

Carlos Alexandre Gonçalves

A functional/cognitive approach to *hápx* and *quasi-hápx* in Brazilian Portuguese suffixation

Abstract: This paper attempts to show, through a set of almost two hundred data collected from various sources, that the phenomenon of *hápx legomenon* also appears in the structure of complex words, especially suffixed words. Anchored in usage-based models, such as Bybee (2010), the work suggests a scale of suffixal unproductivity, ranging from what we call suffixal *hápx*a (suffixes with type 1 frequency), such as *-é* for *maré*, ‘tide’, to the more common unproductive suffixes, such as *-idão*, for *mansidão*, ‘meekness’, for example. Along the way, there are the elements we call quasi-*hápx* suffixes, those that form a very limited family of cognates, such as *-onho*, with only four tokens: *risonho*, ‘laughing’, *tristonho*, ‘a little sad’, *medondo*, ‘scary’, and *enfadonho*, ‘tedious’. The chapter also demonstrates the power of analogy in word creation, as this cognitive ability can lead to the formation of mirror words, blurring the line between *hápx legomenon* and productivity.

Keywords: morphology; *hápx legomenon*; (un)productivity; usage-based models

1 Introduction

Based on the seminal work of Gonçalves *et al.* (2022), this chapter widens the discussion about so-called *hápx legomenon*, a nomenclature firstly introduced by Bauer (2001) for isolated morphological formations. This is the situation with the word *casebre*, ‘small, simple house’, which is easily related to *casa*, ‘house’, and can, therefore, be broken down into *cas-*, a root, and *-ebre*, an exclusive formative of this word.

This chapter is not an English version of the text written with my undergraduate research fellows, whom I would like to thank here¹. There are several reasons

¹ I would like to thank former students Marcos Antônio Gomes Filho and Sarah Batista Santos, who are now Portuguese teachers, for the enormous amount of work they put into collecting the data and for their interesting insights into the analysis.

Carlos Alexandre Gonçalves, UFRJ (Federal University of Rio de Janeiro) and researcher of CNPq (Nacional Council for development of Research), e-mail: carlosvictorio@letras.ufrj.br

for this: (a) there is a whole section on the history of the term *hápx legomenon* in the literature, including its adoption by morphologists; (b) many new examples have been included in this analysis; and (c) the theoretical assumptions are not the same.

Our analysis suggests the existence of *hápx* suffixes (HS) and *quasi-hápx* suffixes (QHS). In very general terms, the former only attach to one root (*lingu-ajar*, *língu-*, ‘language’ + *-ajar*, ‘way of speaking characteristic of an individual or a group’, while the others attach to a very small number of roots (*di-urno*, *not-urno*, *di-*, ‘day’; *not-*, ‘night’, ‘what happens during the day or night, respectively’. The research is supported in the so-called usage-based models, especially in the works of Bybee (2010); Fauconnier; Turner (2002) and Traugott and Trousdale (2013). The central notions used here are, in particular, construction, productivity and analogy. Since we propose an extension of the concept of *hápx legomenon* to the morphological structures of Portuguese, we describe more closely the morphological elements that, in suffix position, constitute unique, isolated sequences.

2 Methodology

Following Gonçalves *et al.* (2022), our data-base were extracted from electronic dictionaries, such as *Dicio* and *Priberam*, etymological dictionaries (Cunha, 1975; Michaellis, 2006) and traditional grammars, such as those by Luft (1969) and Cunha (1970). Moreover, we used advanced online search tools such as *Palavras.net* and *Palavrasque.com*, which allowed us to find all the other words structured by the formatives here studied by searching for a given sequence in ending position, through the search tools available on these websites. After that, in order to prove that the data collected are not archaisms, because they are always analyzable, in Bybee’s terminology², because they have a high degree of compositionality (Bybee, 2010), we tracked the complex words with the *hápx* suffix (HS) and *quasi-hápx* suffix (QHS) in the search tools of the social network X (old Twitter) and collected real interaction data to exemplify the current uses.

Twitter (nowadays X) is an environment rich in verbal language in which users from all over the world create millions of tweets a day, making them a valuable source of data for linguistic analysis. This communication can be a reproduction of an informal conversation, with ordinary users talking to each other, or an institutional channel, where formal and cultured language is used. Overall, there is a

² Compositionality is a semantic parameter and refers to the degree to which the meaning of the whole is predictable from the meaning of its component parts. Analyzability is a morphosyntactic parameter for recognizing these parts.

convergence of various discursive models and genres that reaffirm X as a textual tool. In our X investigation we found several cases, some new, of the morphological units analyzed here, such as *bebum*, ‘very drunk’. This adjective was created from the verbal root *beb-* (*beb-er*), ‘to drink’, plus an ending not found in any other word in the language: *-um*. This new form refers to someone who is exaggeratedly drunk, who gets drunk often or has a tendency to get drunk easily.

Our main aim is to show that these units are far from being interpreted as morphological fossils and are far more numerous than they appear, as we show in the work of 2022. Our *corpus* today has over two hundred formations, which is almost double the amount of work done in 2022.

3 Theoretical framework: Brief remarks

Since much of the theoretical framework used here has already been extensively described in the Presentation, this section presents a very general overview of what are known as usage-based models.

Usage-based models of language involve several models of linguistic analysis within a broader functional/cognitive framework, which emerged in the late 1980s and which presupposes a deep relationship between linguistic structure and usage. The term *usage-based* was coined by Langacker in (1987), in his seminal text, which is one of the most representatives of the Cognitive Grammar. Usage-based language models soon became a significant new trend in linguistics from the early 1990s and among their most influential advocates are Michael Tomasello, Joan Bybee and Adele Goldberg.

Based on these theoretical contributions, it is possible to postulate that the structure of language is forged in experience, both historical and in the current day, and that it derives from general cognitive processes. Thus, the view of language adopted is no longer centered solely on empirically proven tokens, i.e. the linguistic material itself, but on instances of greater abstraction.

Usage-based models emphasize the holistic and contingent approach to linguistic uses, considering that items do not occur or produce meaning in isolation, as contextual relationships need to be considered. This reorientation corresponds to contemporary Functionalism and sees linguistic structure as derived from general cognitive processes, according to Bybee (2010). In this context, linguistic uses are understood as the product of experience, routinization and perspectivization in and through language, among other motivations. From words to idioms, all symbolic units are constructions: form-meaning pairings (Goldberg, 1995).

4 Organization of the manuscript

First of all, we define *hápax legomenon* from its use in Ectodotics to its incorporation into morphological studies. In the sequence, we try to relate the notion of *hápax legomenon* to the lexical (un)productivity issue and thus we propose a distinction between *hápax suffixes* (HS) and *hápax-quasi* suffixes (QHS), showing that the concept of suffixoid, used in Portuguese morphology by Rocha (1998), is not appropriate to this data. Finally, we apply some concepts of usage-based models to Portuguese data, explaining, finally, the power of analogy in the creations of schemas (constructions).

5 The notion of *hápax legomenon* (HL)

Hápax Legomenon (ἅπαξ λεγόμενον), plural *legomena*, is a Greek expression that literally means ‘being said once’. It has come to refer to a term that is used only once in a given context, whether it is the works of a particular author, a particular work of literature, or even within all the known writings of a particular language. For instance, The Bible contains hundreds of *hápax legomena*, that is, words that only appear once in this sacred text. This expression became widely used in Ectodotics, a branch of philology whose task is to edit texts as faithfully as possible to the original or to the author’s will, mainly aiming at the elimination of transcription errors. According to the website Gotquestions, “There are only a small handful of words in the Bible that are completely unknown in other literature. There are about 400 such *hápax legomena* that appear in the Old Testament and about 25 in the New Testament” (<https://www.gotquestions.org/hapax-legomenon.html>). In many cases, such *hápax* can be considered *ex-nihilo* formations, words without any apparent motivation, semantic or formal, such as ‘parathalassios’, a *hápax legomenon* used in Matthew 4:13 and nowhere else in the New Testament.

As many of their origins are obscure and their etymology is unknown, it is very difficult to talk about *ex-nihilo* formation. As Gonçalves (2019) shows, an *ex-nihilo* formation constitutes the creation of a new root and, in the same way as Szymanek (2005), he doesn’t believe that this happens very often and he asks himself whether there really is a formation created out of nothing.

Generally speaking, root creation is motivated (there can always be a reason for a new coinage) and is rarely *ad hoc* or echoic, since the new form usually resembles one or more pre-existing forms or comes from borrowings, acronyms, abbreviations or even inversions in the order of the segments, such as the current slang *yag* for ‘gay’. In a real *ex-nihilo* root-creation, however, there appears to be no lexico-

logical way of accounting for the formation of a word: it has no known precursors, as with the trade name ‘Kodak’ (invented in the US in 1888 by George Eastman), as observes Baayen and Lieben (1991). In other words, there is no direct relationship between *hapax* and *ex-nihilo* formations, although the phenomena may coincide in some cases, because, from a functional/cognitive perspective, there is no space for arbitrariness (Fauconnier; Turner, 2002): all formations are motivated, even if, in some cases, we may not know the motivation.

Morphologists soon appropriated the term *hapax legomenon* to characterize unproductive formations. Thus, in morphology, *hapax legomenon* is always described when addressing the issue of lexical (un)productivity, since it can involve complex formations, some based on (1) unexpected suffixation, (2) morphological reanalyzes or (3) non-concatenative processes, such as lexical blends. All examples are from literary texts written in Portuguese:

1. Chuv-inh-enta (Carlos Drummond de Andrade)
chuv(a) (rain) + Diminutive + Evaluation = ‘intermittent and annoying rain’
2. Enxada-chim (Guimarães Rosa) << espada-chim (espada (sword) + *-chim*
Enxada (roe) + termination without clear content = ‘rural worker who handles the hoe like a swordsman’
3. Sonolent-idão (Mia Couto) = ‘sleepy slowness’
Sonolent(o) (drowsy) + **lentidão** (slowness). The parts in bold belong to the two bases; they are ambimorphemics.

Hay and Baayen (2002) show that the *hapax* is an important measure of lexical productivity since it exemplify many complex words with a single occurrence in several natural languages. Pierrehumbert and Granell (2015) take a brief look at how the notion of *hapax* has been used by morphologists in the last two decades. In general, the authors almost always try to explain the etymology and check which word-formation processes make up *hapax*. For example, justifications for the use of *hapax* are proposed in Baayen and Renouf (1996). Hay and Baayen (2002), in turn, seek to observe what processes are at play: compounding, derivation, onomatopoeia, acronym, abbreviation, anagram, among others. The nature of *hapax*, however, is still little discussed in the literature, which is why we intend to contribute to studies in the area.

One of the peculiarities of *hapax* formations is, of course, unproductivity. A *hapax legomenon*, being an isolated construction, is subject to very low application and consequent fading of use over time, often becoming an archaism (Lieber, 1991). This theme, however, is not much discussed in the literature, since lexical

constructions of great recurrence in the language are most commonly analyzed in the sphere of word formation.

Evidently, in order to determine what an unproductive morphological process consists of, it is essential to consider what is commented on about productivity, for the purpose to establish, in opposition, what are the fundamental aspects of the other extreme, considering productivity, at this point, as profitability (Corbin, 1987). In this case, productivity is something that can be measured and put on a scale.

Productivity is one of the most subjective and controversial notion in linguistics³, even being held as “a matter of some dispute” (Bauer, 2001: 10); it can be a reflection of the “creativity that allows one to say something that has never been said before” (Hockett, 1958: 34); of the “current availability of an element or process” (Ralli, 2010, p. 245); of “new lexical formations, regardless of the volumes of words they generate” (Gonçalves, 2019: 114).

In this chapter, because of the functional/cognitive orientation, we consider productivity to be a gradient parameter and it is related to the expansion or restriction of the (sub)schemas of a construction. Productivity has been closely related to the token and type frequencies (Bybee, 2010) of a given grammar construction. Token frequency refers to the number of times the same unit occurs, while type frequency refers to the number of expressions of a particular pattern, i.e. the types of elements that fill the slots of a construction.

Traugott and Trousdale (2013) draw a parallel between token and type frequency and add that an increase in the range of constructs is a sign of increased productivity. At the other extreme, the low frequency of a certain construction can contribute to its disappearance. For this reason, we have only worked with words where the base is perfectly recognizable, despite the opacity of the formative to its right. For this reason, we turned to electronic dictionaries to find words with the endings here analyzed. With this procedure, for *-ebre*, for *casebre*, ‘poor house’, we discarded atomic words like *lebre*, ‘hare’, and *febre*, ‘fever’, or even some apparently complex ones, like *célebre*, ‘famous’, whose accent does not fall on the suffix.

6 Lexical hápax and morphological hápax

In this chapter, we distinguish between two kinds of *hápax*: (a) *lexical hápax* and (b) *morphological hápax*. The former, as the name suggests, affects words, while the latter involves affixes that fail in one of the main characteristics of linguistics

³ We prefer to use the term “notion”, rather than concept, since “concept” at all, since the latter refers to a number of different concepts, some related to frequency, some to availability, newness etc.

forms: recurrence (Bloomfield, 1933). Due to the greater variety of data in Portuguese, we will focus on cases where the non-recurring element appears on the right edge of the word, thus behaving like a suffix. How should we refer to these elements? Are they in fact suffixes or can they be called *suffixoids*, as Rocha (1998) does? If we consider the position, they are obviously suffixes, but different suffixes, which appear combined with a single root.

Suffixoids are generally defined as “a wordfinal segment that has characteristics of both a free morpheme and a bound morpheme” (Marchand, 1969: 171). The same idea appears in the following quotation: “The term *suffixoid* is often used for free morphemes taking on suffix-like properties” (Lightfoot, 2011, p. 75). This is undoubtedly the case with the Brazilian Portuguese *X-metro* and *X-mania* constructions, as ‘metro’ and ‘mania’ can be used as free forms (words) or bound forms of compounds:

4. Maria tem **mania** de louça limpa.
“Mart has a thing for clean dishes”
5. Maria tem brin**comania** (*brinco* ‘earring’ + *-mania* ‘compulsion’)
“Mary has a thing for wearing earrings”
6. Maria comprou um **metro** de seda.
“Mary bought a meter of silk”
7. O **bafômetro** pegou Jonh novamente (*bafo* ‘puff’ ~ ‘smell of alcohol’ + *- metro*, ‘device’ = “The device that measures alcohol content caught John again”

The situation of *-metro* and *-mania* is very different from forms such as *-aréu*, *-anzil* or *-oz*, which have no link to any free form of the language. As they are attached after a root, these elements can be considered suffixes. However, they display a peculiar behavior: they are exclusive to complex words which, in Bybee’s (2010) terms, are analyzable, because their bases have a degree of transparency. In fact, it is possible to recognize morphological structure in words like these because the transparency of its roots: *fog-o*, ‘fire’, in *fogaréu*, ‘excessive fire’, *corp-o*, ‘body’, in *corpanzil*, ‘sculptural body’), and *fer-a*, ‘beast’, in *feroz*, ‘ferocious’. In the absence of a nomenclature for these particles, which are abundant in Portuguese, we have resorted to the term *hápax* suffix, a suffix exclusive to a single complex word in the language.

Even though HS occupy the post-base position and share several similarities with genuine representatives of the category *suffix*, such as the fact that they depend on the base to express their content, as well as the ability to change class, as

we'll show further ahead. These morphological units do not prototypically belong to the category of suffixes, since they are neither prolific nor available (Bauer, 2001)⁴. Suffixation, in turn, is a morphological process that affects a large number of words and is generally available for the creation of new lexical items, as is the case with the vast majority of suffixation cases in Portuguese, especially those that change the lexical label of the base.

However, it's not always easy to immediately tax a morphological process as productive/unproductive or simply to reduce productivity to factors such as frequency, for example (Bauer, 2001). The productivity of “marginal” morphological processes is still productivity, just on a low scale (Traugott; Trousdale, 2013: 57). A HS occupies the lowest level of productivity, since its occurrence is 1 (one), attaching itself to one – and only one – base (root or word) of the language, in expression and meaning (Gonçalves, 2019). In the table below, we list some unique units of a single word. For reasons of space, we've selected only the most representative formatives and, therefore, we also have not included the cases already mentioned. We have made it clear that only cases where the basis is transparent are taken into account. For example, in addition to *bocarra*, ‘very big mouth’, there are other words ending in *-arra* in Portuguese, but none of them are analyzable because they don't have a base that can be linked to a root or a word, as is the case of *guitarra*, ‘guitar’ and *gambiarra*, ‘gimmick’, among others.

Table 1: Main cases of HS in Portuguese. Source: own elaboration.

Complex word	root	Hápax suffix	base/product category	meaning
<i>andarrilho</i>	andar 'to walk'	<i>-ilho</i>	V → N	'someone who walks a lot, with no destination'
<i>bocarra</i>	boc- 'mouth'	<i>-arra</i>	N → N	'very big mouth'
<i>boliche</i>	bol- 'ball'	<i>-iche</i>	N → N	'bowling'
<i>chatonildo</i>	chaton- 'boring'	<i>-onildo</i>	Adj → Adj	'someone too boring'

⁴ Bauer (2001) chooses to divide the concept of productivity into two forms: on the one hand, availability, which relates to the system and makes a particular morphological process or affix available; on the other hand, profitability, which relates to the norm and is linked to the number of occurrences. In the second situation, according to the author, there is the possibility that a given affix may appear in a large number of derivative products.

Table 1 (continued)

Complex word	root	Hápx suffix	base/product category	meaning
<i>choramingar</i>	chorar 'to cry'	<i>-mingar</i>	V → V	'to cry too much, all the time'
<i>esconderijo</i>	esconder 'to hide'	<i>-ijo</i>	V → N	'a place where something is hidden'
<i>felizardo</i>	feliz 'happy'	<i>-ardo</i>	Adj → Adj	'very fortunate'
<i>ferrolho</i>	ferr- 'iron'	<i>-olho</i>	N → N	'bolt'
<i>jazigo</i>	jaz- 'to die'	<i>-igo</i>	V → N	'grave'
<i>linguajar</i>	ling- 'language'	<i>-ajar</i>	N → N	'kind of language'
<i>longínquo</i>	long- 'far'	<i>-inquo</i>	Adv → Adj	'remote'
<i>mamilo</i>	mam- 'breast'	<i>-ilo</i>	N → N	'nipple'
<i>moçoila</i>	moç- 'young woman'	<i>-oila</i>	Adj → Adj	'young girl'
<i>nevasca</i>	nev- 'snow'	<i>-asca</i>	N → N	'blizzard'
<i>pelanca</i>	pel- 'skin'	<i>-anca</i>	N → N	'wrinkled skin'
<i>pedregulho</i>	pedr- 'stone'	<i>-egulho</i>	N → N	'boulder'
<i>rabiola</i>	rab- 'tail'	<i>-iola</i>	N → N	'kite'
<i>rapariga</i>	rapa(z)- 'young man'	<i>-iga</i>	N → Adj	'young girl'; 'feminine of young man'

As can be seen, most of the HS cases involve semantic lexicalization, presenting idiosyncratic meanings, and have a partially compositional interpretation. All the examples can be considered representative cases of the HS class, since all the endings are unique and exclusive to each word in Table 1. These cases are abundant in Portuguese and cannot be considered isolated. The Table 1 is a good example of many formatives with the same status as *-anzil*: those which, in the suffix position, occur in single, isolated words and therefore lack recurrence.

Two interesting cases of HS are those involving attachment to the same lexical base, as exemplified below. Even though some words have entered the language

by borrowing, there is no denying the morphosemantic correlation between the groupings⁵:

8. a. -ujo *mar-ujo* ‘sailor’
 -esia *mar-esia* ‘sea air’
 -isco *mar-isco* ‘seafood’
 -é *mar-é* ‘tide’
 ‘sea’
9. b. -oça *carr-oça* ‘wagon’
 -ossel *carr-ossel* ‘carousel’
 -uagem *carr-uagem* ‘carriage’
 ‘car’

Still thinking about gradualness, there are many suffixes in Portuguese, which, although they are not *hápax*, appear in combination with a very small number of bases. For this reason, reinforcing the relationship between *hápax legom-enon* and productivity, Gonçalves *et al.* (2022) used the term *quasi-hápax* suffixes (QHS) to label units like *-onho*, linked to only four roots. In all of them, a name became an adjective and its meaning can be generically referred like as ‘X (someone, in 10a and 10b, or something or someone in 10c and 10d) has the appearance of X’.

10. a. *ris-onho* – ‘laughing’
 ‘smile’
 b. *trist-onho* – ‘a little sad’
 ‘sad’
 c. *med-onho* – ‘dreadful’
 ‘fear’
 d. *enfad-onho* – ‘boring’
 ‘tedious’

5 There is no way to deny the relational link between the words in (9) because of the high transparency of the base. On the other hand, the meanings that the so-called suffixes carry are too specialized and, therefore, are far from creating word series, resulting in a productive pattern and perhaps because (9b) are no longer so frequent at the current stage of the language, at least the last two. Even remotely, the formations refer to *carr-* ‘carr-’ (‘car’) they preserve, from this lexical item, one or another property, such as, for example, transport, the conduction of passengers and mobility itself.

The case of *X-onho* constructions is not isolated in the Portuguese: there are a number of other suffixes that form very small families of cognates. If we extend the notion of HL a little further, we arrive at *quasi-hápx* suffix (QHS): a rarely recurring element that applies to very few lexical units in the language, having very low token frequency. As Gonçalves *et al.* (2022: 165) note, “they resemble *hápx* suffixal, (a) by being of obscure and/or poorly accessible etymological origin, (b) adjoining transparent bases, and, albeit to a lesser extent, (c) belonging to the sphere of unproductivity”. The Table 2 (below) lists some of these elements and the number of complex words in the language. Note that the frequency of types varies from 2 to 10.

Table 2: Main cases of QHS in Portuguese. Source: own elaboration.

Hápx suffix	types	Complex words	base/product category	meaning
-uço	2	dent-uço ‘tooth’ ping-uço ‘alcohol’	N → Adj	‘who has forward teeth’ ‘who drinks too much’
-aca	2	velh-aca ‘old person’ fri-aca ‘cold’	Adj → Adj Adj → N	‘deceptive woman’ ‘excessive cold’
-urno	3	di-urno ‘day’ not-urno ‘night’ tacit-urno ‘implicit’	N → Adj	‘during the day’ ‘during the night’ ‘gloomy’
-icho/a	3	barb-icha ‘beard’ rab-icho ‘tail’ lagart-ixa ‘lizard’	N → N	‘small beard’ ‘small tail’ ‘gecko’
-iz	3	aprend-iz ‘to learn’ ‘chamar-iz’ ‘to call’ pet-iz ⁶ ‘to request’	V → N	‘the one who learns’ ‘the one who call’ ‘the one who request’

⁶ Although the root of ‘petiz’ (“beggar”) is a *doublet*, according to Soledade (2005), there are strong semantic and phonological connections between *pet-* and *ped-*: the presence of /t/ in place of /d/ is not enough to make the formation opaque.

Table 2 (continued)

Hápax suffix	types	Complex words	base/product category	meaning
-az	5	fug-az 'to escape' viv-az 'to live' compr-az 'to buy' costum-az 'to used to' mord-az 'to bite'	V → Adj	'fleeting' 'vivacious' 'shopper' 'habitual' 'ferine'
-ucho/a	5–10	gord-ucho 'fat' pequerr-ucho 'small' papel-ucho 'paper' cart-ucho 'letter' fof-ucho 'cute'	Adj → Adj	'low-fat' 'too little' 'small paper' 'cartridge' 'very cute'
-engo/a	5–10	mulher-engu 'woman' monstr-engu 'monster' mol-engu 'mushy' andar-engu 'to walk' fuderengu 'to fuck'	N → Adj V → Adj	'womanizer' 'monstrous' 'too soft' 'who walks a lot' 'who fucks a lot'
-acho	5–10	ri-acho 'river' pen-acho 'bird feather' fog-acho 'fire' popul-acho 'people' quebr-acho 'to brake' mingu-acho 'wane'	N → N V → N	'small river' 'plume' 'small fire' 'popular people' 'axe-breaking' 'gourd with water'

Considering that the notion of productivity is gradual (Traugott; Trousdale, 2013), we could propose the same for unproductivity, based on the notion of *hapax* applied to morphology: in Portuguese, there is a scale of unproductivity that starts with HS, with token frequency 1, passing through QHS, whose frequency varies from 2 to 10 constructs (lexical items), until reaching the unproductive suffixes of more general use in the language. Undoubtedly, there are unproductive suffixes with a greater number of types than others. For example, the suffix *-idão*, analyzed in Basilio (1980) and in Rio-Torto (1988), is considered unproductive for not forming new words nowadays. However, it accounts for a very large contingent of derivatives (around 200), unlike, for example, *-ejo*, which has no more than fifteen analyzable examples:

- | | |
|----------------------------------|---------------------------|
| (11). cort-ejo ‘cortege’ | lugar-ejo ‘small town’ |
| ‘court’ | ‘place’ |
| grac-ejo ‘joke’ | azul-ejo ‘tile’ |
| ‘fun’ | ‘blue’ |
| sertan-ejo ‘who lives in sertão’ | fest-ejo ‘celebration’ |
| ‘sertão’ | ‘party’ |
| benfaz-ejo ‘that’s good for you’ | andejo ‘walker’ |
| ‘to make well’ | ‘to walk’ |
| vilar-ejo ‘small village’ | animal-ejo ‘small animal’ |
| ‘village’ | ‘animal’ |
| real-ejo ‘who makes predictions’ | gargar-ejo ‘gargling’ |
| royal | ‘throat’ |
| malfaz-ejo ‘that’s bad for you’ | sacol-ejo ‘jerk’ |
| ‘to make bad’ | ‘bounce’ |
| murmur-ejo ‘mumble’ | |
| ‘whisper’ | |

In the Figure 1, we propose a scale of unproductivity for suffixes in Portuguese. Note that at the beginning of the scale are the HS, suffixes with type 1 frequency. The most common unproductive suffixes are at the top of the scale, while QHS and less common unproductive suffixes are in the middle.

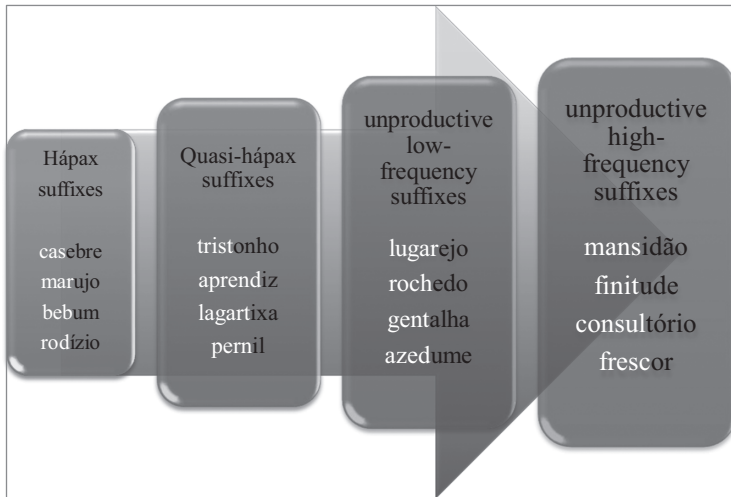


Figure 1: The continuum of unproductivity in suffixation. Source: own elaboration.

7 Hápax suffix and productivity

Contrary to what Rocha (1998) observes, the inventory of HS is not closed: we create nowadays these morphological units, which shows that the phenomenon itself⁷ is productive. In fact, formations like *beb-um*, ‘very drunk’, *chato-nildo*, ‘very tedious’ and – why not? – *sanduba*, ‘sandwich’, are quite recent in the language. This latter lexical item differs from the formers only in that it presupposes a shortening (*sand-*) prior to the adjunction of the non-recurring affix. Still, all of them are HS cases.

The set of morphological units that resemble those listed Table 1 is far from small and, more importantly, the inventory of these particles can expand with the inclusion of new members. In addition to those listed above, formations like *burr-aldo*, ‘very stupid’, and *natur-eba*, ‘natural food advocate’ are brand new. There is no doubt that the relationship between words with HS is horizontal: one construction (simple or complex) is anchored in another, so as to favor some compositionality in meaning and full conditions of isolability of the parts. The situation of *natureba* is very representative because the root *natur-* appears in complex constructions like *natur-al*, ‘natural’, and *natur-eza*, ‘nature’, even though the form *natura* is unusual:

⁷ When we say that the phenomenon itself is productive, we mean that the suffixation of a single word is not marginal in the language.

it appears only in the Latin expression *in natura*. Another situation very common with HS is its passage to the QHS group through the action of analogy.

8 HS and analogy

Let us now turn to the analysis of *chatonildo*, ‘a very, very boring person’. In this lexical item, appears a bounded root *chat-*, the same found in *chat-inho*, ‘a bit boring’, *chat-ear*, ‘to annoy’, and *chat-ice*, ‘bummer’, for example. However, the suffix *-nildo* is exclusive to this morphologically complex unit, and there is no other derived word in Portuguese that has this ending as an affix, unless proper names, like ‘Hermenegildo’, obviously non-analyzable and non-compositional. Even if we thought that this formation passed through *chaton-*, a possible variant of *chatão*, ‘very boring’, with the augmentative suffix *-ão*, which alternates with *-on* in derivatives, we would still have a unit without recurrence: *-ildo*, also found in proper names, like ‘Gildo’ and ‘Hildo’, again neither analyzable nor compositional. In turn, the relationship between the word ‘chato’ and ‘chatonildo’ may be an indication that the unit in question is actually *-nildo*.

Chatonildo is easier to learn, then, because it has a previously known part; and it is easier to process, because of the extra activation that comes from *chato*, a complete word. An interesting fact about SH is that they can be involved in analogical creations. A very recent formation anchored in *chatonildo* is *alienildo*, a word that is also anchored in *alien-ado*, ‘alienated’. As can see in the following meme⁸, this construction certainly builds on *chatonildo* and establishes a new relational link with it: *alienildo* seemed of *chatonildo*, which refers to a very boring person. However, the *alienildo* was never interested in politics, much less in Brazilian history, but he is now an “engaged” (sic!) person and spreads fake news through a single vehicle: Bolsonaro what’s up groups:

This example shows that there is no rigid boundary between HS and productivity, which allows us to state that there are no unbridgeable borders between the completely unproductive and the productive, because speakers can use isolated words to express a point of view. The *Alienildo* in Figure 2 is an excellent example of this: it is a spontaneous formation that portrays the profile of the current Brazilian followers of the former president of Brazil. According to Bauer (1988: 96), morpho-

⁸ First phrase: This is Alienildo. Alienildo was never interested in politics or Brazilian history. Second phrase: But now Alienildo is a committed guy. He spreads Bolsonaro fake news and he is against the “little flu” social isolation measures; he says that the OMS is communist.



Figure 2: Alienildo⁹. Access: march 10, 2024.

logical constructions can be “clearly modeled on an existing complex word, not giving rise to a productive series”. Such is the case of *Alienildo*, mirror formation of *chatonildo*.

Perhaps it is the mechanism of analogy that accounts for the only transparent cases of words ending in *-urno*: *diurno*, ‘daytime’, and *noturno*, ‘nighttime’. One form supports the other; they are “twins” that feed off each other. In Cognitive Linguistics, the human conceptual system has great symbolic potential for constructing meanings. According to Fauconnier and Turner (2002: 27), this is possible due to the operations carried out by the three I’s of the mind:

Identity (perception of equivalence and oppositions between concrete or abstract things); Integration (complex process that has dynamic and structural properties, of rapid operation for the purpose of categorization) and Imagination (simulations, such as: fiction, dreams, hypothetical scenarios, fantasies)”. In English, the form is more interesting because I’s is pronounced ‘eyes’, referring to the “eyes” of the mind. Fauconnier; Turner (2002: 27).

In this section, we will deal with some cases of analogy, conceived here as one of the fundamental principles of cognitive linguistics (Langacker, 1987). For Fauconnier and Turner (2002), this principle acts when the speaker needs to explain a certain element on the basis of another, already known element. In this way, it is a first-or-

⁹ Source: <https://www.facebook.com/cidadao.riosulenses/photos/a.282357075169682/2171797196225651/?type=3>

der principle that facilitates the acquisition and recognition of a new domain, since it “allows a parallelism to be made between elements with different meanings with the aim of verifying and making a new element in the world emerge, categorized in a more appropriate way” (Furtado, 2011: 54).

We have been drawing attention to the fact that SH and QHS are characterized by high semantic density, in Ralli (2010)’s terms, which prevents them from forming series of words. Undoubtedly, since they became so specialized semantically, that they do not present sufficient semantic coherence (Aronoff, 1976) to be applied on a large scale, hence their unproductivity. However, some of them can be accessed by the speakers to create words via analogy. In fact, it is not uncommon to find cases of SH and QHS functioning as the analogical item of mirror creations, as in the example of *Alienildo* from the previous section:

11. a. natureba (“who eats natural food”)
 ↓
 mixtureba (“overmixed”)
 decoreba (“memorize mechanically”)
- b. fumacê (“heavy smoke”)
 ↓
 lamacê (“sludge”)
 miserê (“extreme poverty”)

Other units create very small families of complex items that can create schemas (Booij, 2010). A good example in Portuguese is the small group of nouns ending in *-aréu*. As seen in 12, there is a new form-meaning pairing, which shows that, albeit incipiently, there has been constructionalization: nouns can be created by *-aréu* whose meaning is “large amount of X”.

12. **-aréu** *fogaréu* ‘big fire’
 povaréu ‘crowd of people’
 aguaréu ‘excessive water’

We can formulate a scheme to account for these forms in Portuguese. This scheme is not yet so well established in the language and that it’s hard to say how much it can create new words. In a quick search for wiki dictionaries, those that the next users can include entries, however, we found some forms that don’t appear in the most authoritative dictionaries, which attest to its demand, when talking about a large quantity of something or someone:

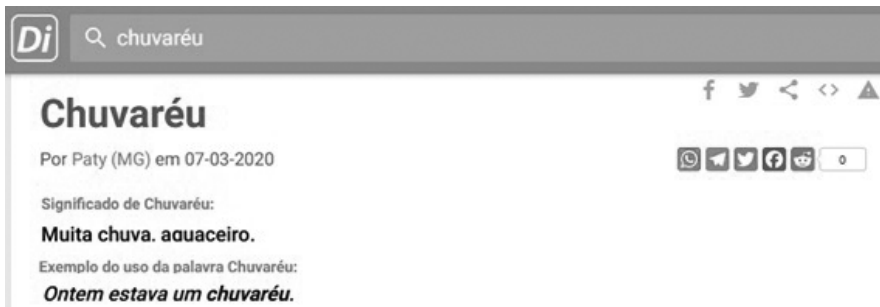


Figure 3: Chubaréu. Source: <https://www.dicionarioinformal.com.br/significado/chubar%C3%A9u/42032/>. Access: 10 march 2024.



Figure 4: Mataréu. Source: <https://www.dicio.com.br/matareu/>. Access: 10 march 2024.

9 Concluding remarks

In this chapter, we hypothesize that morphological suffixation have units belonging to the *hápx* suffixal (HS) and *quasi-hápx* suffixal (QH) categories, which once functioned as synonyms for other formatives, have, over time, lost their ability to compete with more usual formatives due to the more specific content expressed by them, leading them to a very high degree of semantic density (Ralli, 2010), which has culminated in their unproductivity. Moreover, the subjectivity, expressing feelings, emotions and points of view of the speakers, drastically reduces the context

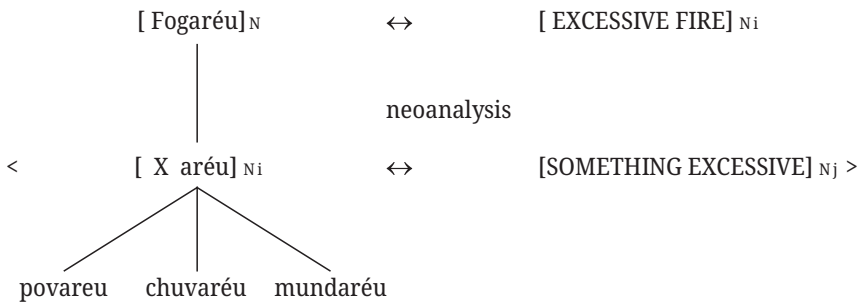
AU: Please provide a better quality image for this figures 3, 4.

of use of these formatives to more specific situations. This confirms Basilio's (1997) proposition that contexts can interdict or favor the use of affixes, which, in turn, can condemn these formatives to unproductivity, since there is a direct relationship between semantic coherence and serial application, as predicted by Aronoff (1976).

We propose a scale of unproductivity for suffixes in Portuguese. This *continuum* begins with type 1 frequency suffixes, those exclusive to a single word in the language (*-ebre*, *cas-(a)*, 'house', *casebre*, 'poor house'). At the other end of the continuum are the unproductive suffixes that have become large lexical families, such as *-idão* (*ampl-(o)*, 'wide', *ampl-idão*, 'vastness'), whose list of derivatives, according to Michaelis' search tool (<https://michaelis.uol.com.br/>), is over 200 words long. QHS and less common unproductive suffixes are in the middle in the scale. The difference between them is that the former has up to ten examples (*-ê*, *miser-*, 'misery'; *miser-ê*, 'poverty'), while the latter varies much more in type frequency.

Another point discussed here was the role of analogy in the expansion of both HS and HQS. In fact, new words can emerge by mirroring others, and can even lead to constructionalization, as is the case with *-aréu*. The following diagram shows how the *X-aréu* constructions are being created. The neoanalysis of *fogaréu* and the its high token frequency led to the *X-aréu* constructionalization, since the root of *fogo*, 'fire', *fog-*, created an empty slot, allowing other items to fill this position.

When a neoanalysis becomes conventionalized in the speaker population, a new microconstruction is created (Traugott; Trousdale, 2013), as can be seen in the following representation, in which a schema [[X]_{s i} aréu]_j is created, which accounts for new formations, such as the constructs *povaréu*, 'a lot of people', and *homaréu*, 'a lot of man':



In this representation, made along the lines of constructional morphology (Booij, 2010), SEM is interpreted as the frame evoked by the source word, *foraréu*. The symbols greater than and less than (respectively, <, >) demarcate the new scheme and the two-way arrow (↔) relates form and meaning within each line, which symbolizes that they are all constructions, i.e. form-meaning-function pairings. In the

second line, neo-analysis operates, responsible for the dismemberment of *-aréu*. That's exactly the point we have a new form-meaning-function pairing and, consequently, a new schema, in which N represents the lexical label of the base and the product (a Noun, in both cases). The subscripts *i* and *j* indicate that both forms are part of the lexicon.

Without wishing to exhaust the subject, we hope to have provided the reader a different perspective on morphological analysis, firstly by showing that there are very special suffixes in Portuguese, exclusive to a single word. Secondly, we also hope to have shown that there is no unbridgeable limit between units with a type frequency 1 and unproductivity, since analogy can lead to the creation of new forms by mirroring another.

References

- Aronoff, Mark. 1976. *Word formation in Generative Grammar*. Cambridge MA, MIT Press.
- Baayen, Harald and Renouf, Antoinette. 1996. Chronically the times: productive lexical innovations in an English newspaper. *Language* 72 (1), 69–96.
- Baayen, Harold and Lieber, Rochelle. 1991. Productivity and English derivation: a corpus-based study. *Linguistics*, 29 (5), 801–844.
- Basílio, Margarida. 1980. *Estruturas Lexicais do português*. Petrópolis: vozes.
- Basílio, Margarida. 1997. O princípio da analogia na constituição do léxico: regras são clichês lexicais. *Veredas*, 7 (1), 9–21.
- Bauer, Laurie. 1998. *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.
- Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart & Winston.
- Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Max Niemeyer.
- Cunha, Antônio Geraldo da. 1975. *Dicionário etimológico da língua portuguesa*. Rio de Janeiro: Nova Fronteira.
- Cunha, Celso Ferreira da. 1970. *Gramática do português contemporâneo*. Belo Horizonte: Bernardo Álvares.
- Fauconnier, Gilles and Turner, Mark. 2002. *The way we think*. New Yorker: Basic Books.
- Furtado, Lilian Ribeiro. 2011. *Análise semântico-cognitiva das substituições sublexicais em português*. Rio de Janeiro: UFRJ.
- Goldberg, Adele. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago; London: The University of Chicago Press.
- Gonçalves, Carlos Alexandre. 2019. *Morfologia*. São Paulo: Parábola.
- Gonçalves, Carlos Alexandre, Gomes Filho, Marcos Antônio and Santos, Sarah Batista. 2022. Sobre bebuns, nudes e naturebas: um estudo sobre hápax sufixal e quase-hápax à luz da Morfologia Relacional. In: Soledade, Juliana, Gonçalves, Carlos Alexandre and Simões Neto, Natival (eds.). *Morfologia Construcional: avanços em língua portuguesa*. Salvador: EDUFBA, 351–369.

- Hay, Jay and Baayen, R Harald. 2002. Parsing and productivity. In Booij, Geert and Van Marle, Jonh (eds), *Yearbook of Morphology*. Dordrecht: Kluwer Academic Publishers, 203–235.
- Hockett, Charles. 1958. *A Course in Modern Linguistics*. The Macmillan Company: New York.
- Langacker, Ronald. 1987. *Foundations of Cognitive Grammar*. Stanford: Stanford University Press.
- Lieber, Rochelle 1991. Productivity and english derivation: A corpus-based study. *Linguistics*, 29 (5):801–844.
- Lightfoot, Douglas. 2011. Testing the suffixoidization of German -mann ‘-man’. *Linguistica* 51 (1.1), 273–283.
- Luft, Celso Pedro. 1969. *Gramática da língua portuguesa*. Porto Alegre: Globo.
- Marchand, Harold. 1969. *The categories and types of present-day English word formation: a synchronic-dia-chronic approach*. Munich: Beck.
- Michaellis, Carolina. 2006. *Moderno dicionário da língua portuguesa*. São Paulo: Melhoramentos.
- Pierrehumbert Janet and Granell, Ramon. 2015. On hapax legomena and morphological productivity. *Yearbook of Morphology*, 24 (1), 203–235.
- Ralli, Angela. 2010. Compounding versus derivation. In: Scalise, Sergio and Vogel, Irene. (eds.) *The Benjamins Handbook of Compounding*. Philadelphia: John Benjamins Publishing Company.
- Rio-Torto, Graça Maria. 1998. *Morfologia derivacional: teoria e aplicação ao português*. Porto: Porto Editora
- Rocha, Luiz Carlos. 1998. *Estruturas morfológicas do português*. Belo Horizonte: Editora da UFMG.
- Soledade, Juliana. 2005. *Semântica morfolexical: contribuições para a descrição do paradigma sufixal do português arcaico*. Salvador: UFBA.
- Szymanek, Bogdan. 2005. The latest trends in english word-formation. In: Štekauer, Paul and Lieber, Rochelle. (eds.). *The Handbook of Word-Formation*. Netherlands: Springer, 429–448.
- Traugott, Elizabeth Closs and Trousdale, Graeme. 2013. *Construcionalization and constructional changes*. Oxford: Oxford University Press.