



ONGOING GRAMMATICALIZATION PROCESS OF THE LEXEME "TIPO" A STUDY BASED ON A SPONTANEOUS SPEECH CORPUS OF BRAZILIAN PORTUGUESE

Saulo Santos

► To cite this version:

Saulo Santos. ONGOING GRAMMATICALIZATION PROCESS OF THE LEXEME "TIPO" A STUDY BASED ON A SPONTANEOUS SPEECH CORPUS OF BRAZILIAN PORTUGUESE. *Revista Inventário*, 2019. hal-04401899

HAL Id: hal-04401899

<https://hal.science/hal-04401899>

Submitted on 18 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

ONGOING GRAMMATICALIZATION PROCESS OF LEXEME “TIPO” A STUDY BASED ON A SPONTANEOUS SPEECH CORPUS OF BRAZILIAN PORTUGUESE

Saulo Mendes Santos
(UFMG - Mestrando)

INFORMAÇÕES SOBRE OS AUTORES
Saulo Mendes Santos possui graduação em Direito pela Faculdade de Estudos Administrativos de Minas Gerais (2015), em Letras/Linguística (2017) pela Universidade Federal de Minas Gerais e cursa atualmente o mestrado em Linguística Teórica e Descritiva, nessa mesma instituição. E-mail: saulo.mendes@gmail.com

RESUMO	ABSTRACT
O lexema TIPO tem sido descrito como um marcador discursivo na literatura (BITTENCOURT, 1999; DAMASIO-LOPES, 2008; CASTELANO; LADEIRA, 2010). No entanto, nem todos os tokens encontrados em um corpus funcionam como marcadores discursivos ou como um item léxico completo. Algumas ocorrências parecem ter um significado gramatical. Este trabalho visa, assim, a apresentar evidências de que o lexema vem passando não apenas por um processo de discursivização, mas também de gramaticalização (HEINE; KUTEVA, 2007). Para tanto, foi utilizado um corpus de fala representativo do português brasileiro, especialmente da diatopia de Minas Gerais (C-ORAL-BRASIL, RASO; MELLO, 2012), sendo todas as ocorrências do lexema TIPO extraídas. Os dados foram então classificados em quatro categorias principais, a saber: item léxico completo, item funcional, unidade tonal contendo apenas o lexema TIPO e unidade tonal contendo apenas a expressão “tipo assim”, de acordo com os traços por cada token apresentado. Os melhores protótipos de cada categoria foram apresentados, os dados foram tabulados e as frequências foram calculadas. Algumas subfunções também foram abordadas. Os dados parecem confirmar que existe, de fato, um processo de gramaticalização em andamento no PB. As frequências de TIPO operando como item discursivo ou funcional estão muito acima da frequência deste lexema funcionando como um item.	The lexeme TIPO (type) has been described in literature as discourse marker (BITTENCOURT, 1999; DAMASIO-LOPES, 2008; CASTELANO; LADEIRA, 2010). However, not all tokens found in a corpus either operates as a discursive marker or as a full lexical item. Some occurrences seem rather to have a grammatical meaning. This paper aims thus at presenting evidence that the lexeme has been undergoing not only a discursivization but also a grammaticalization process (HEINE; KUTEVA, 2007). To do so, a speech corpus representative of Brazilian Portuguese, especially the diatopy of Minas Gerais (C-ORAL-BRASIL, RASO; MELLO, 2012), was surveyed and all occurrences of the lexeme TIPO were extracted. The data were then classified into four main categories, namely, full lexical item, functional item, tonal unit containing only the lexeme TIPO, and tonal unit containing only the expression “tipo assim” (like that, sort of), in accordance with the features each token presented. The best prototypes of each category were presented, and the data was tabulated, and frequencies were calculated. Some subfunctions were also addressed. The results seem to confirm that there is, indeed, a grammaticalization process ongoing in BP. The frequencies of TIPO operating as discursive or functional item is far above the frequency of this lexeme functioning as a full lexical item.

PALAVRAS-CHAVE	KEY-WORDS
Linguística de corpus; Gramaticalização; Lexema TIPO	Corpus Linguistics; Grammaticalization; Lexeme TIPO

INTRODUCTION

The occurrences of the lexeme TIPO (type) in its non-canonical use have been described as discourse markers in the literature, thus undergoing a process of discursivization. However, some occurrences of this lexeme, as observable in spontaneous speech corpora, can be accounted for neither as a full lexical item nor as a discourse marker. The present paper aims, thus, at providing evidence that the lexeme TIPO is undergoing a grammaticalization process, taking the role of a functional item. Furthermore, additional evidence for the discursivization process of the terms “tipo” (type) and “tipo assim” (sort of/like that) are presented.

To accomplish this goal, a spontaneous speech corpus representative of Brazilian Portuguese, especially the diatopy of Minas Gerais, was sampled. All the occurrences of the lexeme were extracted and then analyzed, as well as classified into four main categories, namely, full lexical item, functional item, tonal unit containing only the lexeme TIPO, and tonal unit containing only the expression “tipo assim”. This paper focuses on the occurrences operating as a grammatical item rather than as a lexical item or a discourse marker. Therefore, the morphosyntactic changes observed as a result of the grammaticalization process are presented. Quantitative data are also provided so that the frequencies of the major categories can be compared.

This paper is divided into three sections besides this introduction and the final remarks. The first section reviews the grammaticalization process according to Heine & Kuteva (2007), as well as sums up other works that have studied the lexeme TIPO. The second section describes the corpus and the sampling method, as well as outlines the methodology. In the third section we analyze the occurrences of each of the four categories and quantify the results. Main findings and directions for future research are summarized in the final remarks.

1 LITERATURE REVIEW

In this section we briefly review the stages of a process of grammaticalization (Subsection 1.1), since the focus of this work are the occurrences operating as a functional item, and other works concerning the lexeme TIPO (Subsection 1.2).

1.1 GRAMMATICAL EVOLUTION ACCORDING TO HEINE & KUTEVA

Grammaticalization has been described by Heine & Kuteva (2007) as a

unidirectional process in which four intermingled parameters take part. The first parameter has a pragmatic nature: extension. Extension gives rise to new grammatical meanings when the use of a linguistic term is extended to new contexts, resulting in context-induced reinterpretation. The second parameter is semantic in nature, namely desemanticization or semantic bleaching, which involves a process of semantic loss or generalization of meaning. The third parameter is decategorialization, which takes place when an item undergoes changing in the morphosyntactic features that characterizes it as a lexical or a less grammaticalized item. The last parameter is a phonetical one: erosion. The more often an item is used, the more it tends to be phonetically eroded, losing, thus, substance. In the subsections to come, we briefly summarize each parameter of a process of grammaticalization.

1.1.1 Extension

Extension is the most complex parameter. It is composed of three different components, namely the sociolinguistic, the text-pragmatic, and the semantic one (HEINE; KUTEVA, 2007). The sociolinguistic component is responsible for triggering a grammaticalization process. Generally, a change is started by individuals heedlessly enacting innovative linguistic uses of items or constructions that may eventually be adopted by other users of a speech community. The innovative use is, then, diffused in what Croft (2000) calls “propagation”. The pragmatic component entails, in its turn, the broadening of a context of use to new sets of contexts. Finally, the semantic component implements the change of meaning. The lexeme acquires the emerging grammatical meaning, and the new meaning may eventually become prevalent to the extent that the item loses its original lexical meanings. Extension is a four-staged process (HEINE; KUTEVA, 2007). This process is briefly sketched in Table 1, below:

Table 1 – A model of extension

Stage	Context	Resulting meaning	Type of inference
1. Initial stage	Unconstrained	Source meaning	
2. Bridging context	There is a new context triggering a new meaning	Target meaning foregrounded	
3. Switch context	There is a new context which is incompatible	Source meaning	Invited (cancellable)

	with the source meaning	backgrounded	
4. Conventionalization	The target meaning no longer needs to be supported by the context that gave rise to it; it may be used in new contexts	Target meaning only	Usual (typically non-cancellable)

Source: adapted from Heine & Kuteva (2007, p. 37)

According to Heine & Kuteva (2007, p. 37), this process takes place gradually as an item moves from a less grammatical to a more grammatical meaning. An item undergoing extension first finds a bridging context. The bridging context is often achieved by means of metaphors. However, the extension of use may eventually find a context where the original meaning cannot be interpreted as a metaphor. The item has reached the switch context. Finally, the new meaning is conventionalized, and it no longer needs the supporting original context.

It is noteworthy that, during a process of grammaticalization, the changing item may be concomitantly found on all the stages that it has already reached. The frequency of tokens on each stage may be regarded as an indicator of the degree of progress or completion of the grammaticalization process.

1.1.2 Desemanticization

Desemanticization takes place as soon as extension starts. This is so since an “expression E in a new context C entails that E loses part of its meaning that is incompatible with C” (HEINE; KUTEVA, 2007, p. 39). Metaphoric use, whereby a concept from a more concrete domain is transferred to a more abstract one, is thought to be one of the main processes leading to desemanticization (LAKOFF, 1987). For instance, names of body parts have undergone a desemanticization process in many languages, becoming locative adpositions, such as “behind” and “on top of” (HEINE; KUTEVA, 2007).

1.1.3 Decategorialization

The next parameter is decategorialization. According to Heine & Kuteva (2007, p. 40), some of the most salient features of a lexeme are lost once the linguistic item undergoing grammaticalization is desemanticized. For instance, a full lexical item may lose a) ability to be inflected; b) ability to receive derivational affixes; c) ability to take on modifiers; d) independence as an autonomous form (cliticization); e) syntactic freedom,

e.g., the ability to be positioned in the sentence in the same way as the non-grammaticalized source item; f) ability to be anaphorically indexed; g) belonging to a grammatical paradigm. Since our hypothesis is that the lexeme TIPO is becoming a grammatical item, analyzing its ability to combine with adjectives and determiners, as well as of being headed by adpositions, is the main purpose of this work.

1.1.4 Erosion

According to Heine & Kuteva (2007, p. 42), erosion is the last parameter to come up in a grammaticalization process. The phonetic substance of an item tends to be lost as a result of the grammaticalization. Once it is extended to new contexts, the item becomes more frequently used and is, as a consequence, phonetically eroded. According to Heine & Kuteva (2007, p. 43), phonetic erosion may affect not only segments, including full syllables, but also suprasegmental features, such as stress, tone, or intonation. Phonetic autonomy may be lost resulting in adjunction to neighboring phonetic units and phonetic simplification. Although analyzing phonetic features of grammatical tokens of TIPO is out of our scope, it is naturally a next step of this work. Since the lexeme can still be found in all contexts of the extension process – that is, as a full lexical item and as a grammatical item – phonetically analyzing occurrences of this item could shed some light on how the process takes places.

1.1.5 The cline of grammaticalization

In the process of grammaticalization, a full lexical item (content word) may turn into a grammatical word (function word). As abovementioned, this process occurs by means of gradual rather than by radical shifts. This overlapping chain of events forms, thus, a cline, down which the class of the item may be “downgraded”. These shifts generally follow similar paths across different languages. Although many clines have been put forward, Hopper and Traugott’s (2003) is the most accepted one. Hopper and Traugott’s (2003) cline is depicted below:

content word → grammatical word → clitic → inflectional affix

A content word has syntactic freedom, prosodic autonomy, and lexical meaning. A grammatical word, in its turn, has syntactic freedom, prosodic autonomy but grammatical meaning. Eventually, a grammatical word may lose its prosodic autonomy and transform into a clitic. Clitics only occur attached to prosodically autonomous words. Their syntactic freedom is, however, preserved to some degree. On the last stage, even the position may be conventionalized and fixed. The clitic loses its relative syntactic freedom

and become an inflectional affix.

We assume that the lexeme TIPO that cannot be accounted for in terms of discourse markers is changing from a content item to a grammatical word. However, there is still no significant loss of phonetic substance. Our hypothesis is that, so far, only extension, desemanticization, and decategorialization obtain.

1.2 PREVIOUS WORKS ON THE LEXEME TIPO

In the remainder of this section, we summarize three previous works that dealt with the grammaticalization/discursivization of the lexeme TIPO.

1.2.1 Bittencourt (1999)

Bittencourt (1999) analyzed the noun “tipo” in Brazilian Portuguese, noting that the lexeme frequently combines with the adverb “assim” (like this). According to the author, who studied an unsystematic corpus of spontaneous conversations, mostly of adolescents and some adults, “tipo” is undergoing both a process of grammaticalization and discursivization. These processes are particularly active among younger speakers.

On the grammatical level, Bittencourt (1999) affirms that the item is assuming the role of a syntactic complementizer. On the discourse level, Bittencourt (1999) affirms that “assim” has acquired functions similar to those of *verbi dicendi*. This category of verbs is responsible for creating, articulating, and modalizing new elocution situations, thus granting dynamicity and coherence to discourse. According to Bittencourt (1999, p. 49), some of the discursive functions found in the corpus were regulating turn management (sustaining, feedback, failover, overlapping of speech), manipulating the speaker's own utterances (hesitation, repetition, self-correction), and negotiating discourse (connivance, discordance, hetero-correction by of the interlocutor).

The author also points out that, in many cases, “tipo” is coalescing with the adverb “assim”. This process has given rise to phonetic repercussions. According to the author, the tonicity of “tipo” has weakened, and the item is constituting a unique group of force with “assim”, thus transforming into a unique phonological word.

1.2.2 Damasio-Lopes (2008)

Damasio-Lopes (2008), in her turn, only examined occurrences of the item “assim”. However, the author also noted the coalescence with “tipo”. According to the author, the expression “tipo assim” may play the role of an attenuation marker. A statement

evaluated as imprecise or doubtful may, thus, be accompanied by the expression to show the speaker's disengagement with respect to the content of the utterance.

1.2.3 Castelano & Ladeira (2010)

Finally, Castelano & Ladeira (2010) studied occurrences of the expressions “*assim*”, “*tipo*”, and “*tipo assim*”, as well as their respective discursive-interactional functions in oral narratives. The authors point out that the most frequently used functions are marking a narrative sequencing, inserting explanatory sequences, and hesitations. On the other hand, footing and inserting reported speech are the least used functions. Grammatical functions were not reported since the scope was rather communicative functions, i.e., the discursivization of the item.

As a matter of fact, granting special attention to the discursive functions of the lexeme TIPO, none of the authors draws further conclusions about the grammatical meaning the item has been acquiring – when it is not used as a discourse marker but rather as a connector – and how its morphosyntactic features have been changing. Presenting grammatical meanings of TIPO and evidence of its use as a grammatical word is, thus, the main purpose of this work.

2 CORPUS AND METHODOLOGY

In this section we outline the corpus examined (Subsection 2.1) and the features analyzed (Subsection 2.2).

2.2 CORPUS

The analysis of the lexeme TIPO was based on a sample extracted from the C-ORAL-BRASIL-I language resource (RASO; MELLO, 2012), which is a corpus representative of the spontaneous speech of Brazilian Portuguese, mainly of the diatopy of Minas Gerais, with wide diaphasic variation. C-ORAL-BRASIL is affiliated to the C-ORAL-ROM project (CRESTI; MONEGLIA, 2005), whose scope was the compilation of comparable corpora of the languages of the Romance genus, namely Italian, French, Spanish, and European Portuguese, having been further expanded to American English (CAVALCANTE; RAMOS, 2016). C-ORAL-BRASIL was chosen for this survey since it yielded from natural settings many innovative uses of the language, such as the one here

examined.

2.3 METHODOLOGY

All the occurrences found for the lexeme TIPO in the corpus were extracted by means of the concordancer AntConc (ANTHONY, 2018) and thereafter classified through the audit of tokens, using the software Winpitch (MARTIN, 2005), according to the following categories:

- I. full lexical item;
- II. functional item;
- III. tonal unit containing only the lexeme TIPO;
- IV. and tonal unit containing only the expression “tipo assim”.

Since the scope of this work is the use of TIPO as a grammatical word, only categories I and II were analyzed in depth. Although taking III and IV into consideration would result in a much more interesting and ample analysis, this step is set aside for future work. This is so because TIPO and “tipo assim” within dedicated tonal units may play a much more complex role on the discourse level. Analyzing discourse markers may require not only morphosyntactic but also prosodic-based criteria since different intonational patterns can convey different pragmatic functions (CRESTI, 2000; RASO, 2014; RASO & VIEIRA, 2014). All the same, the tokens of all four categories were tabulated and quantified for comparison purposes.

We examined the presence or absence of seven features. They were:

- I. being alone within a tonal unit¹;
- II. being collocated with the lexeme ASSIM within a tonal unit;
- III. being collocated with a maximum of one functional item, such as “mas” (but) or “e” (and);
- IV. exhibiting inflections;
- V. being collocated with determiners;
- VI. being collocated with adjectives;
- VII. typical syntactic position of conjunctions or preposition with respect to a Noun Phrase.

¹ Here defined as the speech flow uttered between two prosodic boundaries perceptibly marked by trained transcribers.

The tokens were then classified as a) full lexical item, exhibiting mainly the features IV, V, and VI; b) functional item, displaying only the feature VII; c) tonal unit containing only the lexeme TIPO, exhibiting at least features I and/or III; and d) tonal unit containing only the expression TIPO ASSIM, exhibiting at least features II and/or III.

In accordance with the transcription conventions of the C-ORAL corpora (CRESTI; MONEGLIA, 2005; RASO; MELLO, 2012), there are two types of perceptibly marked prosodic boundaries: non-terminal and terminal. The former parses internally an utterance and has a continuative prosodic profile. It is marked through a single slash “/” in the transcription. The latter divides the speech flow into utterances, thus displaying a terminal prosodic profile, which is marked through double slash “//” (MONEGLIA; RASO, 2014). Thus, for determining whether tokens of TIPO were in dedicated tonal units, we queried the patterns “ tipo /” for the initial position of the utterance, “/ tipo /” for the medial position, and “/ tipo //” for the final position. The same procedure was applied for “tipo assim”.

In the next section, we present the results of this work.

3 ANALYSIS AND RESULTS

In the following subsections each category is instantiated by its most prototypical tokens and discussed, being presented in descending order of frequency. We assume here that there are more and less prototypical tokens, even though for this work there was a great deal of convergence on the features exhibited by each category. A total of 305 tokens was found within the corpus. Translations and the code of multimedia files are provided below each example.

3.1 TONAL UNIT CONTAINING ONLY THE LEXEME TIPO:

The tokens of this category exhibited four main discursive functions, namely emphasizing, exemplifying, modalizing, and introducing explanations, thus playing an important role as a dialogic unit governing discourse. Below, we exemplify the most important functions observed within the sample:

(a) Emphasizing

ela é muito fofa / mas / **tipo** / a boca dela é horrorosa / assim //

she's so cute / but / like / her mouth is hideous / let's say so //

(bfamd113)

(b) Modalizing

falta / **tipo** / uma hora e quinze / mais ou menos //

it will take / **something like** / one hour and fifteen / more or less //

(bfamcv32)

(c) Exemplifying

tipo / se é um / um gato //

for instance / if it's a / a cat //

(bfamcv04)

(d) Introducing an explanation

o Paulo tá com maior dúvidas existenciais / **tipo** / na última aula eu fiquei quinze minutos aqui //

Paulo is having an existential crisis / **I mean** / last class I stayed for fifteen minutes here //

(bpubdl08)

For determining the exact function each token has, a much more complex and fine-grained prosodic acoustic analysis, which is out of our scope, should be carried out. However, based on the audit of the tokens, it is possible to affirm that most tokens seem to take on the exemplifying function, as in “tipo / se é um / um gato”.

Since the tokens of this category occurred within dedicated tonal unit, contrasting them to the rest of the utterance, they were not classified along with functional items, which are, on the other hand, part of the syntactic arrangement of the utterance. Within a dedicated tonal unit, TIPO operates mainly as a discourse marker or as parenthetical insertion (SANTOS & BOSSAGLIA, 2018). For further information on prosodic accounts of discourse markers and extra-clausal constituents we refer the reader to Raso (2014), Raso & Vieira (2016), Gobbo (2019), and Santos & Bossaglia (2018).

Since prosody is not the focus of this work, phonetic properties of such tokens are not dealt with. On the other hand, the analysis of features showed that none of them presented features such as IV, V, and VI. In conjunction with the prosodic isolation, the observed functions suggest that TIPO has been undergoing a discursivization process, as already pointed out by other authors.

3.2 FUNCTIONAL ITEM

The tokens of this category typically exhibited two functions, i.e., exemplifying and comparing. The analysis showed that seldom a lexeme exhibiting functional meaning displayed any of the morphosyntactic features of a full lexical item, namely inflections, the presence of determiners, collocation with prepositions, and collocation with adjectives. Furthermore, they tended to take the position commonly attributed to prepositions and connectors, although never connecting two finite clauses. The following examples instantiates the tokens classified as functional items:

(e) Comparing

ah / não / mas não era **tipo** essa sua não / é **tipo** aquela minha
oh / no / but it was not **like** yours / it's **like** that of mine
(bfamcv28)

(f) Comparing

ai faz um negócio **tipo** um leque / assim / tem um negócio dourado //
so, you do a thing **like** a fan / like that / there's a golden thing //
(bfamcv28)

(g) Exemplifying

tipo tempero / por exemplo //
such as seasoning / for instance //
(bfammn32)

Although a fine-grained phonetic survey was out of our scope, the audit revealed a few tokens exhibiting coalescence between TIPO and the following determiners or demonstrative pronouns. Example (h), below, might be regarded as evidence that there may be some degree of erosion, although an in-depth phonetic study should be carried out among speakers of other regions to confirm that this phenomenon cannot be accounted for by the typical slope of the Mineiro dialect towards shortening final vowels.

(h) **tip'a** outra pessoa

like the other person
(bfamcv23 – adapted)

Another interesting evidence of the ongoing grammaticalization process is the one displayed by example (i), below. The syntax of a full lexical item would block such a construction since its grammatical form would be “do tipo dessa” (of the type of this,

such as this). However, this token displays a syntax that may be interpreted as an intermediary stage between a full lexical item and a functional item. The tokens exhibiting functional meaning showed almost without exception a syntax where no preposition was present either before or after the lexeme. This was the only token presenting such a characteristic.

- (i) aqui / o' // eu faço uma pia **do tipo** essa aqui / o' //
look here // I can do a sink **of the type** this here / look //
(bfamdl20)

An ongoing decategorialization process implies morphosyntactic losses, as observed by Heine & Kuteva (2007). The tokens of this category were marked by having no inflections, no determiners, no collocation with adjectives, and typical syntactic position of conjunctions or preposition with respect to a Noun Phrase. We interpret these features as evidence that the lexeme is undergoing decategorialization although it can also be used as a full lexical item.

The item is also clearly undergoing a desemanticization process. The full lexical item was diffused to Latin from the Ancient Greek lexeme ΤΥΠΟΣ and, thereafter, transmitted to Portuguese. Its most prototypical meanings are:

- a) Set of traits or characteristics common to a class of beings or things. = ESPÉCIE, SORTE² (SPECIES, SORT);
- b) Concept that expresses the essence common to a set of things or beings, and serves as a paradigm. = MODELO³ (MODEL);
- c) Thing or being that has the characteristics that are considered exemplars of its class. = SÍMBOLO⁴ (SYMBOL);

The meaning of this lexeme very adequately suits the functions of identifying and comparing the characteristics of other full lexical items. Since its canonical use allows identifying a hyperonym of a class of entities or events, it can be used to compare items,

² "tipo", in Dicionário Priberam da Língua Portuguesa [on line], 2008-2013, <https://priberam.pt/dlpo/tipo> [consultado em 05-07-2018].

³ "tipo", in Dicionário Priberam da Língua Portuguesa [on line], 2008-2013, <https://priberam.pt/dlpo/tipo> [consultado em 05-07-2018].

⁴ "tipo", in Dicionário Priberam da Língua Portuguesa [on line], 2008-2013, <https://priberam.pt/dlpo/tipo> [consultado em 05-07-2018].

as well as to identify an item as a hyponym of the class it identifies. Furthermore, since TIPO has a very abstract meaning – exhibiting almost a set of features characteristic to a functional item –, it follows that almost no meaning has been lost on the desemanticization parameter.

What seems to happen here is rather a generalization of use that leads to higher frequency of occurrence, to consequent morphosyntactic losses, and, to some extent, to phonetic erosion. Therefore, we assume that the functional item is being syntactically used as a preposition or as a conjunction, the meaning of which is the same as in the complex preposition “do tipo de” (of the kind of) and the conjunction “como” (like).

Discussing if TIPO as functional item is a preposition or a conjunction is out of the scope of this work. However, it is noteworthy that no token of this category was used to connect sentences.

3.3 FULL LEXICAL ITEM

In contrast with the functional use, the lexeme TIPO used as a full lexical item may display all the features characteristic to a normal Noun Phrase. These features are inflections, collocation with determiners, and collocation with adjectives and prepositions. Examples (j), (k), and (l), below, illustrates the canonical use of TIPO:

- (j) preocupação por mostrar // não // ter / diversos **tipos** diferentes de fazer uma pintura //
worried about showing // no // to have / several different **manners** of painting //
(bfamd109)
- (k) mas é o **tipo** da coisa que e' &po
but this is the **sort** of thing that he &can
(bfamcv30)
- (l) porque / cada um / tem um / uma afinidade com um **tipo** de coisa / gosta / de um **tipo** de comida /
because / each one / has a / an affinity for a **type** of thing / is fond / of a **type** of food /
(bfammn08)

3.4 TONAL UNIT CONTAINING ONLY THE EXPRESSION TIPO ASSIM:

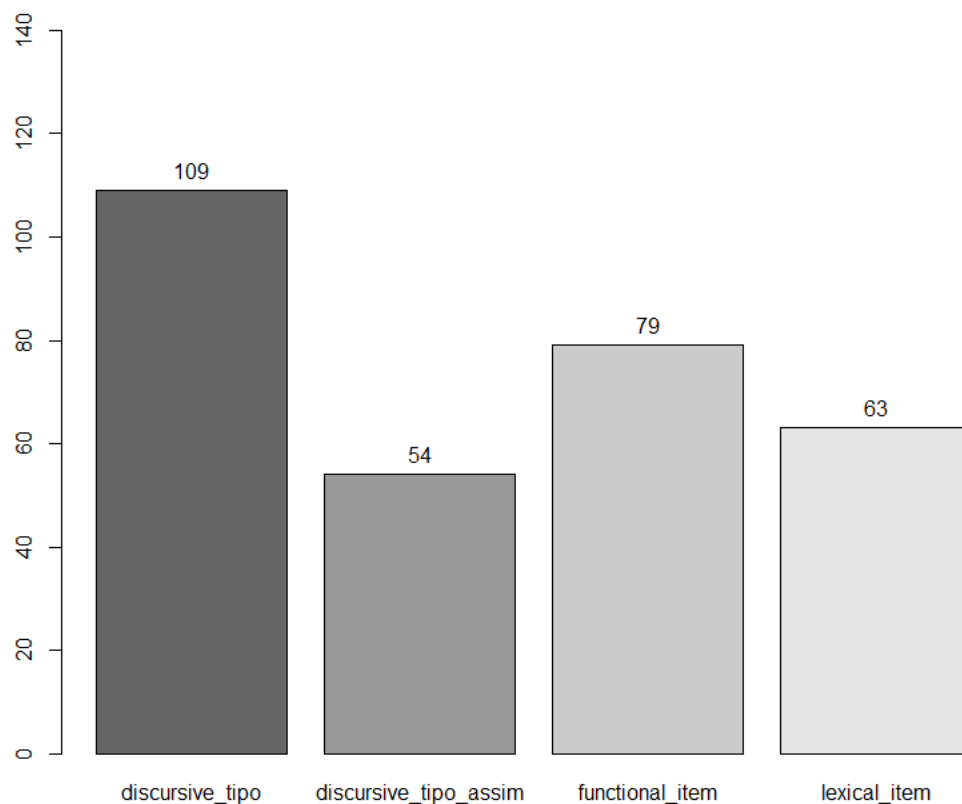
The tokens of this category exhibited functions very similar to the ones described in literature. Its functions are also very similar to TIPO within a dedicated tonal unit (Subsection 3.1). The analysis carried out in Subsection 3.1 also applies to “tipo assim”. Prosodic properties of this expression should also be better accounted for in order to check if prosodic acoustic characteristics may correlate with specific functions. Example (m) illustrates this category:

- (m) nem só com mulher // **tipo assim** / ocê tá puto pra caralho /
not only with women // I mean / you’re fucking angry /
(bfamcv23)

3.5 FREQUENCIES

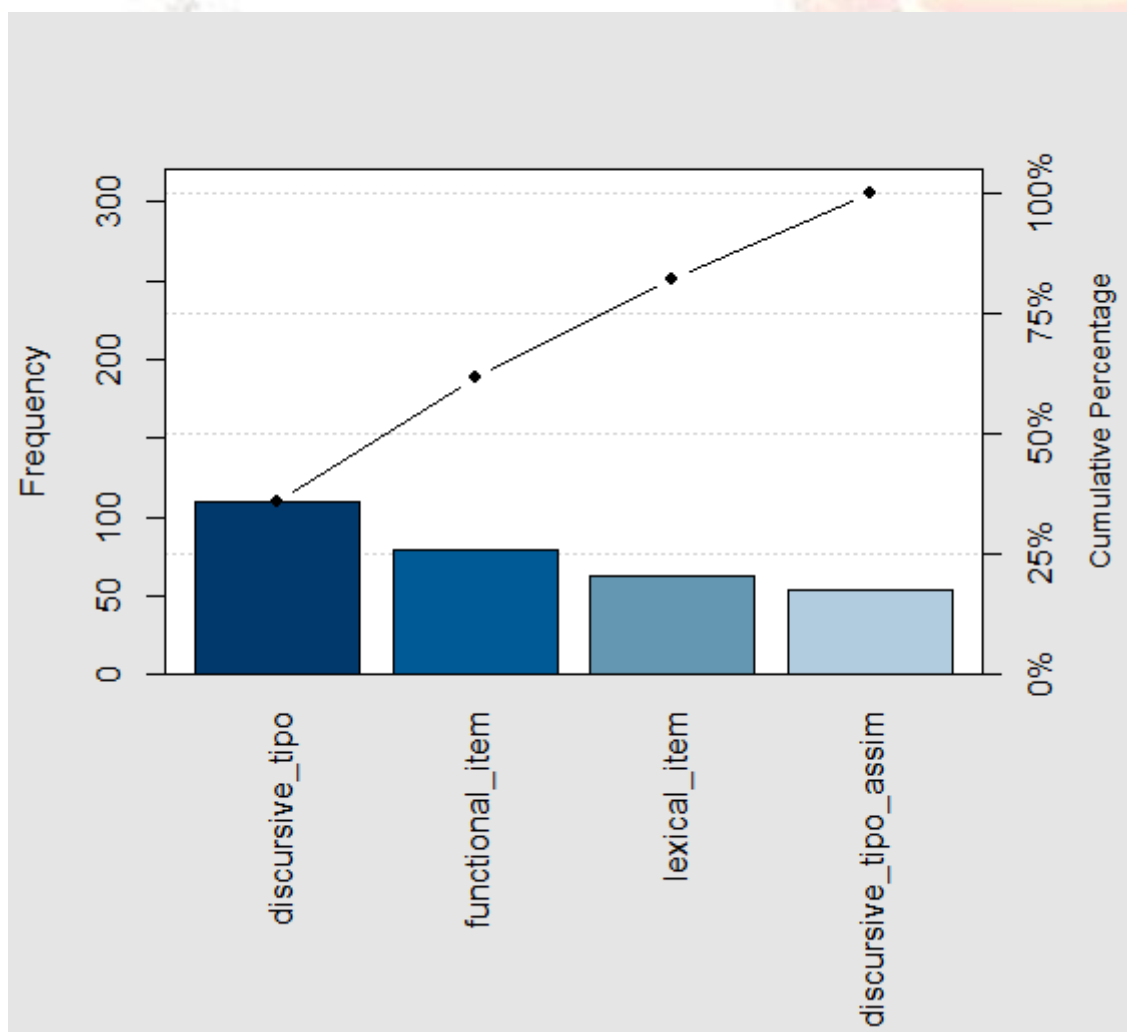
As aforementioned, a total of 305 occurrences of the lexeme TIPO were found on the corpus. The query included inflected forms, such as “tipos” (types) or “tipinho” (little type), although the latter was not found in the corpus. No token was discarded. According to Raso & Mello (2012), C-ORAL-BRASIL-I has a total of 208.130 words. Thus, the ratio of occurrences to the total of words was approximately 1 occurrence for every 682 words. Graph 1, below, displays the frequencies of the lexeme by category:

Graph 1 – Frequencies by category



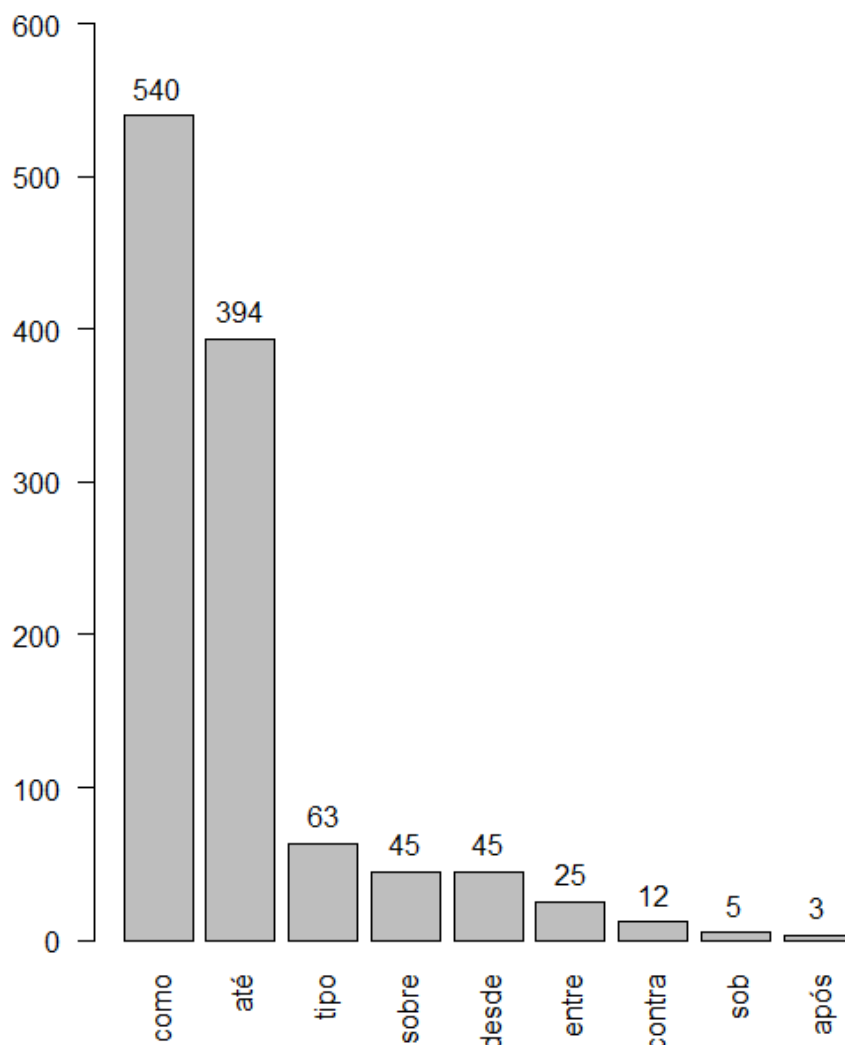
As seen in Graph 1, TIPO bearing discursive functions was the most frequent type. It is followed by the functional category. The canonical use of TIPO records 63 occurrences. In the last position, “tipo assim” records 54 occurrences. The non-canonical use largely outnumbers the canonical use. Graph 2, below, displays a Pareto chart of frequencies:

Graph 2 – Pareto chart of frequencies by type



Since they are used to articulate the language, one of the most prominent characteristics of functional items is that functional types tend to occur much more frequently than lexical types in corpora. Together, occurrences of TIPO operating as a discursive or functional item, coalesced with ASSIM or not, reach 242 occurrences, i.e. approximately 79 percent of all occurrences. The lexeme is, thus, four times more used as a functional or discursive item than as a full lexical item. For the sake of comparison, we summarize in Graph 3, below, the frequencies of some less frequent prepositions and the conjunction “como” (like) occurring within the corpus:

Graph 3 – Frequencies of chosen prepositions and conjunction “como” (like)



As it can be observed, TIPO as a functional item (63 occurrences) is more frequent than many other prepositions⁵. It also represents 11,6% of the occurrences of “como” (like), a conjunction with similar functional meaning, as aforementioned.

Although it is unlikely that TIPO completely loses its original lexical meaning, the results suggest that the lexeme is undergoing processes of grammaticalization and discursivization in Brazilian Portuguese. When the lexeme operates as a functional item, it has the same meaning of the complex preposition “do tipo de” (of the kind of) and “como” (like). Discussing the class of the emerging functional item is out of the scope of this work.

⁵ Como (like); até (until); tipo (like); sobre (about/over); desde (since); entre (between); contra (against); sob (under); após (after).

FINAL REMARKS

In this paper we have provided additional evidence that there are processes of grammaticalization and discursivization of the lexeme TIPO. In the spoken language this process seems to be fairly advanced, considering that the lexeme occurs much more frequently as a functional or discursive item than as a full lexical item in a speech corpus representative of Brazilian Portuguese. TIPO seems to be undergoing a generalization of use, whereby it is acquiring a new meaning of “do tipo do” (of the kind of) or “como” (like). It seems also to be going through a decategorialization process since no token operating as a functional item has displayed the features characteristic to the non-grammaticalized source item, i.e., inflections, the presence of determiners, and collocation with adjectives. With far less evidence it seems possible to say that the functional item has also undergone some degree of phonetic erosion, although a more phonetically accurate research is needed to confirm this hypothesis. Some of the subfunctions of the functional and discursive items were set out, e.g., exemplifying, comparing, modalizing, and introducing explanations.

All the same, there is still a great deal of work to be done. The functions of the tokens occurring in dedicated tonal units are to be better understood. They may have discursive but probably also functional meanings. Different prosodic realizations of the lexeme TIPO in dedicated tonal units may correlate with specific functions. A phonetic study in an informationally tagged speech corpus could shed light on the prosody of such items. Furthermore, the class of the emerging functional item should be analyzed with respect to the occurrences found in the corpus. Finally, the study of speech corpora of the formal register and written language corpora could provide additional evidence of the extent to which this process of grammaticalization has reached.

REFERÊNCIAS

ANTHONY, L. (2018). **AntConc** (Version 3.7.5), Computer Software, 2018. Available from: <http://www.laurenceanthony.net/software/antconc/>. Accessed on: 05/05/2018.

BITTENCOURT, V. O. Gramaticalização e Discursivização no Português Oral do Brasil: o caso “Tipo (Assim)”. **SCRIPTA**, Belo Horizonte, v. 2, n. 4, 1º sem., 2019, p. 39-53.

CAVALCANTE, F.; RAMOS, A. The American English spontaneous speech minicorpus.

Architecture and comparability. **CHIMERA: Romance Corpora and Linguistic Studies**. 3.2, p. 99-124, 2016.

CASTELANO, K. L.; LADEIRA, W. T. Funções discursivo-interacionais das expressões “assim”, “tipo” e “tipo assim” em narrativas orais. **Letra Magna**, ano 6, n. 12, 1º sem., 2010.

CRESTI, E.; MONEGLIA, M. **C-ORAL-ROM**: integrated reference corpora for spoken Romance languages. New York, Amsterdam: John Benjamins, 2005.

CROF, W. 2000. **Explaining language change**: an evolutionary approach. London: Longman, 2000.

DAMASIO-LOPES, L. R. (2008). **A emergência do marcador discursivo “assim” sob a óptica da gramaticalização**: um caso de multifuncionalidade e (inter)subjetivização, 2008, 247 fls. (Dissertação de mestrado) Universidade Estadual Paulista, São José do Rio Preto: 2008.

GOBBO, O. R. **Marcadores discursivos em uma perspectiva informacional**: análise prosódica e estatística, 2019, 158 fls., (Dissertação de mestrado) Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.

HEINE, B.; KUTEVA, T. **The genesis of grammar**: A Reconstruction. New York: Oxford University Press, 2007.

HOPPER, P. J.; TRAUGOTT, E. C. **Grammaticalization**. 2nd ed. Cambridge: Cambridge University Press, 2003.

MARTIN, P. WinPitch Corpus: a text to speech analysis and alignment tool. In: Cresti, E. and Moneglia, M. **C-ORAL-ROM**: integrated reference corpora for spoken Romance languages. New York, Amsterdam: John Benjamins, 2005.

MONEGLIA, M.; RASO, T. Notes on Language into Act Theory (L-Act). In: RASO, T; MELLO, H. (Eds.). **Spoken Corpora and Linguistic Studies**. Amsterdam/Philadelphia, John Benjamins Publishing Company, 2014, p. 468-494.

RASO, T. Prosodic Constraints for Discourse Markers. In: RASO, T; MELLO, H. (Eds.). **Spoken Corpora and Linguistic Studies**. Amsterdam/Philadelphia, John Benjamins Publishing Company, 2014, p. 411-467.

RASO, T., MELLO, H. **C-ORAL-BRASIL**: corpus de referência do português brasileiro falado informal. I. Belo Horizonte: UFMG, 2012.

RASO, T.; VIERA, M. A. A description of Dialogic Units/Discourse Markers in spontaneous speech corpora based on phonetic parameters. **CHIMERA: Romance Corpora and Linguistic Studies**, Vol. 3, No 2, 2016.

SANTOS, S. BOSSAGLIA, G. Morphosyntactic, prosodic, functional and distributional description of the information unit of Parenthesis in spoken Brazilian Portuguese. **CALETROSCÓPIO**, v. 6, p. 37-63, 2018.

Título em português:

**PROCESO DE GRAMATICALIZAÇÃO EM CURSO DO LEXEMA
“TIPO”: ESTUDO BASEADO EM UM CORPUS DE FALA
ESPONTÂNEA DO PORTUGUÊS BRASILEIRO**