

## COMMAS IN TEXTS OF STUDENTS FROM AGE 11 TO 15-YEARS-OLD: A LONGITUDINAL STUDY

Luciani TENANI\*  
Tainan CARVALHO\*\*

- **ABSTRACT:** In this article, we investigate the presence and the absence of commas in a longitudinal sample of texts of the genre report written by 11 to 15 years-old students from a school in the countryside of São Paulo State. Our hypothesis is that the comma is used when syntactic boundaries match the prosodic ones. The prosodic constituents are formed from syntactic and phonological information, according to principles defined in the Prosodic Phonology theoretical approach. The characterization of the conventional commas and the non-conventional absences of commas was made from the identification of syntactic boundaries in which they should or could be placed, taking as reference a Brazilian Portuguese grammar. The results were that the conventional comma increases during the years of schooling, as expected. However, the presence and absence of comma rates changed depending on the type of syntactic structure and on the occurrence of intonational phrase boundaries. Comma occurred predominantly in positions where intonation phrase boundaries match major syntactic boundaries, as a sentence boundary. In turn, the comma systematically does not occur in positions where the syntactic and prosodic boundaries do not match.
- **KEYWORDS:** Commas; syntax; prosody; scholar texts; middle school.

### Introduction

In this paper, we address the comma from a linguistic perspective. Comma is one of the punctuation marks which helps organize and rank information of a written text and supports meaning construction in and between enunciations (DAHLET, 2006). According to Chacon (1998), this mark, along with the other punctuation marks, acts not only in the syntactic dimension of language – a dimension which frequently

---

\* Universidade Estadual Paulista (UNESP), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, São Paulo, Brasil. Departamento de Estudos Linguísticos e Literários. luciani.tenani@unesp.br. ORCID: 0000-0002-8487-0825.

\*\* Universidade Estadual Paulista (UNESP), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, São Paulo, Brasil. Programa de Pós-Graduação em Estudos Linguísticos. tainan.carvalho@unesp.br. ORCID: 0000-0001-6105-3120.

receives priority in the approach to comma in traditional grammars –, but also in other linguistic dimensions, such as phonological, textual and enunciative. In this paper, we shall discuss in detail the relations established by comma between the syntactic and prosodic dimensions based on Chacon (1998), yet – in contrast to this author – we shall adopt the model of Prosodic Phonology (NESPOR; VOGEL, 1986, 2007), which will be presented in the following section, to detect relevant prosodic structure boundaries in order to describe the functioning of the existing commas and, furthermore, set forth hypotheses concerning the absence of commas.

According to Dahlet (2006), commas can have two ways of functioning, one being called “simple schema”, when the comma delimits only the right border of a linguistic structure, and the other, “double schema”, when two commas are used, one at the right border and the other at the left border of a given structure. Carvalho (2019) has analyzed occurrences of commas based on both types of schemas. From that study, we have selected the cases of commas in simple schema, given their characteristics found in the analyzed sample, to be described in the section where we present the results. One example of investigated data is the presence of comma after the adverbial adjunct in: “*Na próxima semana, haverá aula de Português*” (**In the next week**, there will be Portuguese classes). In this example, after the adverbial adjunct “Na próxima semana”, there should be a comma, according to the grammatical rules of Brazilian Portuguese (BECHARA, 1999), since the adjunct is displaced to the first syntactic position in regard to the main clause. If this comma is not used at the syntactic boundary, we observe it as data indicating absence of comma. Therefore, two kinds of data shall be considered: when the punctuation mark is present, classified as conventional usage, and when the mark is absent, classified as non-conventional absence. These categories shall be detailed ahead.

The comma analysis will be performed on a longitudinal sample of texts written by students from 11 to 15 years-old, that primarily correspond to Middle School (henceforth, MS), at a public school in the interior of São Paulo, in the context of a university extension project. The selection of the longitudinal sample is supported by the Curricular Proposition of the State of São Paulo (2008), the document published at the time when the texts were collected, which establishes the teaching of punctuation in MS, and, more specifically, in the final years of the education cycle. According to this document, students are expected to finish MS being proficient in basic notions of punctuation mark usage (and, therefore, of comma). Based on the results, we have put into discussion to what extent this institutional expectation is fulfilled in the analyzed longitudinal sample.

In sum, three main questions are addressed in this paper, namely:

- 1) What are the characteristics of the use of commas in simple schema in productions written by students throughout MS?
- 2) Which syntactic structures support the presence of comma and which ones favor the absence of comma? and

- 3) to what extent could the presence as opposed to the absence of comma be related to the prosodic organization of the enunciations?

Answers to these questions shall be provided in this paper, which has been structured as follows: in the next section, we shall present the theoretical assumptions which support the study of comma as a complex linguistic object; in the section “Analysis material and methodological decisions”, we describe the longitudinal text sample, provide details about the text collection process and explain the quanti-qualitative procedures followed in order to identify, gather, classify and analyze commas, performed by Carvalho (2019); in the section “Syntactic-prosodic characteristics of commas”, we carry out the data analysis and discuss syntactic and prosodic characteristics detected in the presence and absence of commas in the texts of the sample. In the final section, we systematize the syntactic and prosodic complexity involved both in the presence and absence of comma in MS school texts and go a step further in regard to Carvalho (2019), by providing considerations about contributions of this research to its field of study and to future research on commas and on the relation between speech and writing, taking into account the prosodic configuration of enunciations.

### **Theoretical assumptions about commas, writing and speech**

Comma can be considered a graphic sign with a complex functioning, since it acts as a syntactic and semantic operator, as Dahlet (2006) argues. Additionally, it can play different syntactic functions, such as segmenting or ranking parts of the enunciation (DAHLET, 2006), or adding, subtracting and reversing information (THIMONIER, 1970). Dahlet (2006) divides the complexity of the comma’s syntactic functioning in two main characteristics: (i) it acts in a simple schema (when the comma is placed at the right border of the syntactic structure – highlighted in the example), such as in (1.1), or in a double schema (when commas are placed at the right and left border of the syntactic structure highlighted in the example), such as in (1.2); and (ii) it acts on the inter-clause (between clauses) and intra-clause (within a clause) domain, as illustrated in the examples in (2.1) and (2.2), respectively. These characteristics set the comma apart from other punctuation marks, given that this syntactic functioning cannot be verified for the other punctuation marks (DAHLET, 2006).

- (1.1) *No final da tarde de sexta, Mariana renovou o documento.*  
[**At the end of Friday afternoon**, Mariana renewed the document]
- (1.2) *Mariana, no final da tarde de sexta, renovou o documento.*  
[Mariana, **at the end of Friday afternoon**, renewed the document]
- (2.1) *Quando Mariana chegou, Pedro já havia ido embora.*  
[**When Mariana arrived**, Pedro had already left]

(2.2) *Mariana, o Pedro chegou.*  
[**Mariana**, Pedro has arrived]

Another characteristic which adds to the syntactic complexity of the comma is the history of its constitution as a punctuation mark, as discussed by Rocha (1997), in a general sense, and also shown by Yano (2018), particularly in the case of commas in the tradition of Portuguese. Analyses of comma usage over time reveal various functions underlying this punctuation mark as part of the punctuation system, see Rocha (1997) and Yano (2018). About the history of comma usage, we would like to highlight two tendencies defined by Dahlet (2006) and described by Soncin (2014), when analyzing school texts, namely the phonocentric and autonomist tendencies of punctuation. They present opposite points of view about the role(s) performed by punctuation marks (and about comma).

The phonocentric punctuation tendency, historically set between Classical Antiquity and the Middle Ages, consists in an answer to the need of creating graphic resources in order to record prosodic aspects of speech in writing (ROCHA, 1997). At that time, punctuation marks originate with the purpose of reflecting characteristics related to speech in a supposedly direct manner. Concerning comma, its primary function is to indicate pauses to breathe and delimit units based on intonation. We can infer that there is a direct relation between speech and writing, so that the comma is assumed to be able to accurately transpose to the written text some of the characteristics of speech.

The autonomist punctuation tendency, on the other hand, emerges around the 19<sup>th</sup> century, along with the emergence of the French press, which provides a greater dissemination of written texts to the society at that time, and, intrinsically, the practice of silent reading. In this context, the punctuation marks are defined as logical-grammatical mechanisms (as Yano (2018) has also described based on the analysis of Portuguese historical texts), supported by the distinction between speech and writing in their semiosis, as different linguistic modalities. As a result, the need to dissociate speech from writing and establishing rules (mainly of syntactic nature) for the use of punctuation marks is reinforced, aiming at the standardization of these uses desired by grammarians. The primary role of comma from that time on is to segment syntactic units which are previously determined by a set of rules, with the aim of making written texts more readable.

Through this brief review of the history of punctuation, one can observe that, alternately, the phonic and the syntactic dimensions of comma are in focus, neither dimension losing their importance for the functioning of this punctuation mark. This history of punctuation, which is part of the history of writing, helps us understand the complex nature of the comma and, likewise, the sometimes contradictory usage instructions which can be identified in the school environment, as pointed out by Soncin (2014).

One example of this contradiction is the treatment which is generally given to the comma within traditional normative grammars which set forth rules for its usage.

In one set of literature, the comma is described as a sign, by its very nature a pause marker (ROCHA LIMA, 1986; FARACO; MOURA, 1997; CUNHA; CINTRA, 2017). The relationship between comma and speech can be noticed in the words of Cunha and Cintra (2017, p. 657): comma is used “to reconstruct the active movement of oral illocution”. On the other hand, in Luft (1998) we find an attempt of dissociation from the characterization of comma as a pause marker, describing it mainly as a sign with a syntactic character, since the author prescribes the use of comma to separate coordinate terms, additive and alternative coordinate clauses, adversative conjunctions and adverbs, repetition of terms, among others. These comma usage rules display yet another variability, according to Soncin (2014): comma uses which are optional in one work might be mandatory in another and vice-versa.

This variability in usage rules and the complex historical constitution of comma have been taken into consideration to characterize commas from various theoretical-methodological perspectives, analyzing various materials (CHACON, 1998; CORRÊA, 2004; ESVAEL, 2005; DAHLET, 2006; PACHECO, 2006; ARAÚJO-CHIUCHI, 2012; SONCIN 2014; CARVALHO, 2019; TENANI; PAIVA, 2020). In the literature about comma usage in written texts in different languages, there are also various approaches (SIMARD, 1993; CATACH, 1994, for French; CHAFE, 1988, for English; STEINHAUER; FRIEDERICI, 2001; FUCHS; KRIVOKAPIĆ, 2016, for German; CASSANY, 1999; VERDEGUER, 2010; ANDRADE, 2018, for Spanish; PRIMUS, 2007, for various languages, among others).

This study is part of the set of studies about comma usage in texts produced by MS students at public schools in Brazil. We follow the same theoretical-methodological framework of studies about punctuation aiming to encompass the constitutive complexity of the comma as part of the punctuation system (CHACON, 1998; CORRÊA, 2004; ESVAEL, 2005; ARAÚJO-CHIUCHI, 2012; SONCIN, 2014). Along with these studies, we have chosen to detach ourselves from the way of understanding comma usage as the phonocentric and autonomist traditions conceive it, considering it exclusively a linguistic syntax or prosody fact. We align with the view of punctuation supported by Chacon (1998), for whom punctuation acts at least in four dimensions of language, namely: (i) the syntactic dimension; (ii) the phonic dimension; (iii) the textual dimension; and (iv) the enunciative dimension. All of these dimensions are intrinsic to punctuation and integrate it simultaneously, meaning that disregarding them results in obliteration of the constitutive complexity of punctuation.

This study gives priority to the analysis of commas in syntactic and phonic dimensions, without disregarding the relevance of other dimensions mentioned by Chacon (1998) for the description and analysis of comma uses. We understand that commas mark syntactic boundaries of written enunciations, acting as a mechanism of segmentation, organization and hierarchization of text portions. However, we assume its role not to be restricted to an exclusively syntactic function, given that comma also establishes symbolic relations (therefore, not in a direct manner, as set forth by the phonocentric tradition) with the prosodic organization of enunciations in a language.

Consequently, the relations between comma, syntax and prosody which we intend to establish here are not conceived the same way as in the phonocentric and autonomist approaches. From our point of view, comma is taken as a graphic sign which builds syntax and is also a “linguistic sign of symbolic processes which take place in writing by means of the relation with orality, in particular through the prosodic domain” (SONCIN; TENANI, 2015, p. 476).

This approach to commas is grounded in a concept of writing and its relation to speech expressed by Corrêa (2004). This author argues that writing has had a heterogeneous character all along since its genesis, and the punctuation marks, as part of writing, also feature this character in their constitution: they belong to the graphic domain and represent phonic features of language. Additionally, as Corrêa (2004) argues, speech and writing can be conceived as modes of enunciation which, from a discursive perspective, take place through oral and literate practices. This theoretical framework is the basis for the approach to the roles of comma so as to simultaneously gain an understanding of speech and writing characteristics. Therefore, the adopted theoretical perspective surpasses the perspectives of the phonocentric and autonomist traditions of punctuation, which are grounded on a dichotomic conception of the linguistic modalities of speech and writing, by following a view that writing is constituted by speech (see details of this approach and its relation with punctuation in CORRÊA, 1994).

Having shown the theoretical foundations which support the concept of comma within the scope of the complex relation between speech and writing, we shall outline the theoretical foundations providing the basis of methodological decisions in regard to the conception that speech has prosodic elements which are organized in a hierarchy of constituents. The model of Prosodic Phonology, set forth by Nespor and Vogel (1986, 2007), lays the foundation for our interpretation that prosodic boundaries, which are phonologically structured, are linked to the positions where commas are (or could be) used. The proposition we assumed is that the prosodic organization is part of the speaker/listener/writer’s grammar and thus the enunciations, whether spoken or written, are produced based on this phonological grammar. Within this framework, enunciations are organized in a hierarchy of seven prosodic constituents, where a higher domain is necessarily comprised in a lower one of the hierarchy (NESPOR; VOGEL, 1986). Studies which have followed this theoretical framework to address prosodic boundaries that are relevant to describe the functioning of commas in MS texts (ARAÚJO-CHIUCHI, 2012; SONCIN, 2014; CARVALHO, 2019) argue that the intonational phrase (or syntagm) domain (IP, Intonational Phrase) is the most relevant to describe uses of commas. This assertion is made based on the observation, in the above-mentioned studies, that the presence of commas occurs in positions where possible IP boundaries could be located, whereas the absence of commas tends to overlap with prosodic boundaries which potentially would not occur due to the restructuring possibilities of the IP domain, accounting for semantic and pragmatic conditions, as well as syntactic and phonological factors, which we will address ahead. In sum, prosody will not be considered here in its phonetic characteristics, but rather in its phonological

organization, following an approach which assumes the interface between syntax and phonology to be the structuring backbone of prosodic constituents.

In the framework of Prosodic Phonology, the formation algorithms of prosodic domains are defined, and the speech data analysis provides segmental, rhythmic and intonational cues to its configuration. In the case of IP, the intonation contour is a key characteristic, while its boundaries are potentially delimited by a pause (TENANI, 2002; SERRA, 2016). IP domain can be predicted based on the relation between syntactic constituents: the higher boundaries of the syntactic structure, such as those of clauses, overlap with IP boundaries, but these ones may also overlap with syntactic ones smaller than that of a clause, such as, for instance, the edges of vocatives, in [Marina],IP [Pedro has arrived]IP. We point out that the basic assumption of this theoretical framework is the possibility of not having isomorphism between the prosodic and syntactic boundaries, as in: [The girl in green shirt]IP [missed the class today]IP. In this example, the subject of the clause “the girl in green shirt” forms one IP and the predicate “missed class today”, another IP. The prosodic boundary occurs between the subject and the predicate.

In this paper, the identification of IP is based on the IP formation and restructuring algorithm adapted by Frota (2000)<sup>1</sup> for Portuguese, based on Nespor e Vogel (1986). Each IP is formed by phonological phrases (PPh), which correspond to noun, verb or adverbial phrase, for instance.<sup>2</sup>

### (3) Formation and restructuring algorithm of the Intonational Phrase (IP):

a) **I Domain:** (i) all the  $\phi$ s in a string that is not structurally attached to the sentence tree (i.e. parenthetical expression, tag questions, vocatives, etc); (ii) any remaining sequence of adjacent  $\phi$ s in a root sentence; (iii) the domain of an intonation contour, whose boundaries coincide with the positions in which grammar-related pauses may be introduced in an utterance.

b) **I Restructuring:** (i) restructuring of one basic I into shorter Is, or (ii) restructuring of basic Is into a larger I. Factors that play a role in I restructuring: length of the constituents, rate of speech, and style interact with syntactic and semantic restrictions.

(FROTA, 2000, p. 57).

The IP formation algorithm comprises information about factors of syntactic and phonological character, while the domain restructuring is subject to performance (rate

---

<sup>1</sup> Frota (2000) uses “I” for intonational phrase and “ $\phi$ ” for phonological phrase. In this paper, we follow the notation which is most often used in literature nowadays, namely: “IP” for intonational phrase and “PPh” for phonological phrase.

<sup>2</sup> For more information about the constitution of the phonological phrase in Brazilian Portuguese, we recommend Sândalo and Truckenbrodt (2003).

of speech and style), taking into account syntactic and semantic restrictions. Among these factors which contribute to IP restructuring, we must point out that intonational contours, pause, rate and style of speech are not retrievable in the written texts which are part of the investigation material, which we shall discuss in detail in the following section. We will consider information which can be analyzed based on the written material, such as length of the constituents (provided in amount of syllables), syntactic branching and semantic interpretation.

## **Analysis material and methodological decisions**

This study has been carried out using a set of 248 texts, written by 62 students over the course of the four years of MS (62 students x 4 texts [1 per school year] = 248 text productions) between 2008 and 2011.<sup>3</sup> The texts have been selected from the Textus Database (TENANI, 2015), which comprises over 5,000 texts, organized in transversal and longitudinal samples. Given the longitudinal nature of this research, the set of analyzed texts belongs to the longitudinal sample of the database.

This database was built with texts produced in the context of a university extension project by UNESP-IBILCE in partnership with a state school in the city of São José do Rio Preto.<sup>4</sup> In order to conduct the extension project which would lead to the creation of the database, financial support was obtained by the first author through the Pro-Rector of Extension and Culture (PROEC) of UNESP, as well as through FAPESP (Processes 13/25777-8 and 13/24767-9). The students of the partner school who participated in the project received the task, at each monthly meeting, over a time of fifty minutes, to write a text according to the topic and genre previously studied in class, collectively as a class, teacher or undergraduate student of Letters participating in the extension project.<sup>5</sup> The students did not receive extra time to rewrite the productions, and one of the instructions was that they should use a ballpoint pen for the text production. Thereby, the students could make erasures, but not wipe out what had been written. This procedure was typical for the partner school, according to Tenani (2016), and allows us to obtain a “photography” of the text production time.

Out of a large number of texts, characterized by different text genres and typologies following the Curricular Proposition of the State of São Paulo (2008), we selected a set of texts from the longitudinal sample which belongs to the same typology: report.

---

<sup>3</sup> The productions were written by students who would complete eight years of Basic Education. This basic level was changed to nine years from 2009 on, and since 2016, children have been required to attend two years of kindergarten. The effect of this expansion on text production (and possibly of comma) could become the object of future research.

<sup>4</sup> As Tenani (2016) argues, the texts of the database provide a sample of children’s writing of students from MS not only in the state of São Paulo, but also of Brazil, given that the partner school of the extension project achieved scores in Portuguese which were slightly higher than the average of the targets set for the state and the country, when considering the Development Index of the State of São Paulo and the Basic Education Development Index.

<sup>5</sup> In accordance with the ethical research procedures, the students were invited to participate in the project and had the option of refusing to write the texts, without being punished for this refusal.



However, the selected productions differ in topic and text genre.<sup>6</sup> We have only selected texts from one typology, due to the fact that, given the database configuration, report is the only typology consistently produced across all school years of MS and, therefore, this text feature could be controlled for the longitudinal study. Thus, we have limited this study to a single textual typology, although investigating potential relations between comma usage and text genre/type is a relevant topic for future research. In chart 1, we have systematically outlined the specificities of the material used in regard to topic and text genre for each school year, which can be freely accessed upon registration in the database website.

**Chart 1** – Topic and genre of the typology report

School year	Topic	Text genre
5 <sup>th</sup>	Internet and use of MSN	Personal letter
6 <sup>th</sup>	Fights between friends	Personal letter
7 <sup>th</sup>	Advantages of immortality	Report of life expectation
8 <sup>th</sup>	Friendship and Internet	Experience report

**Source:** Textus database (TENANI, 2015).

Concerning the analysis methodology, this study comprises a combination of a quantitative and qualitative approach, the quantitative approach being driven by the interest in systematizing the patterns of the presence and the absence of comma from a longitudinal perspective, and the qualitative being motivated by the interest in showing the relation between the systematicity of presence and absence of commas and the prosodic organization of enunciations, notably the IP boundary.

As stated before, we shall address only the simple schema of comma usage. Two possibilities of comma usage following this schema will be analyzed: the conventional presence and non-conventional absence, typically called, respectively, “hit” and “error” from a normative point of view.<sup>7</sup> The terminology of conventional versus non-conventional used reveals our proposition to stand apart from the perspective which assesses the written production in terms of hit and error, and instead establish a dialog with the grammatical convention, seeking to show which linguistic characteristics act so as to cause presence or absence of the comma, aiming thereby to capture the process

---

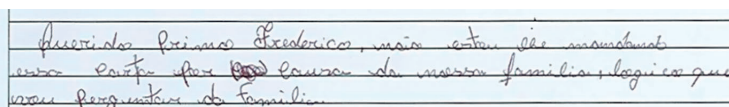
<sup>6</sup> We chose to keep the terms “genre” and “text typology” as used in the text classification of the Textus database. This classification follows the guidelines of the Curricular Proposition of the State of São Paulo (2008) which has a concept close to that of Textual Linguistics about genre and typology (MARCUSCHI, 2001). Nevertheless, from a discursive perspective (BAKHTIN, 1992), this notion of text genre and typology does not hold. We have pointed out these theoretical differences and do not go further into this discussion, since we will not analyze characteristics of text genre.

<sup>7</sup> Another possibility of using the comma in a simple schema is the non-conventional presence, likewise considered to be an “error” from the traditional perspective. The non-conventional presence is marked by insertion of comma in boundaries which are not expected by the grammatical tradition concerning comma usage, as the separation between subject and predicate, for instance. A discussion about this case, not featured in this paper, has been made by Carvalho (2019).

of learning the language (and writing conventions). For the definition of presence or absence of comma, we use the syntactic rules set forth in the work *Moderna Gramática Portuguesa* (‘Modern Portuguese Grammar’) by Bechara (1999). The choice of this grammar is due to the fact that it is a work which guides some of the written school practices and because Bechara (1999) engages, even if briefly, in discussing the relation between punctuation marks and speech characteristics, without linking the comma directly to the pause to breathe and not restricting himself to the syntactic criterion to prescribe the comma usage.

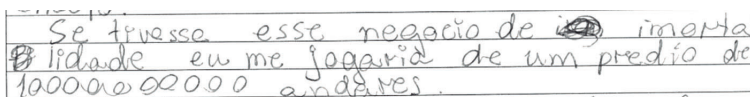
In figures 1 and 2, in the sequence, we provide examples of the considered types of data. Fig. 1 provides an example of the conventional presence of comma in simple schema, as the right border of the vocative (Querido primo Frederico [Dear cousin Frederico]) should be delimited by a comma, according to Bechara (1999). In Fig. 2, there is a non-conventional absence of the comma, since no comma was used at the boundary of the adverbial phrase (Se tivesse esse negócio de imortal [If there were such stuff as immortality]), which is displaced in comparison to the direct order subject–verb–complement (I would throw myself [...]), according to Bechara (1999).

**Figure 1** – Example of conventional presence of the comma



Source: Textus database. Text: Z08\_5C\_01M\_04.<sup>8</sup>

**Figure 2** – Example of non-conventional absence of the comma



Source: Textus database. Text: Z10\_7C\_14F\_01.

The data identified based on Bechara (1999) was organized systematically by: (i) type of data; (ii) type of syntactic boundary and (iii) school year in which the data was gathered. After identifying and classifying the data, quantitative analyses were performed with the aid of statistical tests. In the statistical analysis, the software Minitab 17, Excel 2010/2013, BioEstat 5.3 and the Action Portal were used. Based on the statistical treatment, we obtained the percentual frequency and the mean value of the types of comma usage per school year. We performed the Post-Hoc Bonferroni test in order to detect for which school years the differences in comma usage were more robust and the Chi-squared association test to assess the possibility

<sup>8</sup> Each text of the database is identified by a code with the following information: school, grade, class, number and gender of the student and number of the proposition. In the example “Z08\_5C\_01M\_04”, we read: Zulmira (Z) school, year 2008, fifth grade class C, student 1, male, proposition 4.

of association between the type of comma use for each syntactic boundary. The Chi-squared test was also followed by the Pearson Coefficient calculation with the aim of investigating how significant the association between comma usage and syntactic boundary could be. In all performed statistical tests, the value of 5% of significance level ( $p < 0.05$ ) was used.

Afterwards, we analyzed the data qualitatively by establishing a relation between the presence and absence of commas, previously identified through syntactic criteria, linking the syntactic boundaries to potential IP boundaries.

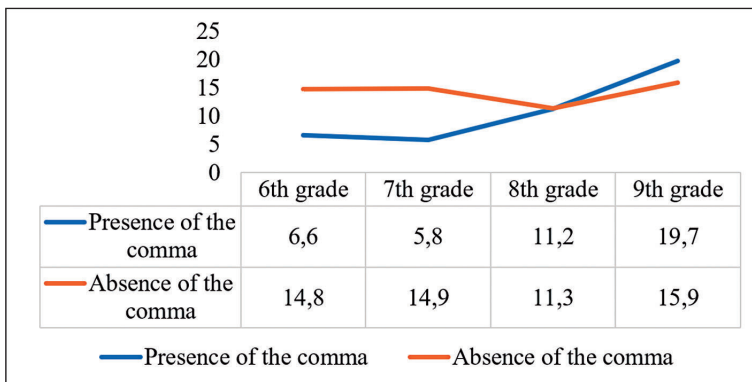
### **Syntactic-prosodic characteristics of commas**

The data description and analysis are organized in three steps, in order to address all variables investigated in this paper. In the first step, we address the overall distribution of conventional and non-conventional commas and the changes in the usage rate of this punctuation mark depending on the school year. Afterwards, we describe the relation between the changes in comma usage observed over time and the identified types of syntactic boundaries. In the last step, we analyze the types of prosodic boundaries involved in the contexts of presence or absence of commas, taking into account the types of syntactic boundaries.

Regarding the contexts in which commas were or should have been used in a simple schema, in the analyzed sample of texts, we recorded 3,285 contexts, 1,422 (43.3%) of which are related to the conventional presence of comma and 1,863 (56.7%) are characterized by its non-conventional absence. Although the non-conventional absence of comma is predominant in the MS sample, we will show that the presence of comma prevails in certain prosodic and syntactic boundaries and, on the other hand, absence of comma tends to be observed when the prosodic IP boundaries can be restructured.

In regard to the longitudinal distribution of the identified contexts, in graph 1, we present the percentual frequencies of presence and absence of comma per school year. The graph suggests a reversed behavior of the distribution of conventional presence as compared to non-conventional absence of that punctuation mark. In the first two years of MS (fifth and sixth grade), the texts are primarily marked by the absence of comma, in accordance with what had been previously reported by Carvalho (2018), developed using a smaller sample of texts of the same textual typology originating from the same database. In the seventh grade, a change in the quantitative data is detected: the texts displayed a higher frequency of contexts in which commas should be used and, in this set, they are predominantly conventional. In the final year, there is a slight predominance of conventional commas over non-conventional commas in the sample.

**Graph 1 – Data percentage per school year**



Source: Author’s elaboration.

This longitudinal data distribution allows us to: (i) capture an expected characteristic, which nevertheless is not featured in literature, about comma in MS, which moves from non-conventional absence in the first year to conventional presence in the final years of the cycle; and (ii) measure the effect of the school years on data about comma in texts from public schools. These results answer the first question: “what are the characteristics of the use of commas in simple schema in productions written by students over the course of MS?” which we intend to address in this paper. Aiming to verify the validity of the result described in (ii), we performed the Kruskal-Wallis test (H), whereby we compared the mean values of presence and absence of comma per school year.<sup>9</sup> The test had statistical support ( $p < 0.05$ ), meaning that the variable “school year” interferes in the presence or absence of comma, so that more years of schooling result in the increase of conventional comma uses.

Since the percentual frequency and the mean value of data types have opposite behaviors and the H-test suggests that the school year has an impact on the occurrence of comma, we also investigated whether there was statistical support to detect which school years have a greater effect on comma usage. Therefore, we performed the Bonferroni Post-Hoc statistical test, whose result can be seen in chart 2. Based on the comparison between the mean values of comma presence and absence per school year, the test shows that using conventional and non-conventional commas in the 8<sup>th</sup> grade of MS differs significantly from the way commas are used in all other school years ( $p < 0.05$ ).

<sup>9</sup> Mean values of conventional presence of comma per school year: 5<sup>th</sup> grade: 0.06; 6<sup>th</sup> grade: 0.05; 7<sup>th</sup> grade: 0.11; 8<sup>th</sup> grade: 0.19. Mean values of non-conventional absence of comma per school year: 5<sup>th</sup> grade: 0.14; 6<sup>th</sup> grade: 0.14; 7<sup>th</sup> grade: 0.11; 8<sup>th</sup> grade: 0.15.

**Chart 2** – Bonferroni Post-Hoc test: comparison between differences in presence/absence of comma across school years

School years	Data types	
	Presence of comma	Absence of comma
	value-p	value-p
6-7	n/s*	n/s*
6-8	n/s*	n/s*
<b>6-9</b>	<b>0.000012</b>	<b>0.000012</b>
7-8	n/s*	n/s*
<b>7-9</b>	<b>0</b>	<b>0</b>
<b>8-9</b>	<b>0.0258</b>	<b>0.01458</b>

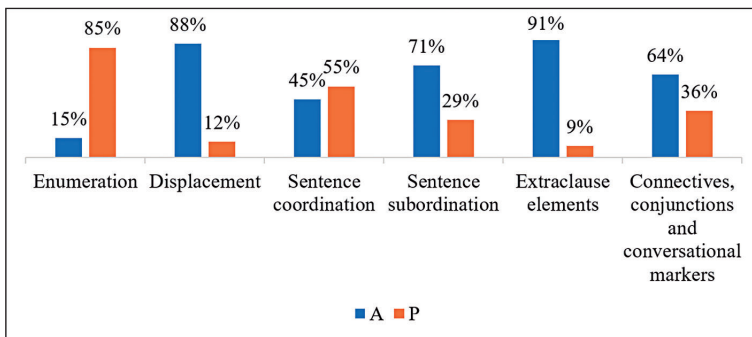
**Source:** Carvalho (2019, p. 98).

**Legend:** \*n/s: contexts in which the difference between the mean values of usage types of the comma was not significant.

In our view, this increase in the conventional use of comma in the 8<sup>th</sup> grade is the result not only of the students' schooling time, but possibly a consequence of the systematic teaching of comma usage in this particular school year, considering the program content in effect at the time when the analyzed texts were produced. Concerning the program content of the 8<sup>th</sup> grade, the Curricular Proposition of the State of São Paulo (2008) established that punctuation should be addressed in three of the four bimesters. We assume that the increase in conventional comma usage by the end of MS could be motivated by the performed didactical/pedagogical work, since these texts have been produced by students who did activities in a state school, following the guidelines of the state's curricular proposition. This is a hypothesis supported by the activity of the first author of this paper as coordinator of a project where reading and text production activities were performed, along with the coordination, so that, through this role, she had systematic contact with the teaching staff in charge of the Portuguese language classes of the school where the activities took place. We ought to clarify that proving this hypothesis is not possible for us due to the fact that we do not have access to the didactical/pedagogical plan of punctuation classes at that school. Obtaining data which could prove or refute the explanatory hypothesis for the results reported above is therefore a topic for future research aiming to measure the effect of activities that teach comma usage in written productions in the school environment.

We would like to underline the fact that we did not only determine a potential positive effect of the didactical/pedagogical work as it led to conventional comma usage, but also oscillation effects between presence and absence of the comma for a given set of syntactic contexts. In graph 2, we present the percentual frequency of presence and absence of the comma for a set of syntactic boundaries detected in the sample.

**Graph 2** – Percentage of data types of comma at syntactic boundaries



Source: Carvalho (2019, p. 102).

Legend: P: presence of comma; A: absence of comma.

We have identified six syntactic contexts for which presence or absence of comma occurs in the investigated texts, which are illustrated below with excerpts from the sample:

a. enumeration:

“*Preenche o formulário com sua idade, nome, gênero, país, avós, irmãos [...]*”;  
[Fill out the form with your **age, name, gender, parents, grandparents, siblings [...]**]

b. displacement:

“*No ano de 2009, eu conheci uma pessoa [...]*”;  
[**In the year of 2009**, I met a person [...]]

c. sentence coordination:

“*Eu acordei, lavei o rosto e escovei os meus dentes [...]*”;  
[**I woke up, washed my face** and brushed my teeth [...]]

d. sentence subordination:

“*Se eu fosse imortal, seria ótimo porque eu poderia fazer tudo o que quisesse*”;  
[**If I were immortal**, it would be great because I could do everything I liked]

e. extraclause elements:<sup>10</sup>

“*Oi, primo, tudo bem com você?*”  
[**Hi**, cousin, everything alright with you?]

f. connectives, conjunctions and conversational markers:

“*Então, ela foi na minha frente*”.  
[**So**, she went ahead of me]

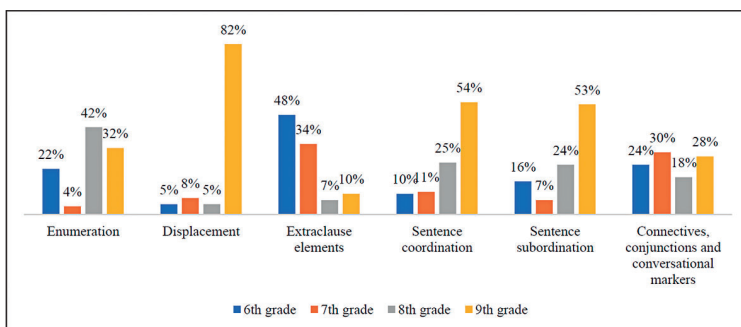
<sup>10</sup> In “extraclause elements”, we grouped comma occurrences at boundaries of elements which are not syntactically linked to the sentence, such as “olá, oi, tudo bem, tchau” [hello, hi, ok, bye], among others.

From graph 2, we would like to highlight that, in the context of “enumeration”, the comma tends to be predominantly conventional (85%), even though there is still absence in some occurrences. The conventional presence of comma is also higher (55%) compared to its absence (45%) in “sentence coordination”. However, in this syntactic context, we find a so-called “oscillation” between having a comma or not at the boundary between clauses.

In turn, comma absence is the predominant characteristic for all other identified syntactic boundaries, although the percentages of absence vary depending on the syntactic boundary: 64% of absence at the boundary of “connectives, conjunctions and conversational markers”; 71% of absence at the boundary of “sentence subordination”; 88% of absence at the “displacement” boundary; 91% of absences at the boundary of extraclause elements”. These results answer the second question: “which syntactic structures support the presence of comma and which ones favor the absence of comma?” presented in the introduction of this paper. Our interpretation is that these results suggest that some syntactic boundaries are identified to a lesser degree by students of MS as locus where commas are expected. Further ahead, we shall analyze these syntactic structures, considering the configurations in prosodic domains in order to determine a possible motivation for these absences.

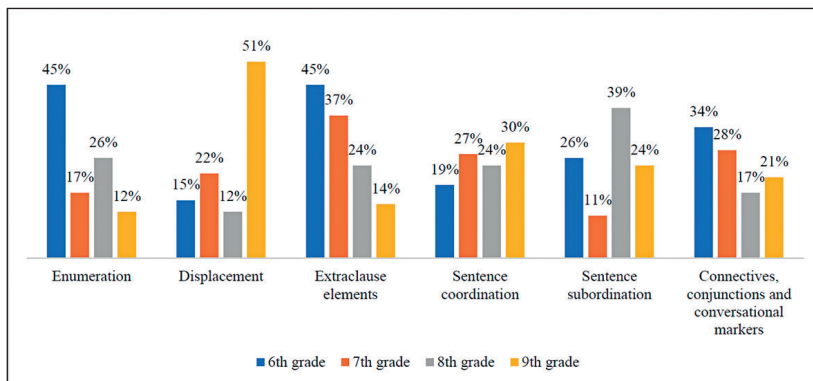
For now, we will resume the longitudinal data description through graphs 3 and 4. In these graphs, we outline the relation between the other variables taken into consideration within the quantitative analysis, namely: the type of comma data, the type of syntactic boundary and the school year. In order to check the association between these variables, we performed the Chi-squared association test ( $X^2$ ). We obtained a statistical confirmation of the association between the variables: for conventional presence of comma, the value  $p=1.03284E-42$  ( $<0.05$ ); for non-conventional absence, the value  $p=1.15098E-40$  ( $<0.05$ ). In both cases, the subsequent calculation of the Pearson contingency coefficient indicates a moderate association between the variables: 0.38 for conventional presence and 0.33 for non-conventional absence. These results support the assertion that presence and absence of the comma tend to vary depending on two factors: in which school year the text is produced and at which syntactic boundary the comma is or should be used.

**Graph 3** – Presence of comma at syntactic boundaries per school year



Source: Carvalho (2019, p. 116).

**Graph 4** – Absence of comma at syntactic boundaries per school year



Source: Carvalho (2019, p. 116).

Patterns in comma usage regarding the school year and the syntactic boundary can be seen in the comparative analysis between the charts 3 and 4. In chart 3, we notice that more commas are used in the 5<sup>th</sup> and 6<sup>th</sup> grade at the same syntactic boundaries, although with little differences in percentage, namely: in the 5<sup>th</sup> grade, there are commas at the boundaries of “extraclause elements” (48%) and of “connectives, conjunctions and conversational markers” (24%); in the 6<sup>th</sup> grade, at boundaries of “extraclause elements :” (34%) and of “connectives, conjunctions and conversational markers” (30%). Since we are dealing with the same students, producing the same textual typology across different school years, we detected the desired increase in conventional presence of comma from the 5<sup>th</sup> to the 6<sup>th</sup> grade when considering the boundary of “connectives, conjunctions and conversational markers”, but found a decrease in the presence of commas from the 6<sup>th</sup> compared to the 5<sup>th</sup> grade, when considering the boundaries of “extraclause elements”. Our hypothesis is that this data points towards other factors at stake, as we intend to explain in the sequence.

Still considering graph 3, we noticed that the highest percentage of commas occurs in the 8<sup>th</sup> grade for all identified syntactic boundaries, except for cases of enumeration, for which a greater presence of commas is found in texts of the 7<sup>th</sup> grade. Furthermore, we observed that these higher presence rates of comma in the 8<sup>th</sup> grade vary depending on the syntactic structure, namely: 82% at the “displacement” boundary; 54% at the boundary of “sentence coordination”; 53% at the boundary of “sentence subordination”; 32% at the “enumeration” boundary; 28% at the boundary of “connectives, conjunctions and conversational markers”; and 10% at the boundary of “extraclause elements”. This set of data leads us to the question: what other factors could be leading to these percentages of presence of comma? We should add that, in the case of “extraclause boundaries”, the data points to a decrease in the presence of comma over the course of the school years, suggesting, at a first glance, a process of regression, and not of learning, in comma usage.



In order to move on in the analysis, it is important to take graph 4 into account, in which the percentages of absence of commas depending on the syntactic boundaries are given. Considering again the boundaries of “extraclause elements”, we found that non-conventional absence of comma gradually decreased throughout the school years. In the same direction of decrease of absence of commas between the initial and final years of MS, we see the data at boundaries of “enumeration” and of “connectives, conjunctions and conversational markers”. At the boundaries of “sentence subordination”, in turn, the data suggests a reasonable stability in the absence of comma between the 5<sup>th</sup> grade (26%) and the 8<sup>th</sup> grade (24%), while there are some increase setbacks in the 7<sup>th</sup> grade (39%), as opposed to a decrease in the 6<sup>th</sup> grade (11%). Another data configuration appears for two other syntactic boundaries: at “displacement” boundaries, the absence of comma concentrates on the 8<sup>th</sup> grade (51%) compared to the other school years; at boundaries of “sentence coordination”, the absence of comma in the 8<sup>th</sup> grade (30%) is higher than in the other school years, even though the absence of commas is distributed in relatively close percentages across school years: 19% (in the 5<sup>th</sup> grade) and 27% (in the 6<sup>th</sup> grade).

From the comparison between the graphs, another result which ought to be highlighted is that, while identifying the syntactic positions where commas are or should be used, we consequently also observed the emergence of a set of syntactic structures in the texts of MS students. We noticed that structures involving a lower syntactic complexity in their way of functioning, such as enumerations, are already common at the beginning of written productions in MS; on the other hand, structures with greater syntactic complexity, such as, for instance, displacements of syntactic structures to the initial position of the sentence or sentence subordination, are more frequent in texts at the end of MS.

We proceed now to establishing a relation between syntactic boundaries, where presence and absence of commas have been located, and prosodic boundaries set forth by the model of Prosodic Phonology (NESPOR; VOGEL, 1986). We found that, in this study, in 100% of the cases, syntactic boundaries in which commas are expected overlap with potential prosodic IP boundaries, according to the formation algorithm of this prosodic constituent. This result confirms previous results by Araújo-Chiuchi (2012), who analyzed texts from the 5<sup>th</sup> grade, by Soncin (2014), who analyzed texts of argumentative typology from the 8<sup>th</sup> grade, and by Tenani and Paiva (2020), who analyzed commas in double schema in texts of argumentative typology from the 8<sup>th</sup> grade. In these studies, the IP boundary also showed a greater relevance for the treatment of types of comma usage. However, we state that prosodic factors underlying the IP domain sometimes lead to presence of the comma and sometimes to its absence in the same syntactic boundary. We shall proceed to discuss this aspect next.

Initially, we will describe a case for which the conventional presence of the comma was more significant and afterwards we will address the boundaries for which absence was predominant. In (4), we provide an example of enumeration structure, which displayed a predominance of conventional commas in the sample. Through the

analysis of these structures, we will show which prosodic characteristics related to the IP domain prompt the presence and absence of comma. We would like to remind the reader that this analysis takes the IP formation algorithm as its foundation, which is formed based on syntactic information, enabling us to link syntactic to prosodic boundaries and consider potential factors which support the restructuring process of the IP domain.

- (4) [Eu prefiro ser normal, e não imortal.]IP [Porque tem o ciclo da vida, nós nascemos, **[bebê,]**IP **[criança,]**IP **[pré-adolescente,]**IP **[adolescente,]**IP **[jovem,]**IP **[adulto,]**IP **[velho]**IP [e depois falece.]IP [Este que é o ciclo da vida.]IP  
 [I would rather be normal, and not immortal.]IP [Because there is the circle of life, we are born, **[baby]**IP **[child,]**IP **[tween,]**IP **[teenager,]**IP **[juvenile,]**IP **[adult,]**IP **[senior]**IP [and then we die.]IP  
 (Text: Z10\_7C\_24M\_01)

According to the IP formation algorithm, this prosodic domain can be defined by the concatenation and/or juxtaposition of elements which are not necessarily syntactically linked to the root sentence. The cases of enumeration, such as (4), provide examples of this circumstance of concatenated and juxtaposed terms, which, within the framework of Prosodic Phonology, have the potential to become non-final IP boundaries. Not only can the listed elements establish prosodic boundaries with this configuration, but, in the above-mentioned example, we also find syntactic boundaries of comma usage. Therefore, our interpretation is that the overlapping between the prosodic boundary and the syntactic boundary helps recognize and identify the locus of conventional comma usage.

For the analysis of prosodic boundaries related to the absence of comma, in turn, we used the examples (5) to (8). In each example, the IP boundaries defined by the formation algorithm of this constituent are indicated. The syntactic structures in question are stressed in bold and the absence of commas is indicated by “( )”.

- (5) [**A**]IP ( ) [*agora lembrei*]IP [*you don't know how to use the Internet:*]IP [**ó**]IP ( ) [*você clica no MSN*]IP [*type in Lala zonasul 2008 @ hotmail.com*]IP. [**af**]IP ( ) [*vai aparecer a minha foto*]IP [**af**]IP ( ) [*vai aparecer o lugar de escrever*]IP [...].  
 [**A**]IP ( ) [now I remembered]IP [you don't know how to use the Internet:]IP [**uh**]IP ( ) [you click on MSN]IP [type in Lala zonasul 2008 @ hotmail.com]IP. [**then**]IP ( ) [my picture will appear]IP [**then**]IP ( ) [the space to write will appear]IP [...].  
 (Text: Z08\_5B\_15F\_04)

(6) [**Hoje**]IP ( ) [*a Gabi não estuda mais aqui por culpa do pai dela.*]IP [*Mas a gente mantém contato pela internet no MSN twitter e Orkut*]IP.

[**Nowadays**]IP ( ) [Gabi doesn't attend school here anymore because of her father.]IP [But we keep in touch over the Internet by MSN twitter and Orkut]IP.  
(Text: Z11\_8E\_31F\_01)

(7) [**Se eu fosse imortal**]IP ( ) [*eu ia fazer coisas,*]IP [*trabalhar,*]IP [*comprar casas,*]IP [*carros,*]IP [*fazenda e a coisa que eu mais quero*]IP [*um cavalo*]IP.

[**If I were immortal**]IP ( ) [I would do stuff,]IP [work,]IP [buy houses,]IP [cars,]IP [a farm and the thing I want most]IP [a horse]IP.  
(Text: Z10\_7C\_0M\_01)

(8) [**Então**]IP ( ) [*ela foi na minha frente,*]IP [*ainda enrolada na toalha,*]IP [*para que pudesse usá-la logo depois que saísse,*]IP [**então**]IP ( ) [*ela foi fazer aquele zig zag correndo para que não pudesse molhar a toalha,*]IP [...].

[**So**]IP ( ) [she went ahead of me,]IP [still wrapped in the towel,]IP [so I could use it right after she left,]IP [then]IP ( ) [she did that zigzag racing so she wouldn't wet the towel,]IP [...].  
(Text: Z11\_8A\_02F\_01)

In these examples, the IP boundaries are defined based on the syntactic constitution of the enunciations, according to the formation algorithm of the domain. Again, here we have examples of non-final IP boundaries. The short length of each structure provided as example acts in the configuration of the IP domain, given that, within the theoretical perspective followed by us (NESPOR; VOGEL, 1986), this factor supports IP phrasing.<sup>11</sup> Depending on the length of a structure, there is context to restructure the boundary, according to the restructuring algorithm. This is a process in which short structures, with a total amount of less than five syllables, are linked to structures adjacent to them to form an IP.<sup>12</sup> This process which we call restructuring by expansion of the IP has the effect of balancing the size of IPs, so that small basic structures join those adjacent to them, generating an IP of greater extension. In chart 3, based on the examples (9) to (12), we look at the IP phrasing predicted by the formation algorithm, and afterwards, at the IP phrasing predicted based on the algorithm of restructuring by expansion – therefore with two IPs – resulting in an IP of greater extension.<sup>13</sup>

---

<sup>11</sup> Prosodic phrasing is a “function of prosody” “which deals with segmentation of the speech continuum into units” (SERRA, 2016, p. 48).

<sup>12</sup> Elordieta *et al.* (2003) and D’Imperio *et al.* (2005) set a parameter to identify structures of short extension and structures of long extension. Along with these authors, we assume that short structures are those featuring up to five syllables; long structures, in turn, present more than five syllables.

<sup>13</sup> In the setting of Prosodic Phonology, the restructuring by separation of IP is also predicted. This is a process in which structures of long extension, phrased in an IP, tend to divide into two IPs, providing optimization and balance between

**Chart 3** – Examples of non-restructured and restructured IP boundaries

Non-restructured IP boundary	Restructured IP boundary
(9) [aí] IP [vai aparecer a minha foto] IP [then] IP [my picture will appear] IP	(9) [aí vai aparecer a minha foto] IP [then my picture will appear] IP
(10) [Hoje] IP [a Gabi não estuda mais aqui] [Nowadays] IP [Gabi doesn't attend school here anymore]	(10) [Hoje a Gabi não estuda mais aqui] IP [Nowadays Gabi doesn't attend school here anymore] IP
(11) [Se fosse imortal] IP [eu ia fazer coisas] [If I were immortal] IP [I would do stuff]	(11) [Se fosse imortal eu ia fazer coisas] [If I were immortal I would do stuff] IP
(12) [Então] IP [ela foi na minha frente] [So] IP [she went ahead of me]	(12) [Então ela foi na minha frente] IP [So she went ahead of me] IP

Source: Author's elaboration.

In all four contexts, there are syntactic boundaries in which, from the perspective of normative grammars, commas should be used. However, the longitudinal sample reveals that this punctuation mark is frequently absent at these boundaries where restructuring by amplification between IPs is expected. The IP restructuring process triggers non-isomorphism between boundaries: there is a syntactic boundary for comma usage, but there is no prosodic boundary, given the small extension of the syntactic structure. It is precisely this prosodic configuration which, in the analyzed longitudinal sample, leads to the absence of comma.

At the end of this analysis about how prosodic boundaries are related to comma placement and absence, we answer the third and last question posed at the beginning: “to what extent could the presence as opposed to the absence of comma be related to the prosodic organization of the enunciations?”.

We have found that the phonological factor “extension” of the constituent affects the prosodic phrasing of the enunciation and, consequently, has an effect on the absence of commas, meaning that the short extension stimulates IP boundary restructuring, leading to possible absence of commas. This IP boundary restructuring in which two IPs are joined, turning into a larger IP, is predicted in the theoretical framework of Prosodic Phonology and, most importantly, is supported by the basic assertion of the theory, according to which there is no mandatory isomorphism between syntactic and prosodic constituents. Additionally, this interpretation is anchored in works about Brazilian

---

parts of the enunciation. As an example, we quote an excerpt taken from the sample, in which the student uses a comma between a subject and a predicate structure: [...] *eu você minha mãe, fomos no canecão lá na represa* (Texto Z09\_6A\_12M\_07) – [I you my mother, went to the Canecão at the dam]. This is an example of non-conventional presence of the comma, a category which is not investigated in this paper, but is discussed in Carvalho (2019). From the syntactic point of view, commas separating the subject from the predicate are considered inappropriate. From the prosodic perspective, we see the possibility of restructuring by separation of IPs in the example at hand, given that the subject of the sentence in the example (I, you, my mother), besides being compound, can be considered to be of long extension, since it presents 7 syllables, which enables the prosodization of this structure as an independent IP, separated from its predicate, which, in turn, is also prosodized in a separate IP, due to its long extension (11 syllables).

Portuguese, which have shown that extension is a factor which affects the prosodic IP phrasing of spoken enunciations (TENANI, 2002; FERNANDES, 2007; SERRA, 2009). Serra (2016), who also addresses extraclause elements, reports, based on a perception study carried out with speakers of Rio de Janeiro variety, that these elements tend not to be perceived as independent linguistic units, even when they are followed by a pause, since they are prosodized along with the structure which follows them, due to their short extension. In this paper, we have shown to what extent the absence of comma results, to a certain degree, from the effect of prosodic phrasing, which can be predicted based on the IP formation and restructuring algorithms. Nevertheless, this result should not be interpreted as a consequence of the inference of phonetic cues perceived by the students as they write. In another respect, we have argued that the prosodic phrasing of the enunciation of IPs so that they are not necessarily isomorphic to the syntactic constituents from which they originate is part of the speaker/listener/writer's phonological grammar.

This line of argumentation dialogs with the perspectives on writing (CORRÊA, 2004) and punctuation (CHACON, 1998) followed in this study, which assume that written enunciations are related to spoken enunciations in a constitutive manner, i.e. writing has in itself characteristics of speech, given its character, its historical constitution and the coexistence of oral and literate practices. From this theoretical approach to writing, the absence of comma is, in part, a consequence of the optimization of the IP configuration depending on the extension of IPs, since this structure configuration affects not only the spoken enunciations (as a performance effect), but equally the written enunciations, due to the fact that these are composed of enunciative features of speech. In this theoretical and analytical confluence, we consider prosody to be part of language (SONCIN, 2014), and not exclusively part of the spoken modality. In this paper, the concept of prosody is not circumscribed to its phonetic characteristics, since we assume prosody to be a subsystem of the phonological grammar, thus underlying spoken and written enunciations. This makes it possible to capture the way how characteristics of prosodic phonology act in comma usage, as we have shown.

## **Final Remarks**

In this paper, we have addressed comma usage in simple schema in a longitudinal sample of texts from MS. The results described above have answered the three questions which guided this study, resumed in the sequence. To the first question, “what are the characteristics of the use of commas in simple schema in productions written by students over the course of MS?”, the answer is: the progressive increase of conventional presence of the comma is concomitant to the increase of schooling time. To the second question, “which syntactic structures support the presence of comma and which ones favor the absence of comma?”, we have presented the following answer: the conventional presence of comma occurs at boundaries of enumerations and sentence coordination;

the absence of comma occurs predominantly at the limits of syntactic displacement of terms or clauses and at the edges of extracause elements. Finally, the third question, “to which extent could the presence as opposed to absence of comma be related to the prosodic organization of the enunciations?”, we have answered based on the finding that commas are used when the syntactic boundary overlaps the IP boundary; the predominance of non-conventional absence of comma was observed in syntactic boundaries which are not necessarily isomorphic to IP boundaries, especially when the syntactic constituent at whose boundary the comma should be used is relatively short, with an extension which is equal to or smaller than five syllables.

These results of the longitudinal analysis provide a contribution both to linguistic description and to native language teaching, since we do not have reports of any other works exploring the analysis of comma in MS from a longitudinal perspective based on a linguistic analysis of syntactic and prosodic structures. In the field of linguistic studies, the contribution lies in presenting a detailed description of a complex object in regard to its syntactic and prosodic functioning. In the educational context, the results provide information on how the students used commas over the course of MS and what are the difficulties concerning mainly the structure type of term or clause displacement that remain at the end of this teaching level. This knowledge lays the foundation to (re)direct classroom practices regarding the teaching of punctuation and text production. The dissemination of this information to teachers of public and private schools has been made by the authors of this paper by offering a university extension course, for instance.

We have established a relation between presence and absence of comma and prosodic characteristics of language, emphasizing the prosodic IP boundary and the effect of its extension, a relevant aspect for addressing the functioning of commas. We have observed that commas occur in positions which are expected from the grammatical point of view: (i) in the locus of IP boundary and (ii) in the locus where IP boundaries are reorganized as predicted by the restructuring algorithm of this domain. This result is a groundbreaking aspect of this work, since we have shown, through prosodic analysis of boundaries where commas are (or could/should be) used, how the prosodic phrasing of enunciations has an important effect on the organization of the written text.

In view of the results of this study, we present some follow-ups to expand the research on this topic. We consider it important to conduct research which: (i) investigates the role of text genre(s) and their repercussions for the presence and absence of commas in school texts; (ii) quantitatively analyzes data through mixed linear models in order to ascertain the interaction of the described factors, such as boundaries of prosodic constituents, type of syntactic structures, semantic-pragmatic characteristics of enunciations, such as emphasis, in combination with social characteristics of the subjects (gender, age, grade); (iii) expands the study to cover texts from different (public and private) school contexts and different social profiles, taking into account oral and literate practices of the participants and the new MS guidelines.

## Acknowledgments

The first author would like to thank CNPq (Process 304.736/2021-0.) for the research productivity scholarship. The second author would like to thank CAPES (Grant Code 001) and FAPESP (Process 16/25504-0) for the grants received during the research development.

TENANI, L.; CARVALHO, T. Vírgulas em textos de alunos de 11 a 15 anos: um estudo longitudinal. *Alfa*, São Paulo, v.67, 2023.

- *RESUMO: Neste artigo, objetivamos analisar presenças e ausências de vírgulas em uma amostra longitudinal de textos do gênero relato escritos por alunos de 11 a 15 anos do Ensino Fundamental II de uma escola no interior do estado de São Paulo. Partimos da hipótese de que a vírgula é empregada não só em razão de fronteiras sintáticas, mas também prosódicas. Os constituintes prosódicos são formados a partir de informações sintáticas e fonológicas, segundo princípios definidos no modelo teórico da Fonologia Prosódica. A caracterização de presenças convencionais e de ausências não convencionais das vírgulas foi feita a partir da identificação de fronteiras sintáticas em que deveriam ou poderiam ser colocadas vírgulas, tomando-se como referência uma gramática da língua portuguesa. Os resultados apontaram que a presença convencional da vírgula aumentou durante os anos de escolarização, como esperado. Porém, as taxas de presença e de ausência da vírgula se alteraram a depender do tipo de estrutura sintática e de haver fronteira de frase entoacional. As presenças da vírgula ocorreram em posições que predominantemente coincidiram com fronteiras de frase entoacional que também eram fronteiras sintáticas de orações. Já as ausências da vírgula ocorreram sistematicamente em posições que não apresentaram coincidência entre fronteiras sintáticas e prosódicas.*
- *PALAVRAS-CHAVE: vírgulas; sintaxe; prosódia; textos escolares; ensino fundamental II.*

## REFERENCES

ANDRADE, E. Insumos teóricos para enseñar las normas que rigen el uso de la coma. *Lenguaje*, Cali, v.46, n.1, p.149-176, 2018.

ARAÚJO-CHIUCHI, A. C. **Os usos não-convencionais da vírgula em textos de alunos da quinta série do ensino fundamental**. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2012.

BAKHTIN, M. **Estética da criação verbal**. São Paulo: Martins Fontes, 1992.

BECHARA, E. C. **Moderna Gramática Portuguesa**. Rio de Janeiro: Lucerna, 1999.

CARVALHO, T. G. **Usos de vírgulas em textos do Ensino Fundamental II**: um estudo longitudinal. 2019. 175f. Dissertação (Mestrado em Estudos Linguísticos) - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2019.

CARVALHO, T. G. Usos de vírgulas em textos do início do Ensino Fundamental II: distribuição e evidências de características prosódicas. **Estudos Linguísticos**, São Paulo, v. 47, n. 2, p. 292-305, 2018.

CASSANY, D. Puntuación: investigaciones, concepciones y didáctica. **Letras**, Caracas, v.58, p.21-54, 1999.

CATACH, N. **La Ponctuation**. Paris: Presses Universitaires de France, 1994.

CHACON, L. **Ritmo da escrita**: uma organização do heterogêneo da linguagem. São Paulo: Martins Fontes, 1998.

CHAFE, W. Punctuation and the prosody of written language. **Written Communication**, Thousand Oaks, n.4, v.5, p.395-426, 1988.

CORRÊA, M. L. G. **O modo heterogêneo de constituição da escrita**. São Paulo: Martins Fontes, 2004.

CORRÊA, M. L. G. Pontuação: sobre seu ensino e concepção. **Leitura: Teoria & Prática**, Campinas, n. 24, p.52-65, 1994.

CUNHA, C. F.; CINTRA, L. F. L. **Nova gramática do português contemporâneo**. Rio de Janeiro: Lexikon, 2017.

D'IMPERIO, M. *et al.* Intonational phrasing in romance: the role of syntactic and prosodic structure. In: FROTA, S.; VIGÁRIO, M.; FREITAS, M. J. (ed.). **Prosodies**. Berlin: Mouton de Gruyter, 2005. p.59-97.

DAHLET, V. M. B. **As (man)obras da pontuação**: usos e significações. São Paulo: Associação Editorial Humanitas, 2006.

ELORDIETA, G. *et al.* Effects of constituent length and syntactic branching on intonational phrasing in Ibero-Romance. In: INTERNATIONAL CONGRESS OF PHONETIC SCIENCES, 15., Barcelona. **Proceedings** [...], Barcelona: Casual Productions, 2003. p. 487-490.

ESVAEL, E. V. da S. **Pontuação na escrita de universitários**: a função enunciativa da vírgula. Tese (Doutorado em Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2005.

FARACO, C. E.; MOURA, F. M. de. **Gramática**: Fonética e Fonologia, Morfologia, Sintaxe, Estilística. São Paulo: Ática, 1997.



FERNANDES, F. R. **Ordem, focalização e preenchimento em português: sintaxe e prosódia**. Tese (Doutorado em Linguística) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2007.

FROTA, S. **Prosody and focus in European Portuguese: Phonological phrasing and intonation**. New York: Garland Publishing, 2000.

FUCHS, S.; KRIVOKAPIC, J. Prosodic Boundaries in Writing: Evidence from a Keystroke Analysis. **Frontiers in Psychology**, Lausanne, n.7, p.1678, 2016.

LUFT, C. P. **A vírgula: considerações sobre o seu ensino e o seu emprego**. São Paulo: Ática, 1998.

MARCUSCHI, L. A. **Da fala para a escrita: atividades de retextualização**. São Paulo: Cortez, 2001.

NESPOR, M.; VOGEL, I. **Prosodic Phonology: with a new foreword**. Berlin: Mouton de Gruyter, 2007.

NESPOR, M.; VOGEL, I. **Prosodic Phonology**. Dordrecht-Holland: Foris Publications, 1986.

PACHECO, V. **O efeito dos estímulos auditivo e visual na percepção dos marcadores prosódicos lexicais e gráficos usados na escrita do Português Brasileiro**. Tese (Doutorado em Linguística) - Instituto de Estudos da Linguagem, Universidade de Campinas, Campinas, 2006.

PRIMUS, B. The typological and historical variation of punctuation systems. **Written Language and Literacy**, Amsterdam, v.10, n.2, p.103-128, 2007.

ROCHA, I. L. V. O sistema de pontuação na escrita ocidental: uma retrospectiva. **DELTA: Documentação de Estudos em Linguística Teórica e Aplicada**, São Paulo, v.13, n.1, p.83-118, 1997.

ROCHA LIMA, C. H. **Gramática Normativa da Língua Portuguesa**. Rio de Janeiro: José Olympio, 1986.

SANDALO, F.; TRUCKENBRODT, H. Some notes on Phonological Phrasing in Brazilian Portuguese. **DELTA: Documentação de Estudos em Linguística Teórica e Aplicada**, São Paulo, v.19, n.1, p.1-30, 2003.

SÃO PAULO (Estado). Secretaria Estadual de Educação. **Proposta curricular do estado de São Paulo: língua portuguesa: ciclo II**. São Paulo, 2008.

SERRA, C. R. A interface prosódia-sintaxe e o fraseamento prosódico no português do Brasil. **Journal of Speech Science**, Campinas, v.5, n.2, p.47-86, 2016.

SERRA, C. R. **Realização e percepção de fronteiras prosódicas no português do Brasil: fala espontânea e leitura**. Tese (Doutorado em Linguística) - Faculdade de Letras, Universidade Federal do Rio Janeiro, Rio de Janeiro, 2009.

SIMARD, M. **Étude de la distribution de la virgule dans les phrases de textes argumentatifs d'expression française**. Dissertação (Mestrado em Linguística) - Université du Québec à Chicoutimi, Saguenay, 1993.

SONCIN, G. C. N. **Língua, discurso e prosódia: investigar o uso da vírgula é restrito? Vírgula!**. Tese (Doutorado em Estudos Linguísticos) - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2014.

SONCIN, G. C. N.; TENANI, L. E. Emprego de vírgulas e prosódia do português brasileiro: aspectos teórico-analíticos e implicações didáticas. **Filologia e Linguística Portuguesa**, São Paulo, v. 2, n. 17, p. 473-493, 2015.

STEINHAUER, K.; FRIEDERICI, A. Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. **Journal of Psycholinguist Research**, New York, v.30, n.3, p.267-295, 2001.

TENANI, L. E. **Prosódia e escrita: uma análise a partir de (hiper)segmentações de palavra**. 2016. 171f. Tese (Livre-docência em Fonologia) - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2016.

TENANI, L. E. **Banco de Dados de Escrita do Ensino Fundamental II**. [São José do Rio Preto: Unesp/GBD, [2015]. Disponível em: <https://www.convenios.grupogbd.com/redacoes/Login#banco>. Acesso em: 31 jan. 2022.

TENANI, L. E. **Domínios prosódicos no português do Brasil: implicações para a prosódia e para a aplicação de processos fonológicos**. 2002. 317f. Tese (Doutorado em Estudos Linguísticos) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2002.

TENANI, L. E.; PAIVA, N. C. Vírgulas em esquema duplo em textos argumentativos: aspectos sintáticos e prosódicos. **Veredas: revista de estudos linguísticos, Juiz de Fora**, v. 24, n. 3, p. 294-321, 2020.

THIMONNIER, R. **Code orthographique et grammatical**. Paris: Hatier, 1970.

VERDEGUER, J. R. **Análisis de los signos de puntuación en textos de estudiantes de educación secundaria**. Tese (Doutorado em Linguística) - Universitat de València, València, 2010.

YANO, C. T. **A história do emprego de vírgula do português clássico ao português europeu moderno**. Tese (Doutorado em Linguística) - Instituto de Estudos da Linguagem, Universidade de Campinas, Campinas, 2018.

Received on February 6, 2022

Approved on May 23, 2022