

Behind the Labels: Criteria for Defining Analytical Categories in FrameNet Brasil

Tiago Timponi Torrent (FrameNet Brasil / UFJF)
Michael Ellsworth (FrameNet / ICSI / UC Berkeley)

ABSTRACT: FrameNet Brasil (FN-Br) aims to develop the Brazilian Portuguese branch of the FrameNet initiative, started by Charles J. Fillmore at the International Computer Sciences Institute, in Berkeley. Since 2007, FN-Br has been developing lexicographic analyses of Brazilian Portuguese based on the principles of Frame Semantics and supported by corpus evidence. This paper reports on the construction of a series of analytical categories supporting lexicographic annotation in FN-Br. We discuss the creation of new phrase type and grammatical function labels for FN-Br, since the original ones were proposed for English. Also, we present the theoretical background behind the creation of such labels.

KEYWORDS: FrameNet Brasil; Lexicographic Annotation; Analytical Categories; Grammatical Functions; Phrase Types

Introduction

FrameNet Brasil (FN-Br) has been developing, since 2007, the Brazilian Portuguese branch of the FrameNet initiative (FILLMORE ET AL., 2003a; 2003b), started in 1997 by Charles J. Fillmore at the International Computer Sciences Institute, in Berkeley. Following the main purposes of the original project, under development for English, FN-Br is creating a lexical resource for Brazilian Portuguese, based on the principles of Frame Semantics (FILLMORE, 1982; 1985) and supported by corpus evidence.

Aside from FN-Br (SALOMÃO, 2009), other non-English framenets are being developed for Chinese (YOU & LIU, 2005), German (BOAS, 2002; ERK ET AL., 2003a), Japanese (OHARA ET AL., 2004), Spanish (SUBIRATS & PETRUCK, 2003) and Swedish (BORIN ET AL., 2010), to name the most advanced initiatives. All of them rely on the same database structure proposed broadly for FrameNet (see BAKER ET AL., 2003 for details): (i) a set of **Frames** composed by a definition and a set of **Frame Elements**; (ii) a set of **Lexical Units** – the pairing of a lemma to a meaning defined in terms of a frame –; and (iii) a set of **Annotated Sentences** that provide evidence for how the Lexical Units are used in the target language in regards to both to their semantic and syntactic valences. In spite of those cases in which a given frame is related to some culturally specific construct (see BERTOLDI, 2011 for a comparative analysis of juridical frames in Brazilian Portuguese and in English), or perspectivized differently by the lexical material in a specific language (see OHARA, this volume) framenets tend to use the same set of frames defined by the original project. Obviously, the same does not hold for the lexical units, since these are specific for each language. As for the categories used for annotating the sentences, although the semantic valences are defined in terms of the Frame Elements, which are dependent to the frame, the syntactic valences are composed of phrase types and grammatical functions, which may vary crosslinguistically.

Although FrameNet-like annotation is designed to require as little theoretical background as possible, because it deals with categories with long tradition in Linguistics, such as grammatical functions and phrase types, and also because it has to make decisions on whether to annotate or not a given constituent in a sentence, one can ask why we work with the set of labels we do and not with another set, or even why we annotate raised subjects and controlling expressions but not anaphoric antecedents.

In this paper, we discuss the definition of these analytical categories for FN-Br, based on those defined for English by the Berkeley FrameNet project. To achieve this goal, this paper begins with a brief presentation of FrameNet's basic concepts, such as **Frame**, **Frame Elements**, **Lexical Units** and **Annotation Sets** in section 1. In section 2, we present the FrameNet Brasil initiative and its relation to FrameNet. In section 3, we discuss the theoretical background supporting the creation of analytical categories for FrameNet-like lexicographic annotation. In section 4, we present the grammatical functions and phrase types used in FN-Br annotations, contrasting them with those used for English. The last section presents the conclusive remarks.

1. FrameNet basics

FrameNet is a computational lexicography initiative, which relies on the theoretical bases of Frame Semantics (FILLMORE, 1982; 1985) and develops syntactic-semantic analyses of lexical items – and, since 2008, of constructions as well (FILLMORE, LEE-GOLDMAN & ROMIEUX, 2012) – supported by corpus evidence (FILLMORE ET AL., 2003a). The main analytical categories used by FrameNet are the **Frame** – together with the Frame Elements – and the **Lexical Unit**.

According to Fillmore's definition in the seminal Frame Semantics paper, a frame can be regarded as

any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text, or into a conversation, all of the others are automatically made available. (FILLMORE, 1982, p.111)

In FrameNet, frames are defined in a way that maintains the basis of Fillmore's definition: they are composed of a textual definition characterizing the broad scene, and of the definitions of the participants and props in the frame, the Frame Elements (FEs). Figure 1 shows the `POSSESSION` frame as it is presented at the FrameNet website.

Some of the FEs are labeled as Core, meaning that they are absolutely necessary for the instantiation of the frame. Non-core FEs are those that may occur when the frame is evoked, usually indicating the circumstances surrounding the instantiation of the frame (see RUPPENHOFER ET AL., 2010 for a complete account of FE types).

In turn, Lexical Units (LUs) are the pairing of a lemma to a frame. In FrameNet, they receive a definition that is either written by the analyst or retrieved from the Concise Oxford Dictionary. LUs are also classified by part of speech. FrameNet annotates for verbs, nouns, adjectives, adverbs and prepositions (RUPPENHOFER ET AL., 2010), as well as for multiword expressions such as *take (one's) life* or *a pain in the neck*, which are classified by part of speech according to their heads.

In the context of Fillmore's (1982) definition, LUs are the linguistic material responsible for introducing a given frame in a text or in a conversation. To analyze how LUs evoke frames and also how FEs are instantiated in real texts, FrameNet annotates sentences in which the target LUs occur, extracting them from large corpora. Sentences are also annotated for their syntactic properties and stored in the database as annotation sets. The sentences that exemplify the frame definition and the FE definitions in Figure 1 show how the FE labels are applied.

Possession [Lexical Unit Index](#)

Definition:

An **Owner** has (or lacks) a **Possession**.

The river's part of my manor, but of course **it** also **BELONGS** **to the River Police**.

It's a nuisance having to abandon **my** **BELONGINGS**, though.

FEs:

Core:

Owner [Own]	The entity that owns a possession.
Possession [Pos]	The thing which is owned by the Owner. Ex.: I HAVE twenty bucks . (This does not include bodyparts or medical conditions, but does include intellectual property, etc.)

Non-Core:

Depictive [Dep]	State of the Owner or Possession .
Duration []	Duration denotes the length of time from the beginning of a continuous situation (the one denoted by the target) to its end. In many cases, the continuous situation is a dynamic action which is ongoing, while in others it is simply an undifferentiated state.
Explanation []	The reason for which the Owner has the Possession .
Manner [Manr]	The way that the Owner possesses the Possession .
Semantic Type: Manner	
Time [Tim]	This FE indicates the Time interval during which the Owner has a possession.

Figure 1: The Possession frame as seen at the FrameNet website (<http://framenet.icsi.berkeley.edu>)

After the annotation sets are analyzed and each part of the relevant example sentences is labeled, the FrameNet Desktop software summarizes the semantic and syntactic valences of the target LUs. Such a summary is presented in the Lexical Entry report, shown in Figure 2, below the information about which frame is paired with the lemma *have.v* in this LU and its definition. The first of the two tables shows the FEs in the `POSSSESSION` frame that were instantiated in sentences containing the target LU, the number of instances annotated for each one, as well as their syntactic realizations in terms of the phrase types in which they occur and of the grammatical functions they play in the sentence. The second table presents the syntactic patterns in which each sequence of FEs observed occur. The numbers in brackets indicate how many instances of each syntactic pattern were observed in the data analyzed.

In section 3, we will discuss the theoretical background supporting FrameNet annotation, and, thus, come back to some of the key concepts presented above.

Lexical Entry

[Annotation](#) [Possession](#)

have.v

Frame: Possession

Definition:

COD: possess, own, or hold

Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realization(s)
Depictive	(4)	AVP.Dep (1) PP[in].Dep (1) PP[in].Obj (1) VPing.Dep (1)
Duration	(2)	PP[for].Dep (2)
Explanation	(3)	PP[in].Dep (2) VPto.Dep (1)
Owner	(153)	NP.Ext (140) NP.Obj (8) CNI.-- (2) DNI.-- (2) INI.-- (1)
Possession	(153)	DNI.-- (1) NP.Obj (146) NP.Dep (4) Sfin.Dep (1) Sing.Obj (1)
Time	(2)	AVP.Dep (2)

Valence Patterns:

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns			
4 TOTAL	Depictive	Owner	Possession	
(1)	AVP Dep	NP Ext	DNI --	
(1)	PP[in] Dep	NP Ext	NP Obj	
(1)	PP[in] Obj	NP Obj	NP Obj	
(1)	VPing Dep	NP Ext	NP Obj	
1 TOTAL	Duration	Owner	Owner	Possession
(1)	PP[for] Dep	NP Ext	NP Ext	NP Obj
1 TOTAL	Duration	Owner	Possession	
(1)	PP[for] Dep	NP Ext	NP Obj	
3 TOTAL	Explanation	Owner	Possession	
(2)	PP[in] Dep	NP Ext	NP Obj	
(1)	VPto Dep	NP Ext	NP Obj	
5 TOTAL	Owner	Owner	Possession	
(5)	NP Ext	NP Ext	NP Obj	

Figure 2: Lexical Entry Report for have.v (<http://framenet.icsi.berkeley.edu>)

2. FrameNet Brasil

Since its foundation in the late 90's, FrameNet has been expanded to other languages, giving rise to several different research initiatives in the Computational Lexicography field. Some of those initiatives, although relying on Frame Semantics and corpus evidence for the construction of their lexicons, have developed their own methodology – and, hence, their own software apparatus – for lexicographic annotation (see, for instance, ERK ET AL., 2003b, on the SALSA annotation tool). Others have chosen to adapt the original FrameNet software, methodology and annotation labels to their target languages. This was the direction followed by Spanish FrameNet, Japanese FrameNet, and FrameNet Brasil as well.

Given such a scenario, in this section, we will present the FN-Br database, discussing how frames and the elements in them are created. We will also present the annotation modes used in both FrameNet and FrameNet Brasil, and, demonstrate how the multilayer annotation is performed in the Brazilian version of the FrameNet Desktop annotation software.

2.1. The FrameNet Brasil Databases

FN-Br currently works with two different databases: a general vocabulary database (GVDB) and a domain specific database (DSDB). The GVDB is being constructed based on Berkeley FrameNet's data release 1.5, which was made available to FN-Br together with the distribution of the FrameNet Desktop software. Frames in this data release, as well as the Frame Elements in them were either directly translated from English into Brazilian Portuguese, or adapted to the later, in those cases in which there are differences regarding perspective, cultural specificity or FE coreness (see, in this respect VAZ, 2012 on the *Communication*, *Communication_manner* and *Communication_means* frames). To ensure the interconnection between FN-Br's GVDB and FrameNet's database, the Brazilian staff maintains the same frame and FE IDs in the translated version. The IDs are the numbers representing each frame and FE in the database tables. By doing so, it will be possible, in the near future, to include FN-Br's GVDB in the FrameSQL tool (SATO, 2008), a search engine capable of comparing frames and their instantiations across languages.

Figure 3 shows the *Posse* frame as seen in FN-Br's GVDB. This frame was translated into Brazilian Portuguese from FrameNet's *Possession* frame, depicted in Figure 1. Since possession is a crosscultural concept, no adaptations were necessary. Besides maintaining the same IDs for the frame and for the FEs, FN-Br also keeps the FE names in English so as to ensure the readability of the frame by non-Portuguese speakers.

Posse		Lexical Unit Index
Definição [Definition]:		
Um Possuidor tem (ou carece de) uma Posse .		
Elementos de Frame [Frame Elements]:		
Nuclear [Core]:		
Posse [Possession]	A coisa que é possuída pelo Possuidor . Ex.: Eu TENHO vinte reais . (Este elemento não inclui partes do corpo ou problemas de saúde, mas inclui propriedades intelectuais, etc.)	
Possuidor [Owner]	A entidade que tem a Posse .	
Periféricos [Non-Core]:		
Descrição [Depictive]	Estado do Possuidor ou da Posse .	
Duração [Duration]	Denota o período de tempo contado a partir do início de uma situação contínua (aquela codificada pelo alvo) até o seu fim. Em muitos casos, a situação contínua é uma ação dinâmica em curso, enquanto, em outros, trata-se simplesmente de um estado indiferenciado.	
Explicação [Explanation]	A razão pela qual o Possuidor tem a Posse .	
Maneira [Manner]	A forma como o Possuidor tem a Posse .	
Semantic Type: Manner		
Tempo [Time]	Este EF indica o intervalo de Tempo durante o qual o Possuidor tem a Posse .	

Figure 3: The Posse (Possession) frame in FN-Br's GVDB (<http://www.framenetbr.ufjf.br>)

The GVDB also contains the constructions created for the Brazilian Portuguese Constructicon (LAGE, 2013), which, for the sake of economy, will not be presented in this paper.

The DSDB, in turn, is being developed within the Copa 2014 FrameNet Brasil project (SALOMÃO ET AL., 2011), which is creating, in cooperation with the Semantec Group at UNISINOS, a multilingual frame-based electronic resource covering the vocabulary of soccer and tourism for Brazilian Portuguese, Spanish, and English. This resource will be used by tourists, journalists and the staff involved in the 2014 FIFA World Cup, to be held in Brazil.

Atração_turística		Lexical Unit Index
Definição [Definition]:		
Um Lugar ou uma Atração , geralmente, apresenta recursos naturais ou artificiais com valor cultural ou histórico que atraem turistas de diferentes origens. O Brasil OFERECE diferentes opções turísticas.		
Elementos de Frame [Frame Elements]:		
Nuclear [Core]:		
Atração [Attraction]	Geralmente, é um espaço físico que faz gerar trânsito de pessoas de diferentes origens para conhecer ou usufruir seus benefícios, o que faz com que seja reconhecido socialmente por apresentar valor histórico ou cultural ou mesmo por oferecer recursos naturais ou artificiais. Em alguns momentos, os benefícios podem ser considerados os próprios atrativos para o Turista . Belo Horizonte OFERECE muitos bares para turistas	
Lugar [Place]	Lugar que abriga recurso natural ou artificial com valor cultural ou histórico.	
Semantic Type: Location		
Turista [Tourist]	Indivíduo ou grupo interessado em conhecer ou desfrutar os benefícios de uma Atração ou de um Lugar .	
Semantic Type: Sentient		
Periféricos [Non-Core]:		
Acompanhante [Co-participant]	Aquele que desfruta os benefícios de uma Atração juntamente com o Turista .	
Semantic Type: Sentient		
Descrição [Depictive]	Características específicas da Atração ou do Lugar .	
Duração [Duration]	O período de duração da atividade turística.	
Semantic Type: Duration		
Finalidade [Purpose]	Alguma ação específica que o Turista pretende realizar com a atividade turística, entretenimento e descanso, por exemplo.	

Figure 4: The Copa 2014 database as depicted in FN-Br's web site (<http://www.framenetbr.ufjf.br>)

As a domain-specific resource, Copa 2014 requires the creation of frames that are more specific than those in the FrameNet database, which covers the general vocabulary. Hence, frames in the Copa 2014 DSDB are being created via a bottom-up approach (see GAMONAL, 2013 for a detailed description of such a process), this is to say that frames were modeled from the sentences extracted from corpora covering the target domains of the dictionary. The result of such an approach is the creation of more detailed frames. If one takes the domain of tourism, for example, while the GVDB has only the *Turismo* (Touring) frame, the DSDB already has 17 frames, as shown in Figure 4.

Since the domains covered by the resource don't seem to vary significantly across cultures (GAMONAL, 2013), Copa 2014 is adopting the same set of frames for the three languages. The translation equivalents for these frames and lexical units, as well as exemple sentences, are all stored in one single database.

2.2. Multilayer Lexicographic Annotation

Annotation procedures adopted by FN-Br follow the guidelines provided by FrameNet's Book (RUPPENHOFER ET AL., 2010) and are performed with the aid of the FrameNet Brasil Desktop, an adaptation of FrameNet's annotation and report generation tool. The FN-Br Desktop allows the creation of frames, FEs and LUs, as well as the creation of subcorpora bound to each LU, which contains the sentences for annotation.

After sentences are imported into the Desktop, they can be annotated for their semantic and syntactic valences. This annotation is performed in a set of layers, which include:

- at least one for the Frame Elements;
- one for Grammatical Functions (GFs);
- one for Phrase Types (PTs);
- one for Other information that is relevant to the annotation but is not covered by the previous layers (information on relative pronouns and their antecedents, for example, is annotated in this layer);
- one for specific properties of the part of speech being analyzed (this layer being named after the part of speech it refers to);
- one for storing extra information on the Sentence, such as whether it is metaphorical or not, for example.

Figure 5 shows an annotation screen in the FN-Br Desktop. In this case, a sentence with the LU *ter* 'have', evoking the *Posse / Possession* frame is annotated for the first four layers, which are the most used in FN-Br's lexicographic annotation.

The sentence depicted in Figure 5 could be translated into English as "*I mean, the person who doesn't have any money is under a lot of stress*". The target LU being annotated is *ter* 'have', which appears inflected in its third person singular present form. This LU evokes, in this sentence, the *Posse / Possession* frame, whose core FEs are the *Posse / Possession* and the *Possuidor / Possessor*. In the first layer – FE –, both the relative pronoun *que* 'that' and its antecedent, the NP *o individuo* 'the person', are marked with the FE *Possuidor / Possessor*. In turn, the NP *dinheiro* 'money' is marked as the *Posse / Possession*. In the GF layer, *o individuo* and *que* are marked as an External Argument, while *dinheiro* is a Direct Object. As for the PT layer, all of the three constituents are marked as Noun Phrases. Finally, in the Other layer, we see that the labels Antecedent and Relative were applied.

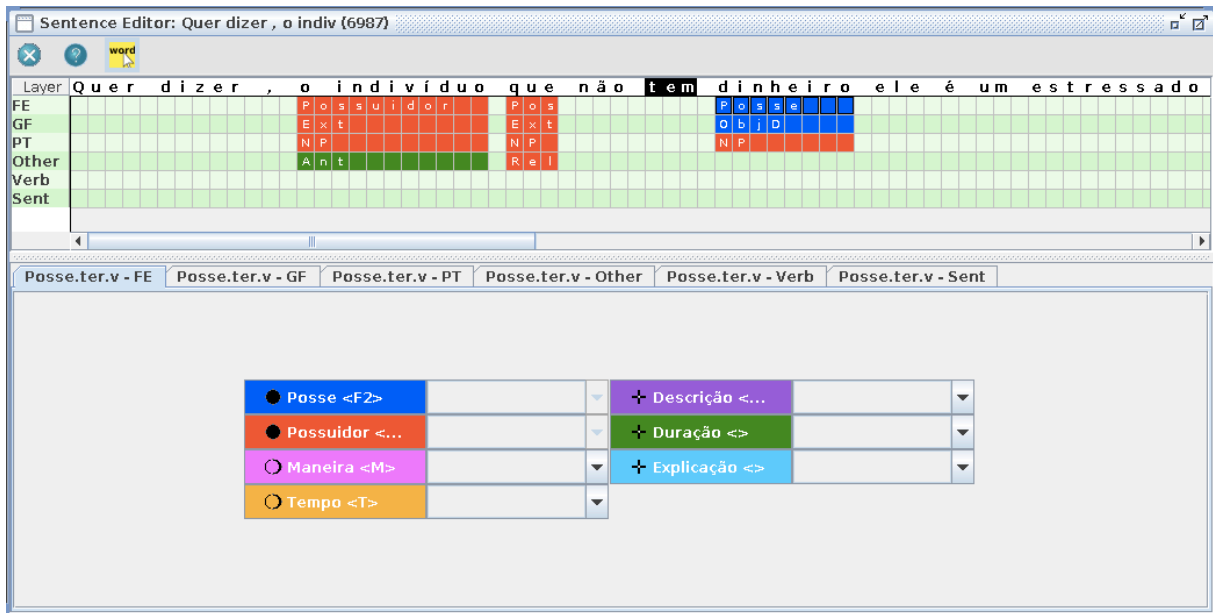


Figure 5: Sentence editor screen in FN-Br's Desktop

The sentence depicted in Figure 5 could be translated into English as “*I mean, the person who doesn't have any money is under stress*”. The target LU being annotated is *ter* ‘have’, which appears inflected in its third person singular present form. This LU evokes, in this sentence, the *Posse* / *Possession* frame, whose core FEs are the *Posse* / *Possession* and the *Possuidor* / *Possessor*. In the first layer – FE –, both the relative pronoun *que* ‘that’ and its antecedent, the NP *o indivíduo* ‘the person’, are marked with the FE *Possuidor* / *Possessor*. In turn, the NP *dinheiro* ‘money’ is marked as the *Posse* / *Possession*. In the GF layer, *o indivíduo* and *que* are marked as an External Argument, while *dinheiro* is a Direct Object. As for the PT layer, all of the three constituents are marked as Noun Phrases. Finally, in the Other layer, we see that the labels Antecedent and Relative were applied.

This example annotation may raise a list of questions, such as:

- Why the rest of the sentence is not annotated?
- Why do the Antecedent and the Relative receive the exact same labels for the first three layers?
- Why is it important to annotate both semantic and syntactic information in FrameNet?

Those questions will be answered as we discuss the theory behind FN's annotation labels in section 3.

3. Behind the Labels: Theoretical Background Supporting FN's Annotation Labels

FrameNet has been designed with the goal of requiring a minimum of theoretical overhead from its users. However, FrameNet views the primary function of annotation as exemplifying how lexical semantics interfaces with syntax; our annotation should help human users and automated machine learning algorithms to map from a phrase to meaning and from a meaning to a phrase. No annotated example can act as a template for building new phrases unless the example has been annotated with categories that generalize beyond the example, and, moreover, with categories that clearly indicate what phenomena they generalize over. Such a description comes perilously close to theory, no matter how carefully phrased.

Nevertheless, we view our policy of “atheoretical” description as keeping us focused concretely on precise matters of how our annotation should be interpreted and generalized by ideal users.

Maintaining clarity about how to annotate in such a way as to support generalization from our examples has required a host of decisions both in our annotation policy and in how data is presented to users. The most basic of these decisions is to separately indicate the contribution of each target. Whether more than one LU is annotated in a sentence or not, each LU is annotated as a completely separate set of layers so that the annotation can be interpreted and generalized to other sentences in which the two LUs did not co-occur. Annotations in which the target LU and its frame explain virtually everything occurring in a sentence are grouped with those in which the target LU and frame explain only a tiny portion of a sentence, to be compared and contrasted for how they map between meaning and form. This simplification is what makes the FrameNet enterprise possible at all. The complications arise from specifying at a gross level how an example is to be generalized, mostly falling under (i) the need to specify how the annotation interacts with syntax (see 3.1) and (ii) the need to indicate both the semantic tendencies of FE fillers and the target-specific syntactic relations of FE fillers.

Inclusion of both semantic and syntactic information about fillers is vital, as it is necessary for human learners, lexical semanticists, and for automatic semantic role labeling. Given the goal of mapping between semantic and syntactic information, the utility of syntactic annotation is obvious. In many cases, syntactic relations unambiguously indicate frame elements, so that in an English sentence representing the Ingestion frame like *plant eats man*, we know that the *man* is the Ingestible, despite the fact that one might have thought that *plant* is semantically a much better fit, simply because the syntactic role of Object cannot map to any other role. In other cases, however, the syntactic analysis is semantically ambiguous and the semantic classes of fillers are the only factor that allows role assignment. In sentences like *I tricked \$50 out of her*, (Manipulate_into_doing) the direct object is analyzed as the Goods, and the prepositional phrase headed by *out of* is analyzed as the Victim, whereas in *I tricked her out of \$50*, the exact opposite assignments are made. Despite the syntactic ambiguity, the assignments remain obvious because money is clearly the kind of thing that is prototypically interpretable as Goods and is not interpretable as Victim, while a person clearly is interpretable as Victim.

Beyond cases where the semantic information of fillers is vital even for the most basic interpretation of a sentence that uses an LU, it is clear that there is a great deal of semantic information in annotation examples that is potentially useful, but not indicated in our structured data. For example, the Ingestion frame includes LUs pertaining both to eating and to drinking and to various kinds of eating and drinking. For second language learners wondering what a verb like *devour* means, they can learn a lot from looking at the annotation of *devour.v* and finding example Ingestibles like *the bodies of dead animals, huge meals, only unbaptised children*, and *all that is mortal of the body* to learn that *devour.v* does not describe a desirable or normal kind of eating, but rather a terrible and inhuman kind.

Because inclusion of both syntactic and semantic information is so vital, our annotation policy emphasizes finding syntactically simple examples and finding semantically informative examples, two constraints that often conflict. The desire to include semantically informative examples is one of the primary motivations for having policies to handle sentences that involve syntactic complications like relative clauses, fronting, passivization, middles, etc. Our annotation aims to indicate only how to map semantics based on the target LU onto phrases headed by the target LU, so all of these constructions, for this part of the project, are merely distractions. The primary goal in our annotation policy is to include as

many informative examples as possible while simultaneously indicating which aspects of the annotation should be considered part of the LU-specific syntax-semantics mapping (the Valence, see 3.2), and which aspects are not to be generalized beyond the current example.

3.1. Locality

As mentioned above, because we are focused on the mapping between syntax and semantics, FrameNet annotates only those parts of a sentence that are syntactically connected to a target LU. The narrowest type of syntactic connection is a **local** syntactic relation: sub-phrases that are part of the phrase headed by the target, as in the direct object in (1). For these phrases, the annotated examples give maximal information on how to build a phrase headed by the target LU, including what type of syntactic sub-phrase can fill an FE, what order it occurs in with respect to other sub-phrases and the target, and how it can interact with other constructions. In principle, these are distinct from other phrases that are connectable to the target LU, for example a verb target LU's subject (in FrameNet called Ext), which is consistently identifiable with a particular role just as local arguments are, even though it is not part of the verb phrase headed by the verb target, may be some considerable distance away from the target due to control relations, and may in fact appear with unusual phrase type due to control relations, as demonstrated in (2). This also includes arguments whose position is entirely due to a construction other than the target LU, as in the questioned direct object in (3). FrameNet policy is to annotate these and, in fact, any phrase in any position that could naturally signal the same semantic relation to the target, regardless of what material occurs in the phrase; this policy is intended to be broad enough to cover every kind of indirect syntactic relation, from the connection of the subject to a verb phrase, to control relations, to heavy NP shift, to gapping, without enumerating all such connecting constructions in advance, a task that is virtually impossible within a language and absolutely impossible in a cross-linguistic setting, as is made all too clear by edge-cases like (4), in which it is perfectly obvious that the possessor of a noun modified by the adjective *favorite* must be supplied as the subject of the infinitive *prepare*, and thus be annotated as filling the `COOK` frame element; it would not be possible to know all such idiosyncratic constructions in advance.

- (1) She was trying to **put** *the cap* back on.
- (2) *Their* attempt to stop the oil by **putting** the cap back on was a simple failure.
- (3) *Whom* do you think he tried to **contact**?
- (4) *His* favorite soup to **prepare** was French onion.

All of the phrases that are syntactically connected to the target are available for annotation. It is important to note that relations that are mediated by anaphoric pronouns and inference, as in (5), we do not annotate. The pronoun *they* refers back to the actions *his attempts at stalling*, and the pronoun itself forms the subject and means action of the purpose infinitive construction which, further, supplies the subject to the verb *anger*; by inference, the agent of the means action of *anger* is the agent of *anger*, so it seems attractive to annotate *his* as the Agent of *anger*.

- (5) *His* attempts at stalling she opposed, whether they were just to **anger** her or not.

The pattern in (5) is different from those in (1-4) in that, depending as it does on inference and anaphora, it is robust to complete syntactic rearrangement, i.e., the *they* could have referred to actions that were mentioned over the course of paragraphs, each with different subjects mentioned or even themselves only resolvable by inference, and the conclusion that the Agent(s) of whatever events are covered by *they* are those who intend to

cause anger. In a nutshell, the pattern in (5) is not about mappings between syntax and semantics, but rather it is about combining the low-level syntax and semantics mappings with semantics to semantics operations that are not restricted by syntax. If we included such examples in our data, we would be implying a syntax-semantics connection that is not there.

For those syntactically connected elements of the sentence, to the extent possible, they are annotated just as if they were local; this policy is based on our goal of representing the syntactic specifications associated specifically with the Target, rather than the syntactic facts that are due to other complications occurring in the sentence. Thus *whom* in (3) is annotated with the grammatical function Object, just as in (1). This policy lets us generalize beyond the peculiar appearance of arguments in examples like (3), since their configuration is predictable from the combination of the process of question formation with simple sentences like (6).

(6) She tried to **contact** *him*.

Labeling with the grammatical function Object is a shorthand for a set of possible realizations and alternations in English. Regardless of where an Object actually appears in an annotated example, it indicates, among other things, that an argument may appear locally immediately after the Target (*curing a wound*), as part of a hyphenated *-ing* compound (*wound-curing*), or non-locally due to interaction with other constructions, and, when the Target is passivized, may appear as the External argument.

Because we do not distinguish between local arguments and non-local ones, FrameNet's policy is easily misinterpreted, and has led to problems especially in the case where a target occurs inside a relative clause, as seen in Figure 5 and in (7-8):

(7) *The woman (that)* she had wanted to **contact** had (apparently) passed away.

(8) *The woman (that)* she had wanted to **talk to** had ...

In (7) and (8), we can see that *woman* is a simple, useful example of the semantic restrictions of an Addressee. However, in (7) *the woman* and the relative-clause marker *that* do not form a constituent, and, worse, the semantically uninformative *that* is the clause-internal argument that should be annotated as the argument of **contact**. The problems only proliferate when we look at examples involving prepositions: in (8), the argument of **talk** appears in three pieces, [*the woman*], [*that*], and [*to*], of which the last is the piece that shows the syntactic requirement for an Addressee to appear as a prepositional phrase with the preposition *to*, and the first is the one that gives useful semantic information.

Since we want to represent both types of information but do not want to mislead learners (either humans or computers) about the syntactic and semantic information, we have an elaborated annotation in this case that indicates how the relative clause is put together. In keeping with representing only the Target's contribution to the syntax, we note that all the separate pieces would occur as a single constituent in a sentence without a relative clause, so we must label all the pieces of the frame element with the phrase type and grammatical function of the single constituent that would occur without the relative clause – NP/Obj for (7), PP/Dep for (8). We then also explicitly mark the relativizing word (*that* in both sentences) with the label Rel and mark the entire modified NP with the label Ant. From this annotation, learners can see that the Ant should be understood as providing useful semantic information, while the other piece or pieces of the FE reveal any useful syntactic information, such as the preposition *to* in (8).

3.2. Valence Patterns

The result of labeling FEs of a Target only with the syntactic relations and syntactic structures called for by the Target specifically is that the annotation represents patterns of mapping from FEs to syntax and vice-versa. However, as seen in examples like (9-10), we cannot assign a mapping from a particular semantic role (*Recipient*) consistently to the same syntactic role – *Obj* in (9), *PP(to)* in (10). Similarly, we cannot consistently map from a syntactic role (*Obj*) to a semantic role – *Recipient* in (9), *Theme* in (10). We can, however, note that the syntactic sequence **give** + NP/*Obj* + NP/*Dep* is associated with the first NP being the *Recipient* and the second NP being the *Theme*, as can be seen from the uninterpretability of (11).

- (9) She **gave** *me* a book.
- (10) He **gave** a book *to me*.
- (11) * She **gave** a book *me*.

The conclusion is that, in many cases, we need to note the entire pattern of mappings of all of the semantic and syntactic elements. These sets of FEs + GFs + PTs we call *valences*. Valences represent the unique patterns of mapping between the totality of syntactic elements (phrase types, grammatical functions, relative order) and the list of frame elements. Because we wish to remain as atheoretical as possible, however, these valences are not equivalent to the minimal lexical valences of Construction Grammar (for minimal lexical entries and valences, see KAY & FILLMORE, 1999, p. 14, e.g.). Despite generalizing over clear, cross-cutting, language-wide phenomena such as relative clause formation, our valences do not generalize over alternations that are potentially lexically idiosyncratic, including passive sentences or any other complex valence alternation that interacts closely with the Target's semantics. For instance, in (12), note that we record *Pat* as the subject, since **rumor** can only appear in the passive, while the conative *at her face* in (13) is labeled Patient/*PP(at)*/*Dep*. This approach allows annotators to annotate without resolving deep mysteries such as the conative construction before proceeding and does not impose any (potentially incomprehensible) analysis on either human learners or automated machine learning applications.

- (12) *Pat* is **rumored** to have a secret.
- (13) The cat **scratched** *at her face*.

4. Annotation Labels in FrameNet Brasil

In spite of the fact that FN-Br follows all the annotation policies discussed in section 3, given the differences between English and Brazilian Portuguese syntax and morphology, it was necessary to re-evaluate and, in many cases, propose changes to set of GF and PT labels used for annotation. This section presents the sets of labels used in FN-Br for assigning GFs and PTs to the linguistic material instantiating the FEs, and also provides discussion of why the sets of labels were defined the way they are presented here.

4.1. Grammatical Functions

As has already been stated in 3.1, FrameNet, and FN-Br as well, annotates syntactic information related to the target LU. Hence, the set of GF labels to be used in annotation is directly related to the part-of-speech of the LU being annotated. Berkeley FrameNet uses the set of grammatical functions presented in Chart 1 (RUPPENHOFFER ET AL., 2010, p.64-71).

External Arguments (*Ext*) are used in FrameNet not only for subjects, but also for expressions controlling subjects, and dependents of a noun governing an infinitival verb.

Arguments of a preposition that are outside of the PP are also marked as Ext under similar circumstances. In turn, any kind of object subcategorized by a verb or preposition receives the Obj label (see 3.1 for detailed discussion on the kinds of structures FN annotates as Obj). The Dep label is applied to all other kinds of complements, such as finite or infinitival clauses, adverbs, prepositional phrases and so on. It may be considered the default GF label in FrameNet. Head can refer both to nouns modified by pre-nominal target adjectives and to an event or relation modified by a target adverb or prepositional phrase. The Gen label is applied to the possessive determiners of a target noun, while Quant refers to pre-nominal definite or indefinite quantifiers. Finally, Appos marks post-target appositional nouns or NPs.

GF Labels	Part-of-speech of the target LU				
	Verb	Noun	Adjective	Preposition	Adverb
External argument (Ext)	X	X	X	X	
Object (Obj)	X			X	
Dependent (Dep)	X	X	X	X	X
Head noun/verb (Head)			X	X	X
Genitive determiner (Gen)		X			
Quantificational det. (Quant)		X			
Appositive (Appos)		X			

Chart 1: Summary of GFs used in Berkeley FrameNet

For FrameNet Brasil, one new GF was created, while one was renamed so as to become more intuitive. In the following paragraphs, we present all GFs used in FN-Br, providing examples.

The External argument (Ext) label in FN-Br is equivalent to the Ext label used in the original project. As can be seen in (14-20), when the target LU is a verb, the Ext label is applied to, respectively, subjects of finite target verbs (14), raised subjects (15), expressions controlling subjects of infinitival verbs (16-18), and arguments of a noun governing an infinitival verb (19-20).¹

- (14) [Maria_{Ext}] **comprou** frutas.
 Maria buy.PAST.3SG fruit
Maria bought fruit
- (15) [Maria_{Ext}] podia **comprar** frutas.
 Maria can.PAST.3SG buy.INF fruit
Maria could buy fruit.
- (16) [Maria_{Ext}] queria **comprar** frutas.
 Maria want.PAST.3SG buy.INF fruit
Maria wanted to buy fruit.
- (17) Antônio viu [Maria_{Ext}] **comprar** frutas.
 Antônio see.PAST.3SG Maria buy.INF fruit
Antônio saw Maria buying fruit.
- (18) Antônio pediu [pra Maria_{Ext}] **comprar** frutas.
 Antônio ask.PAST.3SG to Maria buy.INF fruit
Antônio asked Maria to buy fruit.

¹ For the sake of clarity, we chose to fabricate examples for the GF labels. Examples are similar to each other on purpose, so that it can be easier for the non-Portuguese-speaking reader to compare the different instantiations of grammatical functions.

- (19) Eu admiro a disposição [de Maria_{Ext}]²
 I admire.PRES.1SG the willingness of Maria
 para **ajudar** os amigos.
 to help.INF the friends.
I admire Maria's willingness to help her friends.
- (20) Eu admiro a [sua_{Ext}] disposição
 I admire.PRES.1SG the your willingness
 de **ajudar** os amigos.
 of help.INF the friends.
I admire your willingness to help your friends.

Examples (14-20) almost parallel those listed by Ruppenhofer et al. (2010, p.65), This is due to the fact that both Brazilian Portuguese and English allow almost the same kinds of external argument configurations. One key difference, however, relies on the fact that, while English does not allow null subjects, at least not in its Standard norm, in Brazilian Portuguese, as in other romance languages, they are very common. Spanish FrameNet has a specific kind of label for null subjects: the External Constructional Null Instantiation – ECNI (SUBIRATS, 2009, p.140), however, such a label is not applied in the GF layer, but it marks FEs that are not overtly expressed. Anyhow, FN-Br decided to label null subjects as constructionally motivated null instantiations (CNI) together with other CNIs such as Imperatives and Instructional Infinitives³.

Following the annotation policies devised for English, FN-Br also applies the Ext label when annotating noun, adjective and prepositional targets. Ext labels are applied to FEs evoked by noun targets in support constructions, such as (21). In the case of adjective targets, FEs that receive the Ext label either are the subjects of a copula (22) or the objects in a verbal-nominal predicate (23). As for preposition target externals, they comprise those elements that are related to the preposition in spite of being outside of the Prepositional Phrase (24). Again, the use of the Ext label in FN-Br is similar to the one proposed for English (RUPPENHOFER ET AL., 2010, p. 66-71).

- (21) [Maria_{Ext}] deu **um soco** no Antônio.
 Maria give.PAST.3SG a punch in the Antônio
Maria punched Antônio.
- (22) [Maria_{Ext}] é **alta.**
 Maria be.PRES.3SG tall
Maria is tall.
- (23) Antônio acha [Maria_{Ext}] **bonita.**
 Antônio find.PRES.3SG Maria beautiful
Antônio thinks Maria is beautiful.
- (24) Antônio viu [o livro_{Ext}] **sobre** a mesa.
 Antônio see.PAST.3SG the book on the table

² This constituent is bracketed [de Maria] on the same principle that has us bracket, in English, “said [to her] to go”. The bracketing in such cases follows a general principle of local of mediation of control in a way that we cannot fully explore here, but we may separately note that this increases the comparability of (19) to (20), in the second of which the possession marking is inherently expressed by *sua* ‘your’.

³ It could be argued, however, that those cases should be annotated as Definite Null Instantiations (DNI), since null subjects in Brazilian Portuguese can be regarded as a mechanism of textual cohesion based on anaphora. Nevertheless, a constructional annotation would allow us to explain the DNI as due to a Subjectless Finite VP construction and would additionally allow us to explain the different subject specification in the Imperative construction entirely with the same mechanisms. Most importantly, using CNI will avoid an implicit claim that such DNIs are in any way lexically specific.

Antônio saw the book on the table.

Differently from what holds for the External, the Object label is not applied similarly for English and for Brazilian Portuguese. The Object label used for English is restricted to NP and wh-extracted objects, all other kinds of internal arguments being Dependents. According to Ruppenhofer et al. (2010, p. 65), the Obj label is only assigned to those NPs that are not headed by a preposition and that can serve as a subject in a passive construction. For Brazilian Portuguese, some changes were necessary so as to more adequately capture the syntactic properties of internal arguments of verb targets. First, two labels were created: the Direct Object (DObj) and the Indirect Object (IObj). The DObj label is used for all cases in which the Obj label is applied for English (25-27), plus those cases of either finite or infinitival clausal objects (28-29).

- (25) Maria **comprou** [frutas_{DObj}].
 Maria buy.PAST.3SG fruit
Maria bought fruit
- (26) [O que_{DObj}] Maria **comprou?**
 What Maria buy.PAST.3SG
What did Maria buy?
- (27) Maria **viu** [nos_{DObj}] comprar frutas.
 Maria see.PAST.3SG us buy.INF fruit
Maria saw us buying fruit.
- (28) Maria **sabe** [comprar frutas_{DObj}].
 Maria know.PRES.3SG buy.INF fruit
Maria knows how to buy fruit.
- (29) Maria **sabe** [que você adora frutas_{DObj}].
 Maria know.PAST.3SG que you adore.PRES.2SG fruit
Maria knows you love fruit.

The reasons for including clausal complements among the pieces of language labeled as DObj are three: (i) they can go through wh-extraction in the same way as NPs can, (ii) they can serve as subjects of passive sentences, and (iii) differently from what holds for English, both NP objects and clausal objects may be separated from the verbs subcategorizing them by intervening linguistic material. For instance, while, in English, a sentence like (30) would sound at least weird, if not ungrammatical, its translation into Brazilian Portuguese in (31) would cause no alarm to speakers.

(30) * Ronaldo passed slowly the ball.

(31) Ronaldo passou lentamente a bola.

The other difference between FrameNet and FrameNet Brasil in regards to the use of the Obj label is related to preposition targets. Annotations carried out for English may present the use of Obj for NPs headed by a preposition. Unlike in English, the NPs depending from a prepositional head in Brazilian Portuguese do not share significant properties with the NPs depending from a verbal head (passivizability, cohesion with head); therefore, FN-Br has chosen to use the default label Dep for complements of prepositions.

Indirect Objects, in turn, cannot be used as subjects of passive sentences, since Brazilian Portuguese only allows direct transitive verbs in passive constructions. According to Castilho (2010), Brazilian Portuguese has two kinds of prepositional complements following verbs, indirect objects and obliques, the differences between those two being the following:

- indirect objects can be replaced by dative pronouns (*me, te, lhe*), while obliques cannot;
- indirect objects are headed exclusively by prepositions *a* and *para*, while obliques can occur with other prepositions;
- indirect objects are usually the second object of a verb, while obliques tend to be the only internal argument of a verb.

The IObj label in FN-Br is applied exclusively to indirect objects, such as the ones exemplified in (32-34), while oblique complements are merged into the Dependent category.

- (32) Maria **deu** um livro [pra mim_{IObj}].
 Maria give.PAST.3SG a book to me.DAT
Maria gave a book to me.
- (33) Maria [me_{IObj}] **deu** um livro
 Maria me.DAT give.PAST.3SG a book
Maria gave me a book.
- (34) Maria **deu** um livro [pra mim_{IObj}] ler.
 Maria give.PAST.3SG a book to me.DAT read.INF
Maria gave a book for me to read it.

The Dependent label, hence, is applied to all other verbal arguments not covered by the Ext, DObj and IObj labels, including obliques (35). It is also used for phrases traditionally classified as verbal adjuncts (36-38). In regards to the other kinds of targets (nouns, adjectives and prepositions), Dep is the label ascribed to the right complements of such targets (39-41).

- (35) Maria **gosta** [de frutas_{Dep}].
 Maria like.PRES.3SG of fruit
Maria likes fruit
- (36) Maria **saiu** [de casa_{Dep}].
 Maria leave.PAST.3SG of house
Maria left the house.
- (37) Maria **comprou** frutas [ontem_{Dep}].
 Maria buy.PAST.3SG fruit yesterday
Maria bought fruit yesterday.
- (38) Maria **veio** [aqui_{Dep}].
 Maria come.PAST.ESG here
Maria came here.
- (39) Maria tem **medo** [do escuro_{Dep}].
 Maria have.PRES.3SG fear of the dark
Maria is afraid of dark places.
- (40) Maria é responsável [pelas compras_{Dep}].
 Maria be.PRES.1SG the responsible for the purchases
Maria is in charge of purchases.
- (41) Maria saiu **para** [comprar frutas_{Dep}].
 Maria leave.PAST.1SG for buy. INF fruit
Maria left to buy fruit.

The Head label, translated into Portuguese as Núcleo, is primarily used for identifying head nouns modified by qualitative adjective targets (42). Relational adjectives (e.g. *algebraic*) are not usually annotated in FrameNet, for they are not usually the frame-evoking element in a noun-adjective expression (RUPPENHOFER ET AL., p.68).

- (42) Maria comprou [frutas_{Head}] **frescas.**
 Maria buy.PAST.1SG fruit fresh
Maria bought fresh fruit.

Gen and Quant labels were also imported to FN-Br's annotation, the only difference being the fact that we renamed Gen as PossDet (from Possessive Determiner) so as to make the label more intuitive for annotators. The linguistic chunks they identify, however, are quite similar to those they mark in English. PossDet is applied to either pre-nominal or post-nominal possessive determiners (43-44), although post-nominal possessive determiners are pretty rare in Brazilian Portuguese, while Quant is applied to either definite or indefinite quantifying expressions accompanying a noun (45-46).

- (43) [Minhas_{PossDet}] **compras** foram entregues.
 my purchases be.PAST.3PL deliver.PART
My purchases were delivered.
- (44) Vejo a felicidade nos **olhos** [seus_{PossDet}].
 see.PRES.1SG the happiness in the eyes your
I see hapiness in your eyes.
- (45) Maria comprou [três_{Quant}] **litros** de leite.
 Maria buy.PAST.3SG three liters of milk
Maria bought three liters of milk.
- (46) Maria gastou [muitos_{Quant}] **milhares** de dólares.
 Maria spend.PAST.3SG many thousands of dollars
Maria spent many thousands of dollars.

As for the Appositive label (Appos), although it seems quite intuitive, one can be tempted to annotate it in regards to verbal targets, in sentences like (47). However, since FrameNet annotation is concerned with describing the valence of target LUs, respecting the constituency of the terms instantiating the FEs, Appos is only used when annotating noun targets, as in (48).

- (47) *Maria, [minha professora de francês_{Appos}], **viajou.**
 Maria my teacher of French travel.PAST.3SG
Maria, my French teacher, has traveled.
- (48) O **professor** [Fillmore_{Appos}] iniciou a FrameNet.
 the professor Fillmore initiate.PAST.3SG theFrameNet
Professor Fillmore initiated FrameNet.

The reason why (47) is a bad annotation while (48) is a good one is the fact that *viajou* 'travel' evokes the Travel frame, in which a Traveler moves from a Source location to a Goal through a Path, and both "Maria" and "minha professora de francês" identify the same FE Traveler. Hence, instead of labeling "Maria" as Traveler/Ext/NP and "minha professora de francês" as Traveler/Appos/NP, we choose to label the whole chunk "Maria, minha professora de francês" as Traveler/Ext/NP. Such a choice is reinforced by the fact that, in relation do verb targets, Appositives always inherit the grammatical function of their antecedents. In (48), on the other hand, we find a completely different scenario. Here, the LU *professor* evokes the People_by_vocation frame, in which a Person, marked by the Appos "Fillmore", is viewed in terms of his vocation, being a professor.

The GF labels Dep and Head, when used for adverbial targets, identify linguistic chunks whose syntactic properties are somehow different from those presented by the chunks

these labels identify when annotating for other part-of-speech targets: Head is used to mark the events or relations modified by the adverbs – this use being extended to preposition targets, when the preposition being annotated for heads a PP with adverbial value –, while Dep is applied to any linguistic material modifying the adverb (usually another adverb). See (49) for examples.

- (49) [Maria explicou tudo_{Head}] [bem_{Dep}] **lentamente.**
 Maria explain.PAST.3SG everything well slowly
 Maria explained everything quite slowly.

Given the examples annotated in this section, we can summarize the GF labels used in FN-Br as presented in Chart 2.

GF Labels	Part-of-speech of the target LU				
	Verb	Noun	Adjective	Preposition	Adverb
External argument (Ext)	X	X	X	X	
Direct Object (DObj)	X				
Indirect Object (IObj)	X				
Dependent (Dep)	X	X	X	X	X
Head noun/verb (Head)			X		X
Possessive det. (PossDet)		X			
Quantificational det. (Quant)		X			
Appositive (Appos)		X			

Chart 2: Summary of GFs used in FrameNet Brasil

Aside from ascribing GF labels to the linguistic material instantiating FEs, FN-Br annotation also captures the morphological properties of such a material in terms of Phrase Types. In 4.2. we present the set of PT labels developed for FN-Br.

4.2. Phrase Types

Unlike GFs, Phrase Types (PTs) tend to cause less controversy in FrameNet-like annotations, since the labels and the linguistic material they identify are not different from those found in other analytical perspectives; in this section we present the criteria used for creating the set of PT labels FN-Br uses.

Because the primary goal of PT labels in FrameNet is to categorize the selectional properties of a valence-slot, the PT labels are designed to refer to patterns of external or distributional syntax. These interact with internal syntax in interesting ways. In Chart 2, we show the phrase types which are relevant for Brazilian Portuguese, arranged by several cross-cutting features of internal syntax: the morphology of the syntactic head of the phrase (noun, infinitive verb), the presence or absence of complements (head only vs. the rest), the presence or absence of a marker (e.g. a particular preposition that serves merely to identify its NP complement as a particular argument of a verb), and by the presence or absence of an external argument (distinguishing, in general, phrases from sentences). In addition, the cross-cutting categories combine with indirect question constructions and relativizing constructions, which, despite the fact that both may appear as finite sentences or infinitival verb phrase, are for convenience listed as if they had a different head morphology than other categories. In particular cases, the marker is sufficiently unusual in how it interacts with the grammar that a specific phrase type is devoted to such marked phrases (e.g. *se* is a marked finite sentence, possessive is a marked NP). Although the morphology of Brazilian Portuguese is not the same

as English, just as in English, these features are those that are selected for particular phrases in valence slots.

Despite variable internal syntax, there are several phrases that can be given the same label because of identical external syntax. Just as in English, there are no valences that differentially select the phrases composed of an Ext with an AP, from the phrases composed of an Ext with a gerund or a participle, nor are these differently selected depending on whether they have an overt marker – thus the label S_{abs} for all of them.

Head Morphology	External (+) Marker (-)	External (+) Marker (+)	External (-) Marker (-)	External (-) Marker (+)
Noun (N)	Noun Phrase (NP)	Prepositional Phrase (PP)	Noun Phrase (NP)	Prepositional Phrase (PP)
Adjective (Adj)	Absolutive Sentence (S_{abs})		Adjective Phrase (AdjP)	Prepositional Adjective Phrase (PPAdj)
Adverb (Adv)	---	---	Adverbial Phrase (AdvP)	---
Infinitival Verb (V_{inf})	---	Prepositional Infinitive Sentence (PS_{inf})	Verb Phrase (VP_{inf})	Prepositional Verb Phrase (PVP_{inf})
Gerund (V_{ger})	Absolutive Sentence (S_{abs})		Gerund Phrase (VP_{ger})	Prepositional Gerund Phrase (PP_{ger})
Participle (V_{part})			Participle Phrase (VP_{part})	---
Finite Verb (V_{fin})	Finite Complement Sentence (S_{fin})	Finite Adverbial Sentence (Sub)	---	---
Subjunctive Verb (V_{subj})	Subjunctive Complement Sentence (S_{subj}) Se ('if') Sentence (S_{se})	Subjunctive Adverbial Sentence (Sub) Se ('if') Sentence (S_{se})	---	---
Interrogative (Interrog)	Interrogative Sentence ($S_{interrog}$)	Interrogative Prepositional Phrase ($PP_{interrog}$)	---	Interrogative Prepositional Phrase ($PP_{interrog}$)
Relative (Rel)	Finite Relative Clause (S_{finRel})	Prepositional Infinitival Relative (PInfRel)	---	Prepositional Infinitival Relative (PInfRel)
Number (Num)	Number (Num)	Prepositional Phrase (PP)	Number (Num)	Prepositional Phrase (PP)
Quotation (Quot)	Quotation (Quot)	---	---	---
Possessive (Poss)	Possessive Noun Phrase (PossNP)	---	---	---

Chart 3: Phrase Types used for annotation in FN-Br.

The first column in Chart 3, besides identifying the morphology of the head in each PT label used in FN-Br, also presents a list of labels, since head-only constituents may occur in the annotation, especially when annotating non-lexical constructions, but also in cases such as in (50).

- (50) uma **discussão** [algébrica_{Adj}]
 a discussion algebraic
an algebraic discussion

Overall, we may note that the alterations to the GF and PT labels of FrameNet as used for English have been modest. However, the flexibility of the FrameNet method and software have made it relatively straightforward to add the GF and PT labels for Brazilian Portuguese while retaining the same clear framework.

Conclusion

In this paper, we aimed to discuss the criteria used for developing the GF and PT sets of labels used in FN-Br annotation. After presenting some FrameNet and FrameNet Brasil basics, we defined the concepts of locality and valence in FrameNet, both of which are key for understanding the motivation behind our decisions on whether to annotate or not a given piece of language. Our policies on locality and valence are further motivated in terms of what information is of use to language learners, computational systems, and linguists.

Having defined the ground upon which FrameNet – and FN-Br – annotation is founded, we presented the sets of GF and PT labels used for Brazilian Portuguese, exemplifying the contexts in which each label in the first set is used. We also carried out specific discussions on the kind of linguistic material covered by some labels, such as DObj, IObj, Dep and Appos, for instance. Far from representing meaningless theoretical labels, each of these categories is concretely grounded in what linguistic alternations are implied by their use, how such phrases are put together, and in what embedding contexts these phrases should be expected to occur.

We believe the discussions presented in this paper may contribute to the understanding of the analytical decisions behind the annotations in the FN-Br databases, thus enabling all users of our data to interpret them properly and further enabling them to proceed to powerful and insightful analyses of their own, firmly planted on the foundation of more than a decade of detailed annotation practice.

Acknowledgements

FrameNet Brasil is funded by CNPq (Grants # 401853/2011-0, 474270/2011-4 and 476292/2012-3), FAPEMIG (Grants # APQ-00567-12 and APQ-03954-10), and UFJF. We thank the FrameNet Brasil staff, for reading and commenting on this paper. We also thank Chuck Fillmore, for being such an inspiration to all of us.

References

- BAKER, C. F.; FILLMORE, C. J. & CRONIN, B. The Structure of The FrameNet Database. *International Journal of Lexicography*, v. 16, n. 3, p. 281-296, 2003.
- BERTOLDI, A. *Semântica de Frames e recursos lexicais jurídicos: um estudo contrastivo*. 2011. 136 p. Tese de Doutorado em Linguística Aplicada, Universidade do Vale dos Sinos, São Leopoldo, 2011.

BOAS, H. C. Bilingual FrameNet Dictionaries for Machine Translation. In: RODRIGUEZ, M. G. & ARAUJO, C. P. S. *Proceedings of The Third International Conference on Language Resources and Evaluation*. Las Palmas, Espanha, 2002.

BORIN, L.; DANNÉLLS, D.; FORSBERG, M.; TOPOROWSKA GRONOSTAJ, M. & KOKKINAKIS, D. Swedish FrameNet++. In: *Proceedings of the Swedish Language Technology Conference*. Linköping, 2010.

CASTILHO, A. T. *Nova Gramática do Português Brasileiro*. São Paulo: Contexto, 2010.

ERK, K.; KOWALSKI, A.; PADÓ, S. & PINKAL, M. Building a Resource for Lexical Semantics. In: *Proceedings of the 17th International Conference of Linguists*. Prague, 2003a.

ERK, K.; KOWALSKI, A. & PADÓ, S. The SALSA Annotation Tool. In: *Proceedings of the Workshop on Prospects and Advances in the Syntax/Semantics Interface*. Nancy, 2003b.

FILLMORE, C. J. Frame Semantics. In: *Linguistics in the Morning Calm*. Seul: Hanshin Publishing Co., p.111-137, 1982.

FILLMORE, C. J. Frames and the semantics of understanding. *Quaderni di Semantica*, v. 6, n. 2, p. 222-254, 1985.

FILLMORE, C. J.; JOHNSON, C. R. & PETRUCK, M. R. L. Background to FrameNet. *International Journal of Lexicography*, v. 16, n. 3, p. 235-250, 2003a.

FILLMORE, C. J.; PETRUCK, M. R. L.; RUPPENHOFER, J. & WRIGHT, A. FrameNet in action: the case of attaching. *International Journal of Lexicography*, v. 16, n. 3, p. 297-332, 2003b.

FILLMORE, C. J.; LEE-GOLDMAN, R. & RHOMIEUX, R. The FrameNet Constructicon. In: BOAS, H. C. and SAG, I. A. (Eds.). *Sign-based Construction Grammar*. Stanford: Center for the Study of Language and Information, 2012.

GAMONAL, M. A. *Copa 2014 FrameNet Brasil: diretrizes para a constituição de um dicionário eletrônico trilingue a partir da análise de frames da experiência turística*. 2013. 146p. Dissertação de Mestrado em Linguística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

KAY, P. & FILLMORE, C. J. Grammatical Constructions and Linguistic Generalizations: the What's X Doing Y Construction. *Language*, v. 75, n. 1, p. 1-33, 1999.

LAGE, L. M. *Frames e Construções: a implementação do Constructicon na FrameNet Brasil*. 2013. 113p. Dissertação de Mestrado em Linguística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

OHARA, K. H.; FUJII, S.; OHORI, T.; SUZUKI, R.; SAITO, H. & ISHIZAKI, S. The Japanese FrameNet Project: An Introduction. In: *Proceedings of the LREC 2004 Satellite Workshop: Building Lexical Resources from Semantically Annotated Corpora*. Lisboa, 2004.

RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R. L.; JOHNSON, C. R. & SCHEFFCZYK, J. *FrameNet II: Extended theory and practice*. 2010. Retrieved from: <http://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>. Accessed in May, 15th 2012.

SALOMÃO, M.M.M. FrameNet Brasil: um trabalho em progresso. *Calidoscópico*, v. 7, n. 2, p. 171-182, 2009.

SALOMÃO, M. M. M.; TORRENT, T. T.; CAMPOS, F. C. A.; BRAGA, R. M. M. & VIEIRA, M. B. *Copa 2014 Framenet Brasil*. Projeto apresentado ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), no âmbito do Edital Universal nº 14/2011. Juiz de Fora, 2011.

SATO, H. New Functions of FrameSQL for Multilingual FrameNets. In: *Proceedings of the Sixth International Language Resources and Evaluation Conference*. European Language Resources Association, Marrakech, p. 758-762, 2008.

SUBIRATS, C. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon In: BOAS, H. C. *Multilingual FrameNets in computational lexicography: methods and applications*. Berlin: Mouton De Gruyter, 2009.

SUBIRATS, C. & PETRUCK, M. R. L. Surprise: Spanish FrameNet! In: *Proceedings of the 17th International Conference of Linguists*. Prague, 2003.

VAZ, F. F. *Estudo comparativo entre frames de Comunicação do Inglês e do Português do Brasil*. 2012. 221 p. Tese de Doutorado em Linguística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2012.

YOU, L & LIU, K. Building Chinese FrameNet Database. In: *Proceedings of the 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering*. Shanghai: ICCS Shanghai Normal University, 2005.

Received: 30/04/2013

Revised: 22/07/2013

Accepted: 30/07/2013

Published: 31/10/2013