
'He *began* practicing'

It is an open question whether *arrancar* will in the future be allowed in this kind of construction. But it probably will. One reason why is that *arrancar* can already be used in a semantically similar construction in which finite *arrancar* is followed by a prepositional phrase introduced by the preposition *com* (**arrancar + com + NP**). When the head of the NP is a deverbal N, the construction is equivalent to the already grammaticalized inchoative periphrastic construction we saw above with the verb *começar* (i.e., **começar + a + *V_{inf}***):

arrancar + com + NP: => **começar + a + *V_{inf}***:

(32) Ele *arrancou com o estudo* do inglês => (32') Ele *começou a estudar* inglês
'He *began with the study* of English' 'He *began studying* English'

(33) Ele *arrancou com o treino* (ontem) => (33') Ele *começou a treinar* (ontem)
'He *began with the practice* (yesterday)' 'He *began practicing* (yesterday)'

Besides being semantically equivalent, both constructions are productive in E.P. But, as **começar + a + *V_{inf}*** is much more common and has a wider distribution, it is expectable that it may come to serve as the model for the expansion of the **arrancar + com + NP** construction to new contexts. In a second phase, as the parallelism (intersubstitutability) of both constructions grows, it is plausible to admit that a construction **arrancar + a + *V_{inf}*** will emerge that will be equivalent to **começar + a + *V_{inf}***.

Another fact that may lead one to think that the grammaticalization of *arrancar* in Portuguese will be pushed forward until the emergence of the construction **arrancar + a + *V_{inf}*** is the situation that can be observed in Spanish, a language quite close to European Portuguese. In Spanish, the grammaticalization of *arrancar* has already gone so far as to have originated the inchoative construction **arrancar + a + *V_{inf}***, although its use is not yet frequent and it still carries with it the semantic feature of **suddenness** that characterizes the oldest and more lexical senses of the verb, both in Portuguese and Spanish. Here are some examples:

- (34) *Arrancó a cantar*
 ‘[She] (suddenly) began singing’ (DRAE)
- (35) *la profesora arranca a hablar en alemán*
 ‘the teacher (suddenly) begins speaking in German’ (CdE)
- (36) *ella arrancó a sollozar*
 ‘she (suddenly) began sobbing’ (CdE)

Again, frequency of use will probably lead to the expansion of *arrancar + a + V_{inf}* to many other contexts. In some of those, the feature ‘suddenness’ will no longer make sense. When this stage is reached, *arrancar* will not be able to convey anything besides the idea of *beginning*, thus being on the way of becoming an inchoative semi-auxiliary verb in the construction *arrancar + a + V_{inf}*. This grammaticalization process has gone further in Spanish than in Portuguese, but it is plausible to admit that both languages may acquire a new inchoative semi-auxiliary verb in the future.

Sources

- CdE – Davies, Mark. (2002-) *Corpus del Español: 100 million words, 1200s-1900s*. [<http://www.corpusdelespanol.org>]
- CdP – Davies, Mark and Michael Ferreira. (2006-) *Corpus do Português: 45 million words, 1300s-1900s*. [<http://www.corpusdoportugues.org>]
- CRPC – *Corpus de Referência do Português Contemporâneo (versão v2.1, Portugal only)*. Centro de Linguística da Universidade de Lisboa (CLUL). [<http://alfclul.clul.ul.pt/CQPnet/>]
- DRAE – *Diccionario de la Lengua Española. Real Academia Española. 22.^a edición, 2001*. [<http://www.rae.es/obras-academicas/diccionarios/diccionario-de-la-lengua-espanola>]

References

- Bergs, Alexander & Gabriele Diewald (eds.) (2008). *Constructions and Language Change*. Berlin and New York: Mouton de Gruyter.
- Bybee, Joan (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Bybee, Joan (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan & David Eddington (2006). A usage-based approach to Spanish verbs of ‘becoming’. *Language* 82 (2), pp. 323-355.
- Diewald, Gabriele (2002). A model for relevant types of contexts in grammaticalization. In: Ilse Wischer & Gabriele Diewald (eds.) *New Reflections on Grammaticalization*. Amsterdam and Philadelphia: John Benjamins, pp. 103-120.
- Goldberg, Adele E. (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago and London: University of Chicago Press.

- Grice, H. P. (1989). Logic and Conversation. In: H. P. Grice. 1989. *Studies in the Way of Words*. Cambridge, MA, and London: Harvard University Press, pp. 22-40.
- Heine, Bernd (2002). On the role of context in grammaticalization. In: Gabriele Diewald & Ilse Wischer (eds.) *New Reflections on Grammaticalization*. Amsterdam and Philadelphia: John Benjamins, pp. 83-101.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer (1991). *Grammaticalization. A Conceptual Framework*. Chicago and London: The University of Chicago Press.
- Hopper, Paul J. & Elizabeth C. Traugott (2003² [1993]). *Grammaticalization*. Cambridge: Cambridge University Press.
- Keller, Rudi (1994). *On language change. The invisible hand in language*. London and New York: Routledge.
- Luraghi, Silvia (2010). Causes of Language Change. In: Silvia Luraghi and Vit Bubenik (eds.). 2010. *The Continuum Companion to Historical Linguistics*. London and New York: Continuum, pp. 358-370.