



# FrameNet annotation as a means to identify genre-relevant linguistic structures

**NATÁLIA SIGILIANO**

Federal University of Juiz de Fora  
natalia.sigiliano@ufjf.edu.br

**TIAGO TORRENT**

Federal University of Juiz de Fora  
tiago.torrent@ufjf.edu.br

## — RÉSUMÉ

Dans la recherche sur l'enseignement du portugais brésilien comme L1, on préconise depuis longtemps de fonder l'étude de la grammaire sur des textes qui présentent une structure linguistique donnée de manière particulièrement saillante. Parmi les exemples d'application de cette approche, on retrouve notamment l'enseignement de l'impératif par le truchement de recettes de cuisine. Cela dit, dans les faits, l'application de cette approche demeure rare ; en effet, les manuels abordent la plupart des notions de grammaire en recourant à des phrases retirées de leur contexte d'usage réel. L'une des raisons qui expliquent cette réalité réside dans le difficile accès à des descriptions systématiques des propriétés grammaticales de textes entiers en portugais brésilien. Cela étant, cette recherche vise à créer un logiciel en accès libre grâce auquel il serait possible de rechercher des structures grammaticales dans des textes annotés pour les extraire à des fins d'enseignement des langues. Le protocole d'annotation employé respecte les directives d'annotation lexicographique et constructionnelle de FrameNet Brasil. Dans cet article, nous présentons une étude pilote basée sur l'annotation d'un guide de voyage et d'un TED Talk. Le résultat ainsi obtenu permet de découvrir les types de structures grammaticales les plus pertinentes chez ces deux genres textuels. En outre, nous considérons l'opportunité de FrameNet pour repérer des structures grammaticales intrinsèquement liées à des genres textuels précis, de même que pour montrer de quelle manière ces dernières correspondent à la structure sémantique du texte.

## MOTS-CLÉS

**annotation dans FrameNet, enseignement des langues, analyse linguistique, genres textuels**

## — ABSTRACT

Research on the teaching of Brazilian Portuguese as L1 has long advocated basing the study of grammar on texts in which a given linguistic structure is highly prevalent. One example of an application of this type is the use of recipes as a means of teaching the imperative mood. However, these kinds of applications are still rare, since manuals approach most grammatical structures using sentences removed from their context of actual usage. One of the factors which contribute to this scenario is the lack of easily accessible systematic descriptions of the grammatical properties of full texts in Brazilian Portuguese. This research, therefore aims to build a freely available computational resource in which texts annotated for the grammatical structures instantiated in them can be searched and extracted for language teaching purposes. Annotation follows the lexicographic and constructional annotation guidelines of FrameNet Brasil. In this paper, we present a pilot study based on the annotation of a travel guide and a TED Talk, showing which kinds of grammatical structures are more relevant to these genres. We also discuss the potential of FrameNet to identify genre-relevant grammatical structures and to show how they pair to the semantic structure of the text.

## KEYWORDS

**FrameNet annotation, language teaching, linguistic analysis, textual genres**

## 1. Introduction

Since the 1980s, strategies for teaching Brazilian Portuguese as a first language for school children have changed significantly. Due to contributions from academic research, changes in the official documentation defining the parameters of the basic school curriculum, and students' lack of interest in studying their language, teachers have put a considerable amount of effort into changing their perspective towards work with Brazilian Portuguese in their classes (Neves 2002; Antunes 2003).

According to Neves (2002: 231), schools are expected to support children in the accumulation of experience and in mastering skills that will provide them with the competence to produce appropriate texts in diverse communicative situations. That is to say that schools, mainly during language classes, should contribute to the students' growing ability to interact in different social contexts.

Until the 1980s, language pedagogy in Brazil was predominantly prescriptive and normative. The focus of classes was on the analysis of sentences isolated from their context, and extracted from literary texts by famous Brazilian and Portuguese authors. These sentences were used as parameters and examples for proposing grammar rules. Such a methodology imposed a great disparity between the language that students were used to in their everyday interactions, and that which was taught in school. After the 1980s, the above mentioned changes led teachers to understand that grammar could be better understood by students if taught based on everyday texts, both written and spoken. This is because, instead of extracting rules from isolated sentences, students are invited to approach the text from the perspective of its communicative purpose. Therefore, studying language structure equates to studying the means to convey ideas and to talk about real issues in society.

Nevertheless, this perspective is still not incorporated into teachers' practices. Many of them report feeling insecure when teaching grammar in a non-prescriptive way, since they were neither trained to do it differently, nor to find alternative teaching methods in the pedagogical materials made available to them. Put differently, although teachers already recognize the importance of the change in methodological perspective suggested by the National Curricula Parameters (Brasil 1998) and by relevant research, they are not capable of implementing it. This is mainly because they do not find support in the teaching materials distributed to schools by the Brazilian

Government. Although such materials sometimes bring innovative strategies for language teaching, they are mostly based on a normative perspective of grammar.

To contribute to the development and distribution of grammar teaching materials grounded on real everyday texts, the research project, which includes this work, aims to develop a freely available computational resource in which texts annotated for the grammatical structures instantiated in them can be searched and extracted for language teaching purposes. Annotation follows the guidelines of FrameNet Brasil. In this paper, we present, in section 2, the theoretical grounding of the work, discussing how the theories of Linguistic Analysis and Genre interrelate. In section 3, we describe FrameNet annotation, and in section 4, we present the materials and methods used in a pilot study based on the annotation of a travel guide, whose results were then compared with those obtained from the annotation of a TED Talk. Section 5 shows which kinds of grammatical structures are more relevant to these genres, and, in section 6, we discuss the potential of FrameNet to identify genre-relevant grammatical structures and to show how they pair to the semantic structure of the text.

## **2. Linguistic Analysis meets Genre Theory**

Grammar teaching is one of the main purposes of Portuguese classes in Brazil. However, teachers have been recognizing the importance of investing time in reading and writing practices, due to the unsatisfactory results of Brazilian students in reading and writing evaluations, and to the inconsistencies present in normative grammars of the language, as pointed out by substantial research in linguistics (Mendonça 2016).

According to Antunes (2007), because language is made of a lexicon and a grammar which materialize in texts shaped by rules of social interaction, to study the grammar of a language, students need to engage in reading and writing practices effectively. This means that learners need to use grammar rules while speaking, reading and writing different textual genres with the adequate level of formality. However, some teachers and the manuals they adopt tend to categorize text comprehension and writing, on one side, and grammar, on the other, as opposing teaching perspectives, hence abandoning work with grammar in classrooms. Nevertheless, as Barbosa (2008) points out, linguistic analysis practices are strongly related to text comprehension and writing practices, and teachers may use a grammar study class to demonstrate

how a given grammar rule or linguistic structure is relevant to a given text (Geraldi 2013). Therefore, the relevance of work with grammar is grounded in the context of production and comprehension of texts.

Bronckart (1999) claims that the appropriation of textual genres is a key mechanism for socialization and participation of the individual in communication activities. In this perspective, genres are defined as socially stable discursive activities that may be used for different kinds of social interaction, including those involving power. It is then essential to recognize the relevance of working with textual genres in school to give students the opportunity to become proficient in comprehending and in producing the texts that regulate life in society.

Schneuwly and Dolz (2004) propose a key for grouping textual genres according to different levels of education. They argue that such a grouping is important because it provides students with (i) different ways to achieve proficiency in writing, (ii) the possibility of defining how different genres operate, (iii) knowledge of the variety of language operations that are deeply connected to a group of genres, and (iv) the possibility of practicing everyday actions mimetically. Although these aspects are equally important for the student to become a proficient user of the language, in this paper we focus on the third aspect, that is, on how the relation between language operations and textual genre groupings can be a valid tool for language teaching.

As pointed out by Bakthin (2000: 302)

we learn how to fit our speech into genre forms, and, when we listen to someone speaking, we can from the beginning, from the very first words, predict the genre, guess how long it will take for her to speak and to finish. (our translation)<sup>1</sup>

In this learning process, learners take into consideration not only the format of the genre or their own interpretation, but also the intrinsic linguistic relationships within the genre, which may be altered to meet reader, media or production requirements. In principle, teaching materials for Brazilian Portuguese should provide students with activities focusing on the recognition of such linguistic aspects of textual genres. Even so, the separation between the study of grammar and text-based study is still very common in such manuals.

Consequently, the association between textual genres and language analysis practices has been a promising and necessary area of study if one aims to positively and effectively change Brazilian Portuguese language pedagogy. Some important, although unsystematic, work in this direction has been done both in academia and in the editorial industry. As a result, some language structures are commonly associated with textual genres in teaching manuals. Table 1 shows the most frequently exploited correspondences.

**Table 1**  
Language Structures vs. Textual Genres

Language Structures	Textual Genres
Imperative Mood	Manuals, cooking recipes, ads
Verbal Tenses	Short stories, chronicles, romances
Adjectives	Reviews, biographies
Passive Voice	Science news, signs, ads

Language structures most commonly associated with textual genres

Although these associations are found in some teaching manuals, they are still very scarce, as most language structures that are relevant for the development of linguistic skills are still studied separately from textual genres in schools. Thus, it is usually the case that teachers and authors of teaching manuals first choose which language structure they want to teach and then associate it with example sentences extracted from aleatory textual genres. For instance, when teaching the indicative mood, manuals tend to show random sentences from any textual genre or grouping that has previously been studied. The same holds true for other structures, such as parts of speech, noun and verbal agreement, and so on: any piece of text may be taken as an example of how the structure functions. However, there is generally no consideration about how this structure particularly contributes to that textual genre.

In a recent study, Sigiliano and Silva (2017) analyzed the grammar study sections of all six teaching manuals approved by the Brazilian Government Program for Teaching Manuals (PNLD) to be used in Junior High classes in 2017.<sup>2</sup> One of the criteria in the analysis aimed to investigate whether the work in grammar sections was based on the analysis of isolated words, sentences or whole texts. Authors found that in 6% of cases, the work was

based on words, in 46% it was based on sentences and in 48% it was based on texts. Although the high percentage of sections in which the work was based on texts might point to the fact that textual genres are taken into consideration for the study of grammar, further analyses demonstrated it is still not the case. The same study showed that, of the sections including texts, in 4% of cases there was not a single activity related to reading and comprehension of a text, and in 21% of cases, text comprehension was only marginally explored. In the end, it is only in 41% of the sections in which texts appear that authors propose any—even marginal—correlation between the grammatical structure being taught and the formal characteristics of textual genres or types (argumentation, description, injunction, narrative). Finally, the research showed that the work involving the reading and comprehension of a text as a means of inviting students to think about grammar usage—the so-called reflexive methodology—tends to appear in the study of morphology and language variation, while the study of syntax tends to adopt a more transmissive methodology.

In light of this, in the next sections, we propose a methodology in which FrameNet annotation is used to build a database of textual genres which are annotated for relevant linguistic structures. In doing so, teachers interested in extending the reflexive methodology into work with other aspects of grammar can extract relevant textual genres for their classes.

### 3. FrameNet Annotation

FrameNet (Fillmore *et al.* 2003) was created in 1997 at the International Computer Science Institute, at Berkeley, as a lexical resource for English. It is based on Frame Semantics (Fillmore 1982) and relies on annotated corpus evidence. Multiple FrameNets have now been founded for other languages such as Spanish, Japanese, German, Swedish, Chinese, Korean, Latvian and Brazilian Portuguese.

In Frame Semantics (and therefore in FrameNet), a frame is defined as a system of interrelated concepts. Each concept in the system is a Frame Element (FE) and its relation to the other concepts in the same system is such that, if one FE is made present, the whole system is also made present (Fillmore 1982: 111). Frames are the background against which meaning is to be construed, and, therefore, according to Frame Semantics, “meanings are relativized to scenes” (Fillmore 1977: 59). That is to say, to understand the meaning of a word, one must access the system on which it is based, or,

in FrameNet terminology, the frame *evoked* by the word. As an example, consider the Giving frame<sup>3</sup>, shown in Figure 1.

Figure 1  
The Giving Frame

**Giving**

**Definition**  
A **Donor** transfers a **Theme** from a **Donor** to a **Recipient**. This frame includes only actions that are initiated by the **Donor** (the one that starts out owning the **Theme**). Sentences (even metaphorical ones) must meet the following entailments: the **Donor** first has possession of the **Theme**. Following the transfer the **Donor** no longer has the **Theme** and the **Recipient** does. Barney gave the beer to Moe. \$300 was endowed to the university to build a new performing arts building.

**Example(s)**

**Core Frame Elements**

**Donor [Donor]** The person that begins in possession of the **Theme** and causes it to be in the possession of the **Recipient**.

**Recipient [Rec]** The entity that ends up in possession of the **Theme**.

**Theme [Thm]** The object that changes ownership.

**Non-Core Frame Elements**

**Circumstances [cir]** The **Circumstances** are the conditions under which the **Theme** is given. I give my services free of charge.

**Depictive [dep]** A description of the **Donor**, **Recipient**, or **Theme** given independently of the giving event per se.

Frame report extracted from the FrameNet Brasil website. Available at <http://www.framenetbr.ufjf.br>.

As Figure 1 demonstrates, as well as a prose definition, each frame is defined in terms of the core and non-core frame elements (FEs) within it. Core FEs are mandatory for the instantiation of the frame, while non-core FEs may or may not be present. In the case of the Giving frame, in Figure 1, a DONOR, a THEME and a RECIPIENT must be conceptually present for the frame to be instantiated, while additional CIRCUMSTANCES may also be provided. Verbs like give, in English, and its equivalent *dar* in Brazilian Portuguese, evoke the Giving frame. Once frames are modeled in FrameNet, sentences containing the words evoking them can be annotated for the FEs in the relevant frame. As an example, consider sentence (1).

(1) Ela deu dinheiro pra mim viajar<sup>4</sup>  
 She give.PST.3SG money for me.DAT travel.INF  
 « She gave the money for me to travel »

Since *dar.v* evokes the Giving frame, (1) could be annotated in FrameNet Brasil as shown in Figure 2.



**Figure 2**

FrameNet Brasil Annotation for the verb *dar* evoking the Giving frame

	E	l	a	d	e	u	d	i	n	h	e	i	r	o	p	a	r	a	m	i	m	v	i	a	j	a	r	.
Dar.dar.v																												
FE																												
GF																												
PT																												

Annotation of sentence (1) in FrameNet Brasil

Usually, sentences containing the lexical units (LUs) evoking frames are annotated over three layers: one for the FE materializing in the constituents within the syntactic environment of the LU, one for the Grammatical Function (GF) of such constituents in the sentence, and one for their Phrase Types (PTs). In Figure 2, we see that the subject, or the External Argument (Ext), noun phrase (NP) *Ela* ‘she’ instantiates the DONOR FE, while the Direct Object (ObjD) NP *dinheiro* ‘money’ is the THEME, and Indirect Object (ObjInd) prepositional phrase *para mim* ‘for me’, the RECIPIENT.

Because a single sentence can be composed of several LUs, each one of them evoking a different frame, sentence (1) can be further annotated for the noun *dinheiro* ‘money’, the preposition *para* ‘for’ and the verb *viajar* ‘travel’, which evoke, respectively, the MONEY, PURPOSE and TRAVEL frames. For each LU an Annotation Set is created. The Annotation Sets for (1) are shown in Figure 3.



- c) Constructionally Null Instantiations, which occur when a given FE is not present due to a construction licensing its omission. CNIs are usually related to constructions such as Imperatives, Passives and Desinential Subjects.

Finally, it might also be the case that a core FE is incorporated by the LU itself, as opposed to manifesting as one of its arguments. Typical cases of Incorporation (INC) occur either with denominal verbs such as *encaixotar* ‘to box’, a verb which evokes the **Placing** frame and incorporates the FE **GOAL**, in which the **THEME** is placed by the **AGENT**. Such cases also occur with nouns denoting entities, in which case the FE representing the entity is incorporated. Sentences (2–4) exemplify cases of CNI, DNI, INI and INC respectively.

(2) A cidade convida a um belo passeio: <sup>LU</sup>comece [pela Praça da Liberdade<sub>PLACE/Dep/PP</sub>]. **AGENT** = CNI | **ACTIVITY** = DNI  
The city invites you to take a beautiful stroll: start at Liberty Square.

(3) [O museu<sub>PLACE/Ext/NP</sub>] <sup>LU</sup>oferece [um espetáculo permanente com efeitos especiais de luz e som<sub>ATTRACTION/ObjD/NP</sub>]. **TOURIST** = INI  
The museum offers a permanent show featuring special light and sound effects.

(4) A região do Araguaia oferece aos <sup>LU</sup>turistas a oportunidade de observarem dezenas de espécies de pássaros muito raras.  
**PEOPLE** = INC  
The Araguaia region offers tourists the opportunity to watch dozens of very rare bird species

In (2), the LU *começar.v* ‘start’ evokes the **Activity\_start** frame. In this frame, an **AGENT** starts performing a given **ACTIVITY**. Note that in the example, neither the **AGENT** nor the **ACTIVITY** is explicitly mentioned in the clause headed by *começar.v*. Because the verb is in its imperative form, the omission of the **AGENT** is licensed by the Imperative Construction. Hence, the **AGENT** is labeled CNI in (2). As for the **ACTIVITY** FE, it refers to the stroll mentioned in the previous clause and, thus, is labeled DNI.

Sentence (3), in turn, shows an example of INI. The verbal LU *oferecer.v* ‘offer’ evokes the **Tourist\_potential** frame, in which a **PLACE** offers some **ATTRACTION** to a **TOURIST**. Note that both the **PLACE** and the

ATTRACTION are instantiated in the sentence, but the TOURIST FE is inferred, since the text is a travel guide.

In (4), the FE PEOPLE in the *People\_by\_leisure\_activity* frame is incorporated by the noun *turista.n* ‘tourist’, which denotes a kind of person.

The original purpose of FrameNet annotation was the provision of corpus evidence supporting the descriptions and structure modeled in frames. Because of its multifaceted nature, however, it has been applied to other tasks in Computational Linguistics, such as question answering (Shen and Lapata 2007) and sentiment analysis (Maks and Vossen 2012).

FrameNet has also been applied to language pedagogy efforts, such as the German Frame-Semantics Online Lexicon—G-FOL—(Boas and Dux 2013) and the Language Acquisition Reusing KORP project—Lark—(Pilán and Volodina 2014). In the former, frames are used to organize the lexica of German and English, building semantically informed correspondences between them, while, in the latter, a semantic role identification exercise generator was built based on Swedish FrameNet data.

Our aim here, however, is to investigate a different use of FrameNet in language pedagogy. Because FrameNet annotation is not only about semantics—that is, the FEs—but also about morphology and syntax—PTs and GFs— we hypothesize that annotating full instances of textual genres will result in a collection of information about the linguistic structures that manifest most commonly in a given textual genre. A freely available and searchable online resource featuring all the annotated texts could then provide teachers with relevant information about which textual genres to use when teaching a given linguistic structure, or, alternatively, which language structures to explore when working with a given genre or genre grouping.

To initially assess the outcome of such an initiative, we conducted a pilot annotation experiment of one travel guide, and then contrasted the results of this annotation with those of a different genre. The experiment’s methods are reported in the following section.

#### 4. Material and Methods

To conduct the pilot annotation experiment we selected one travel guide

produced by the Brazilian Government Agency for Tourism—Embratur—for the major international sports events held in Brazil in 2014 and 2016 (Brasil 2012). The guide features 28 chapters: the first is an introduction to Brazil, and each subsequent chapter is dedicated to a Brazilian State and the Federal District of Brasília. The total number of sentences in the guide is around 4 900.

The FrameNet Brasil WebTool was used in the annotation task. A group of four annotators were given the task of annotating the verbal, nominal, adjectival and adverbial LUs in the sentences. Because annotators could not create new frames, annotation was restricted to the LUs evoking a pre-existing frame in the FrameNet Brasil database.

For the first round of annotations, the round reported in this paper, annotators were given the goal of annotating a total of 100 sentences each, during a four-week period. Weekly group meetings were held so that doubts and questions could be shared and answered.

Because the travel guide genre is usually categorized as being part of the injunctive type (Calvi 2010: 18; Costa 2008: 108–109), we anticipated that Imperatives and other instruction-giving strategies—such as instructional infinitives—would have a strong presence in the annotation.

To compare the results obtained in this annotation effort to those from a different genre, we used data from the shared annotation task of the Multilingual FrameNet Project (Baker and Ellsworth 2017). In this task, each FrameNet team was asked to annotate a transcript of the TED Talk “Do schools kill creativity”, which had been translated into Brazilian Portuguese. Annotators were instructed to use only the frames from FrameNet Data Release 1.7, so, once again, no new frames could be created. The Brazilian Portuguese version of the TED Talk features 271 sentences, seven of which had no frame-evoking elements. Hence, 264 sentences were annotated.

Results of the first annotation round are presented and discussed in the following section.

## **5. Results and Discussion**

In the first round of annotation of the travel guide, a total of 380 sentences were annotated for 276 different LUs, evoking 175 different frames. The remaining 20 sentences did not feature any LU evoking an already

existing frame in FrameNet Brasil and therefore were not annotated. Because each sentence is usually annotated for more than one LU (see Figure 3), and because LUs can occur more than once in the text, the total number of Annotation Sets for the travel guide was 628.<sup>5</sup>

For the sake of comparison, the 264 sentences in the TED Talk were annotated for 282 different LUs, evoking 171 different frames, and yielded 526 Annotation Sets.

The results indicate that the TED Talk transcript serves as a good point of comparison for the annotation of the travel guide, since the variation in the number of frames and LUs annotated was minimal, while the difference in the number of Annotation Sets (102) is compatible with the difference in the number of sentences for each genre (116).

The first criterion to be analyzed was the distribution of verbal, nominal, adjectival and adverbial LUs in both texts. Table 2 shows the number and percentage of LUs per part of speech in the annotated sentences in the travel guide, compared to the TED Talk.

**Table 2**  
Lexical Units per POS

POS	Travel Guide		TED Talk	
	Quantity	Percentage	Quantity	Percentage
Verb	182	65.9%	92	32.6%
Noun	78	28.4%	102	36.2%
Adjective	12	4.3%	39	13.8%
Adverb	4	1.4%	49	17.4%
TOTAL	276	100%	282	100%

Number of LUs annotated per POS

Note that 65.9% of LUs in the travel guide are verbs, compared to only 32.6% in the TED Talk. Moreover, among the 78 noun LUs in the travel guide, 15 represent events and are deverbal nouns. In the TED Talk, the number of deverbal nouns is 11. Hence, the travel guide features a total of 197 eventive

LUs (71.4%), in comparison with just 103 (36.5%) in the TED Talk. This is because the main communicative purpose of a travel guide is to provide tourists with information on places they can visit or things they can do. On the other hand, the TED Talk aims to convey an idea, which is supported by a higher number of nouns denoting entities. Sentences (5–7) illustrate this difference. The first two, from the travel guide, are typical descriptions of tourist attractions (5) and service providers (6); the third, from the TED Talk, presents one of the speaker’s remarks (7).

(5) A praia é ideal para famílias e para quem quer descansar e nadar nas suas piscinas naturais.

*The beach is ideal for families and for those who want to rest and swim in its natural pools.*

(6) A pousada localizada no seringal também oferece visitas guiadas.

*The bed and breakfast located in the rubber tree forest also offers guided tours.*

(7) Se você for pensar, todo o sistema de educação pública ao redor do mundo é uma extensão do processo de ingresso à universidade.

*If you think of it, the whole system of public education around the world is a protracted process of university entrance.*

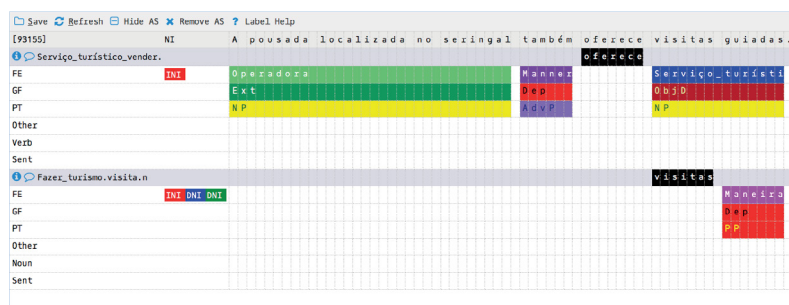
Note that (5) and (6) feature a greater number of verbs (descansar ‘rest’, nadar ‘swim’, oferecer ‘offer’) and an eventive noun (visitas ‘tours’), while the focus of (7) is on the NP sistema de educação pública ‘system of public education.’

Put in typological terms (Bronckart 1999), (5) and (6) are mainly injunctive, since the description of the beach is written with the purpose of suggesting it as a tourist destination for families with children, and that of the bed and breakfast to suggest the guided tours as something visitors can engage in. Example (6) is also explicative, however. Sentences like (5) and (6) are abundant in the travel guide and exemplify Calvi’s (2010: 25) definition of descriptive travel guides as a genre whose purpose is to inform and give advice on what to do.

Sentence (6) exemplifies a second distinctive property of travel guides, which relates to how the injunctive communicative purpose of this genre manifests grammatically. Our intuition was that imperatives would be very

frequent in this genre, and, therefore, the number of CNI labels applied to the TOURIST or other agentive FEs—such as the one shown in (2)—would be high. However, the most frequent NI-type structures found in the text were precisely the ones exemplified in (6). In the first type, a given PLACE, BUSINESS or TRAVEL INFRASTRUCTURE offers some ATTRACTION or SERVICE to the TOURIST. Because the text being annotated is a travel guide and, therefore, meant to be read by tourists, the TOURIST FE was neither overtly expressed in the sentence, nor could it be identified with some specific referent in the preceding context. It was therefore labeled as INI, as shown by the red label in the FE layer of the first Annotation Set of Figure 4. In the second type, an eventive noun indicates some ACTIVITY in which the TOURIST can engage. The ACTIVITY FE is incorporated by the LU, while the TOURIST FE is also labeled as INI, as in the second Annotation Set of the same figure. The total number of INIs in the travel guide was 7 times higher than that of CNIs.

Figure 4  
FrameNet Brasil Annotation for Sentence (6)



This analysis corroborates that of Calvi (2010: 18), to the extent that this author claims that “in fact, the production of a travel guide presupposes the travel experience, which, nevertheless, is hidden, allowing the infinite repetition of the same itinerary.”<sup>6</sup> In FrameNet terms, the hidden nature of the travel experience is manifested in the null instantiation of the tourist and, moreover, in its promotion from one specific traveler—the one who wrote the guide—to a generic traveler, who will be conceptually instantiated by every reader of the travel guide.

In summary, at least for the first round of annotation, the omission of referentially generic arguments was considerably more frequent in the travel guide than that of imperative subjects. This finding provides initial evidence



to the claims that (a) a thorough syntactic-semantic analysis of each textual genre can reveal their most relevant grammatical properties, and (b) that a freely available online resource in which users can browse genres and their relevant linguistic structures may be useful for helping teachers choose texts for preparing language study activities for their classes.

## **6. Conclusions and Outlook**

In this paper, we presented a preliminary experiment deploying FrameNet annotation as a means of identifying genre-relevant linguistic structures. Approximately 400 sentences from a travel guide to Brazilian cities produced by Embratur were annotated by four annotators. Annotation was compared to that of a TED Talk.

Initial findings show that, contrary to our assumptions, the ellipsis of semantically generic arguments seems to be the most relevant linguistic structure found in the travel guide annotated, with the number of INIs being approximately 7 times more frequent than the number of CNIs, which stand for constructionally null arguments, such as those licensed by imperatives.

Our findings point to the fact that the long-term goal of the project to which this experiment is circumscribed—that of creating an online resource of annotated textual genres—may prove to be pertinent in helping teachers find relevant textual material for preparation of language classes.

The next steps in the project include annotating at least one full text of different textual genres belonging to different types, as well as making the texts and the annotations available and searchable online.

## NOTES

1. Aprendemos a moldar nossa fala às formas de gênero e, ao ouvir a fala do outro, sabemos de imediato, bem nas primeiras palavras, pressentir-lhe o gênero, adivinhar-lhe o volume (extensão) e prever-lhe o fim.
2. PNLD is a program run by the Ministry of Education through which course books and teaching manuals produced by Brazilian authors and publishers are evaluated and, if approved, purchased by and distributed to public schools and their students. Every two years, PNLD publishes a call for proposals in which evaluation criteria and content parameters are defined.
3. For notational purposes, frame names will be typed in *Courier* font, while frame element names will be shown in *SMALLCAPS*. All frames mentioned in this paper can be consulted at the FrameNet Brasil website at <http://www.ufjf.br/framenetbr/dados/vocabulario-generico/>.
4. All example sentences provided in this paper were extracted from the FrameNet Brasil corpus.
5. All annotations produced by FrameNet Brasil can be found at <http://webtool.framenetbr.ufjf.br/index.php/webtool/report/frame/main>
6. En efecto, la elaboración de la guía presupone la experiencia del viaje, que sin embargo permanece oculta, permitiendo la infinita repetición del mismo itinerario.

## RÉFÉRENCES

ANTUNES, Irandé (2007): *Muito além da gramática: por um ensino de línguas sem pedras no caminho*. São Paulo: Parábola Editorial.

BAKER, Collin and ELLSWORTH, Michael (2017): Graph Methods for Multilingual FrameNets. In *Proceedings of ACL 2017—Workshop on Text Graphs*. Vancouver, Canada.

BARBOSA, Jacqueline (2008): *O lugar da gramática no currículo de língua materna*. Santa Fé: Ediciones UNL.

BOAS, Hans Christian and DUX, Ryan (2013): Semantic frames for foreign-language education: Towards a German frame-based dictionary. *Veredas On-Line*. 17(1):81-100.

BRASIL (1998): *Parâmetros curriculares nacionais*. Ensino Fundamental II. Brasília: MEC.

BRASIL (2012): *Guia de Cidades 2012*. São Paulo: Empresa Brasileira de Comunicações e Produções.

BRONCKART, Jean Paul (1999): *Atividades de linguagens, textos e discursos: por um interacionismo sócio-discursivo*. Traduit par MACHADO, Anna Rachel et CUNHA, Péricles da. São Paulo: Educ.

CALVI, Maria Vittoria (2010): Los géneros discursivos en la lengua del turismo: una propuesta de clasificación. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*. 19:9-32.

COSTA, Sérgio Roberto (2008): *Dicionário de Gêneros Textuais*. Belo Horizonte: Autêntica Editora.

FILLMORE, Charles J. (1977): The case for case reopened. *Syntax and semantics*. 8:59-82.

FILLMORE, Charles J. (1982): Frame Semantics. In *The Linguistic Society of Korea, ed. Linguistics in The Morning Calm*. Seoul: Hanshin.

FILLMORE, Charles J., PETRUCK, Miriam Rachel-Lea, RUPPENHOFER, Joseph, et al. (2003): FrameNet in action: The case of attaching. *International Journal of Lexicography*. 16(3):297-332.

GERALDI, João Wanderlei (2013): *Portos de passagem*. São Paulo: WMF Martins Fontes.

MAKS, Isa et VOSSEN, Piek (2012): A lexicon model for deep sentiment analysis and opinion mining applications. *Decision Support Systems*. 53(4):680-688.

MARCUSCHI, Luiz Antônio (2008): Gêneros textuais: configuração, dinamicidade e circulação. In *KARWOSKI, Acir Mário, GAYDECZKA, Beatriz and BRITO, Karim Siebeneicher, ed. Gêneros textuais: reflexões e ensino*. Rio de Janeiro: Nova Fronteira.

NEVES, Maria Helena de Moura (2002): *A gramática: história, teoria e análise, ensino*. São Paulo: Editora da UNESP.

MENDONÇA, Márcia (2006): Análise lingüística no ensino médio: um novo olhar, um outro objeto. In BUNZEN, Clécio and MENDONÇA, Márcia, ed. *Português no ensino médio e formação de professores*. São Paulo: Parábola Editorial.

PILÁN, Ildikó and VOLODINA, Elena (2014): Reusing Swedish FrameNet for training semantic roles. In *Proceedings of LREC 2014*. Reykjavik, Iceland.

SCHNEUWLY, Bernard. and DOLZ, Joaquim (2004): *Gêneros orais e escritos na escola*. Traduit par ROJO, Roxane Helena Rodrigues and SALES CORDEIRO, Gláís. Campinas: Mercado de Letras.

SHEN, Dan et LAPATA, Mirella (2007): Using Semantic Roles to Improve Question Answering. In *Proceedings of EMNLP-CoNLL 2007*. Prague, Czech Republic.

SIGILIANO, Natália Sathler and SILVA, Wagner Rodrigues (2017): Diagnóstico de propostas de análise linguística em livros didáticos aprovados em programa oficial. In Magalhães, Tânia, Garcia Reis, Andreia Rezende et Ferreira, Helena, ed. *Concepção discursiva de linguagem: formação e ensino*. Campinas: Pontes Editora

TORRENT, Tiago Timponi and ELLSWORTH, Michael Justin (2013): Behind the Labels: Criteria for defining analytical categories in FrameNet Brasil. *Veredas*. 17(1):44-65.