CHAPTER 4

Towards continuity between the lexicon and the constructicon in FrameNet Brasil

Tiago Timponi Torrent, Ely Edison da Silva Matos, Ludmila Lage, Adrieli Laviola, Tatiane Tavares, Vânia Gomes de Almeida and Natália Sigiliano

This chapter presents the theoretical discussion, analytical procedures, and corresponding computational implementations carried out by FrameNet Brasil to take the principle of the continuity between grammar and the lexicon to the computational domain by deeply integrating two frame-based resources under development for Brazilian Portuguese: a lexicon and a constructicon. To achieve this goal, we start by discussing the continuity principle in the Construction Grammar paradigm, that is, the founding idea that both the lexicon and what is usually referred to in most traditions as grammar can be accounted for in terms of constructions. After, we present the computational solutions that led to the development of a new common database structure for the FrameNet Brasil Lexicon and Construction: FN-Br 2.0. Finally, we present the analyses of two constructions in Brazilian Portuguese: the Dative with Infinitive Construction, which illustrates currently available features of FN-Br 2.0, and the Inceptive Aspect Construction, which imposes a series of new challenges to the model.

Keywords: FrameNet Brasil, constructicon, lexicon, database structure, construction-to-construction relations, construction-to-frame relations

1. Introduction

The development of a Constructicon in FrameNet Brasil followed the path inaugurated in 2008 by the Beyond the Core Project, whose main purpose was to design a resource that, in complementation to the Berkeley FrameNet Lexicon, could account for those phenomena located beyond the semantic and syntactic affordances of lexical units (Fillmore, Lee-Goldman & Rhomieux, 2012, p. 311–314; Petruck & Lee-Goldman, this volume). Hence, building on the analyses of Brazilian Portuguese constructions carried out by various researchers at the Federal

DOI 10.1075/cal.22.04tor

University of Juiz de Fora (see Miranda & Salomão, 2008 for a collection of those analyses), the Brazilian Portuguese Constructicon initiated in 2010, by analyzing the `para_infinitive` Family of Constructions (Torrent, 2015).

At first, we attempted to deploy an adapted version of the same software used by Berkeley FrameNet to create both the Lexicon and the Constructicon: the FrameNetDesktop. We started both resources by not only expanding the Berkeley FrameNet database for frames into Brazilian Portuguese, but also by adapting the labels in the annotation tool to the morphological and syntactic characteristics of our language (Torrent & Ellsworth, 2013). Inasmuch as the work advanced, however, some questions related to basic assumptions of Construction Grammar arose:

1. How to model the fact that constructions may evoke frames?
2. How to model the fact that constructions are related to each other in a network?
3. How to model productive constraints on which kind of lexical material can fill a given slot in a construction?
4. How to model the continuity between grammar and the lexicon?

Those issues led us to rethink some aspects of our analytical approach, which, in turn led to the need of revising the database structure and software tools used to model the analyses. The solutions adopted in this process and the new software apparatus developed to model them – the FN-Br 2.0 – are the main topics of this chapter.

We start by presenting some FrameNet Brasil basics in Section 2. In Section 3, we discuss the implications of the continuity approach to the development of the lexicon and the constructicon, and present the computational effort carried out to cope with those implications: FN-Br 2.0. Section 4 brings two sample analyses of Brazilian Portuguese constructions that explore features of FN-Br 2.0 presented in the previous section, and present new challenges to the same model. The last section brings our conclusions.

2. FrameNet Brasil[1]

FrameNet Brasil started in 2007 as a research project in the Graduate Program in Linguistics at the Federal University of Juiz de Fora. Because of that specific characteristic, the work in the Brazilian branch of framenet has always been deeply connected to the development of M.A. theses and PhD dissertations discussing the analytical solutions adopted by framenet to deal with phenomena that have been the focus of Cognitive Semantics and Construction Grammar for a long time.

1. http://www.framenetbr.ufjf.br

From 2010 on, the former project – now a lab housing several research initiatives – evolved also into different directions, among which, we highlight three:

1. expanding the Berkeley FrameNet lexical database into Brazilian Portuguese;
2. creating a repertoire of Brazilian Portuguese constructions;
3. developing domain specific multilingual applications of framenet to non-specialist users.

Those initiatives led so far to the development of two databases: the FN-Br database, containing a general vocabulary lexicon and a constructicon; and the m.knob database, containing a multilingual – Brazilian Portuguese, English, Spanish and French – lexicon covering the vocabularies of Tourism and Sports (Torrent, Salomão, Campos et al., 2014; Costa & Torrent, 2017). This second database supports the FrameNet Brasil World Cup Dictionary, a web app designed to help tourists during the 2014 FIFA World Cup (Torrent, Salomão, Matos et al., 2014), and the Multilingual Knowledge Base, an app providing travel recommendations and sentence translations for tourists (Paiva & Torrent, 2017).[2]

The next sections present the FN-Br database in more detail, since this is the one in which the constructicon is included.

2.1 The FN-Br lexicon

The FN-Br lexicon is being expanded from the Berkeley FrameNet data release 1.7. The expansion process into Brazilian Portuguese includes:

1. adapting the annotation tool to the specificities of this language;
2. translating the names and descriptions of frames and frame elements;
3. adapting frames and frame elements to Brazilian Portuguese, in cases where it is necessary;
4. populating the database with the lexical units.

The adaptation of the annotation tool to Brazilian Portuguese led to the definition of the labels to be used in the annotation of the grammatical functions (GFs) and phrase types (PTs) of the linguistic material instantiating the frame elements (FEs) that manifest in the local context of target lexical units (LUs) (see Petruck and

2. The m.knob Lexicon comprises 87 trilingual frames, 70 of which did not exist in the Berkeley FrameNet data release 1.7. A total of 5,251 LUs are associated to the frames: 1,669 for Brazilian Portuguese, 2,551 for English, 930 for Spanish and 101 for French. The m.knob database has more than 13,000 annotation sets. The FrameNet Brasil World Cup Dictionary can be accessed at http://www.dicionariodacopa.com.br and the Multilingual Knowledge Base at http://mknob.com.

Lee-Goldman, this volume, for a concise explanation of the lexicographic annotation process in framenet). In this process, documented in Torrent & Ellsworth (2013), the properties of some labels were changed, while other labels were created, such as the Indirect Object label, for example.

Also, the criteria for applying the labels concerning Null Instantiations required adaptations. Although Fillmore (2007, p. 147–148) defines the difference between Definite (DNI) and Indefinite (INI) Null Instantiations primarily in terms of the kind of informational status of the FEs marked with these labels, both in his work and in that by Ruppenhoffer et al. (2016, p. 28–29), there is a clear proposal of treating this difference in terms of properties of the LUs. In other words, the main difference between a DNI and an INI is that, while the first is a zero anaphora to which it is possible to identify a referent, the latter is an existential omission that does not call for the identification of a referent in the context. However, FrameNet treats specially the indefinite type as a valence property of the LU. Hence, some verbs such as *eat* and *bake* would license INIs of the FEs Ingestibles and Heating_instrument, respectively.[3]

For example, with *eat*, INIs would be licensed in sentences such as the one presented in (2) as an answer to the question in (1), meaning that the speaker has already eaten something else before and, therefore, will not try the cake. Note that the thing eaten by the speaker is not the cake being offered, since, if it were, the answer would be the one presented in (3).

(1) Would you like to try some of this delicious cake?

(2) No, thanks, I already ate.

(3) No, thanks, I already ate it.

The same does not hold for Brazilian Portuguese, since the answer in (4) would be suitable for both scenarios.

(4) Não, obrigado, eu já comi.
No thanks I already eat.past.1sg
No, thanks, I already ate./No, thanks, I already ate it.

Therefore, due to the fact that verbs in Brazilian Portuguese – in general – license Direct Object omissions where there is either an anaphoric reference or an existential one, the informational status of the omitted FE is the only criterion taken into consideration, with no further attempt of capturing LU-specific properties

3. Frame Element, as well as Construction Element names are represented in Small caps in this chapter.

concerning NIs: when it is possible to identify a specific referent, the DNI label is applied; when not, an INI label is used.

The last adaptation concerns the Constructional Null Instantiations. In FrameNet, the CNI label is used for those cases in which a grammatical construction licenses the omission of the constituent that would be assigned a FE label. Because English requires verb subjects to be overtly expressed in declarative sentences, and Brazilian Portuguese doesn't, CNIs in FrameNet Brasil include omitted subjects.

Processes 2, 3 and 4 in the expansion are performed simultaneously. As new LUs are added to the FrameNet Brasil database, frames and FEs – both their names and descriptions – are translated and/or adapted into Brazilian Portuguese. Linkage to the original English database is maintained both by the use of the same IDs for each expanded frame and FE in the FN-Br database – so that machines can track which Brazilian Portuguese frame is linked to which English frame –, and by the maintenance of the English names for each frame and FE in the report, next to the translated names – so that non-speakers of Portuguese can understand the general structure of the frames. So far, the FN-Br lexicon comprises 472 frames, 2,896 LUs and 2,386 annotated sentences.

Except for the points discussed above, the annotation process in the FN-Br database broadly follows the same guidelines defined by Ruppenhoffer et al. (2016) for both lexicographic and full-text annotation.

2.2 The FN-Br constructicon

Built in parallel with the FN-Br Lexicon, the FN-Br Constructicon followed, from the beginning, the directions of the Beyond the Core Project, since it also started by computationally representing constructions that had already been studied by Construction Grammarians. An example of such kind of construction is the `dative_with_infinitive` cxn (Torrent, 2015). Sentence (5) provides an instance of a construct licensed by the `dative_with_infinitive` cxn highlighted in bold.

(5) Ela deu **[dinheiro$_{\text{HEAD}}$]** **[para** **mim** **viajar$_{\text{PARA_SINF}}$]**
She give.PAST.3SG money to me.DAT travel.INF
She gave the money for me to travel.

A traditional Berkeley Construction Grammar representation of such a construct is presented in Figure 1.

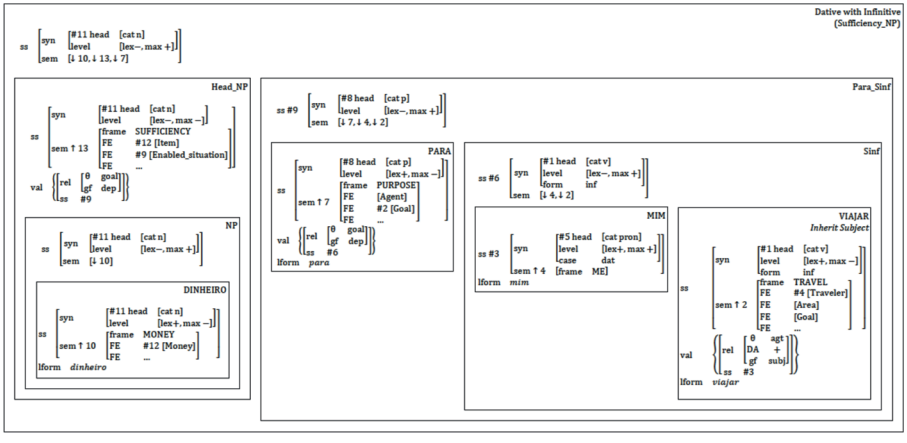

Figure 1. Berkeley-construction-grammar-style diagram for *"dinheiro para mim viajar"*

Each box in Figure 1 indicates one constituent in the construct. Three kinds of information may be provided for each constituent: its syntactic-semantic features (ss), its valence requirements (val) and its lexical form (lform). Except for the lform, the properties of each constituent are specified in terms of an Attribute Value Matrix (AVM). Values may be expressed as binary features (+/−), as items in a closed list (v, dat, subj, Traveler …), or go unspecified (…). Numbers 1 to 13 are unification indices, when preceded by an upward of downward arrow, those numbers indicate that some semantic import of the constituents is projected up to a daughter sign or to the whole construct.

Hence, by reading Figure 1 one may state that:

1. The construction licensing this construct is composed of two daughter signs, a HEAD_NP and an infinitival sentence headed by *para* (PARA_SINF), a preposition generally used in Portuguese for indicating purpose and direction. The HEAD_NP evokes the `Sufficiency` frame, which is composed of two core Frame Elements (FE), the ENABLED_SITUATION and the ITEM enabling it.[4]
2. The PARA_SINF sign is, in turn, composed of two daughters, its head, the preposition PARA (#8) and an infinitival sentence (SINF), which fulfills the valence requirements of PARA, as indicated by the unification index #6. The syntax and semantics of SINF unify with that of VIAJAR (#1, #2), the infinitival head verb in this sentence.

4. In FrameNet, FEs may be assigned three different coreness statuses: (a) core, when the frame does not exist without the FE; (b) peripheral, when the FE adds circumstantial information relevant, but not necessary, to the frame; and (c) extra-thematic, when the circumstantial information is not dependent on the frame at any level, and is usually introduced by a construction that happens to occur in the sentence.

3. In turn, the PARA_SINF sing fulfills the valence requirements of the HEAD_NP (#9). Semantically, it unifies with the ENABLED_SITUATION FE, while the ITEM FE is mapped to the MONEY FE in the `Money` frame evoked by DINHEIRO (#12).

Since Figure 1 brings the representation of a construct – i.e. of a piece of language licensed by a construction (Fillmore et al., 2012) – it does not match entirely with the representation of the `dative_with_infinitive` cxn, which is more generic: virtually any infinitival verb may occupy the head slot of the PARA_SINF sign, as well as any noun that can serve as a resource enabling a situation may be the head of the NP. Also, the subject of the infinitival verb may be left unspecified, because there are apparently no restrictions that apply to this slot. The […] notation indicates that those features are left unspecified in the construction. On the other hand, the preposition *para* must be always present.

Moreover, although the lexical constructions filling the slots are not present in the representation of the `dative_with_infinitive` cxn itself, the unification restrictions proposed in Figure 1 are still valid, since they are dependent on the whole construction, not on each specific lexical possibility on its own. In other words, because the `Sufficiency` frame is evoked by the construction, the FE ITEM will be mapped to whichever Noun heads the NP, while the FE ENABLED_SITUATION will be described in terms of whichever infinitival verb in the PARA_SINF daughter.

The nested boxes diagram of the construction itself is presented in Figure 2.

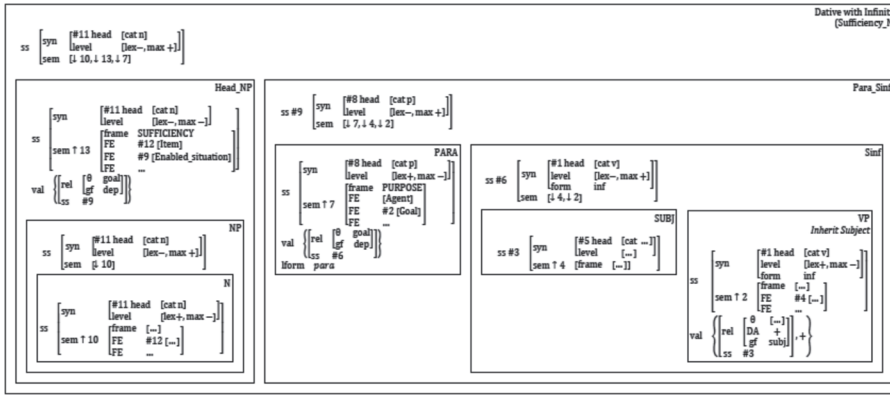

Figure 2. Berkeley-construction-grammar-style diagram for the `dative_with_infinitive` cxn

The challenge of developing a constructicon is that of computationally representing the restrictions and possibilities of constructions such as those depicted in Figure 2, which includes, for instance, how the elements in the construction relate to one another, or how the formal pole relates to the semantic pole. Hence, computational models of constructions must address, from the beginning, both constituency and

unification. In the Berkeley FrameNet Constructicon, constituency is addressed by the creation of Construction Elements (CEs), the daughter signs of the construction. When the construction evokes a frame, the CEs refer to the FEs in the frame evoked by the construction. Unification is handled by annotating instances of the construction in a multilayer fashion. During this annotation process, labels specifying the Grammatical Function (GF) and Phrase Type (PT) of the valents in the construct are aligned to the CEs (Fillmore et al., 2012, p. 321–324).

However, in the FN-Br Constructicon, construction modeling is somehow different to the one carried out in the original Berkeley Constructicon. The most prominent of those distinctions were formalized by Lage's (2013) annotation policies, which aim to provide objective criteria for both construction creation and annotation.

In regards to the creation of constructions and their daughter signs, the FN-Br Constructicon creates CEs based more on formal aspects than on functional/semantic features. Hence, if a construction evokes a frame, instead of creating CEs that refer to the FEs in the evoked frame, in FN-Br we create CEs such as HEAD_NP or PARA_SINF, and then link those CEs to the FEs in the appropriate frame, as it will be shown in Section 4. In other words, instead of addressing the unification of syntactic and semantic features only in annotation and in the prose description of the constructions and CEs, we also formalize it via relations in the database when such unification is part of the general properties of the construction. In this sense, besides being different from the Berkeley Constructicon, the process of construction creation in FN-Br also differs form that of the Swedish Constructicon, in which constructions feature generic semantic roles (such as Agent, Patient and so on).

For instance, in the case of the `dative_with_infinitive` cxn, a construction to frame relation in the database maps the CEs HEAD_NP (NP_NÚCLEO) and PARA_SINF to the FEs ITEM and ENABLED_SITUATION (SITUAÇÃO_HABILITADA) in the `Sufficiency` (`Suficiência`) frame, respectively. When it comes to annotating an instance of the construction, as in (5), such a relation automatically assigns the relevant FE labels as those of the CE are applied to the construct, as can be seen in Figure 3.

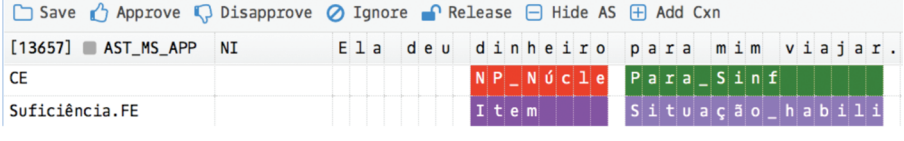

Figure 3. An annotated example of the `dative_with_infinitive` cxn in the FN-Br web tool

Although the annotation depicted in Figure 3 captures the form-meaning unification processes that are constant in the `dative_with_infinitive` cxn, it does not capture all the processes that are relevant to the meaning of the sentence. There are other aspects of such a meaning that are contributed by the lexical constructions in the sentence and their valence affordances. In practical terms, it means the annotation of a sentence such as (5) would have FEs that are evoked by the LUs *dar.v* 'give', *dinheiro.n* 'money', *para.prep* 'for', *viajar.v* 'travel' and some others evoked by the `dative_with_infinitive` cxn. Figure 4 shows how sentence (5) is fully annotated in FN-Br.

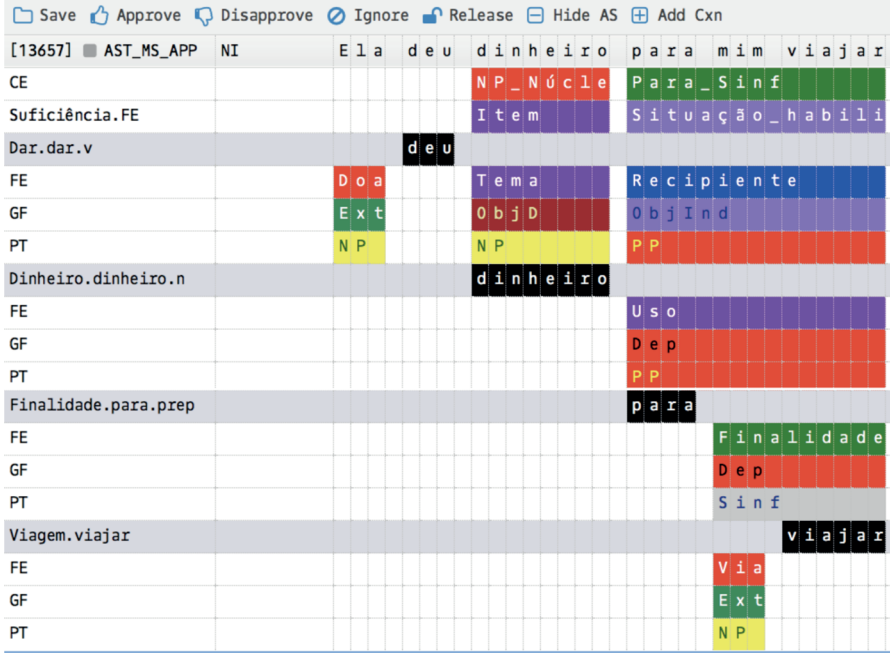

Figure 4. The complete annotation of *Ela deu dinheiro para mim viajar* in the FN-Br web tool

As it may be seen in Figure 4, there are several layers of annotation associated with the sentence:

1. Evoked by the verb *dar* 'give', the core FEs Donor (Doador), Theme (Tema) and Recipient (Recipiente) in the `Giving` (`Dar`) frame are assigned, respectively, to *Ela* 'she', *dinheiro* 'money' and *para mim viajar* 'for me to travel', together with the relevant labels for the GFs and PTs for each valent.

2. In the case of the LU *dinheiro* 'money', evoking the `Money` (`Dinheiro`) frame, the FE Money is incorporated by the LU, meaning its realization is in the LU itself, not in its local syntactic context. Again, a non-core FE, this time the Inherent_purpose (Uso), is instantiated by the infinitival sentence headed by *para*.[5]
3. The `Purpose` (`Finalidade`) frame, evoked by *para* 'for', has as core FE the Goal (Finalidade) represented by the infinitival sentence.
4. Last, the `Travel` (`Viagem`) frame, evoked by *viajar* 'travel', is instantiated via the core FE Traveler (Viajante), linguistically manifested by *mim* 'me'.[6]

Although each LU, as well as the `dative_with_infinitive` cxn, evokes a particular frame and contributes to the meaning of the sentence in a specific way, the frames evoked interact with one another. Moreover, although LUs and constructions may be referred to as different units of language, such a distinction is more due to the historical development of Berkeley FrameNet (as a lexicon to which a constructicon is being added) than to a conceptual difference between those types of units: both kinds of units may be treated similarly when it comes to annotation, and, most importantly, both kinds of units are constructions. Hence, the improvements FrameNet Brasil has been making in its analytical tools, such as the web annotation tool depicted in Figures 3 and 4, aims to allow for deeper integration between the two kinds of annotation (lexical and constructional) and between the databases derived from each of them. In other words, our current effort is to adequately model the continuity between Brazilian Portuguese grammar and lexicon by developing a database system in which Lexicon and Constructicon interact. Let's turn now to this specific issue.

5. We will return to the question about whether the Imposed_purpose and Inherent_purpose FEs should be included in the annotation for the LUs *dar* 'give' and *dinheiro* 'money' by the end of this chapter.

6. All frames in this example are the same for both English and Portuguese, and have been expanded from Berkeley FrameNet into FrameNet Brasil with no adaptations to the structure.

3. Modeling the continuity between grammar and the lexicon

Fillmore, Kay & O'Connor (1988), while introducing the model sustaining their analysis of the *let alone* constructions, state one of the most recognizable pillars of Construction Grammar: the continuity between grammar and the lexicon. Such a statement not only set a clear differentiation between the constructionist approaches and the preceding derivationist paradigm, but also allowed each Construction Grammar to develop one single set of analytical tools to account for lexical items, idioms, argument structures and so on. As the constructionist approach that later came to be known as Berkeley Construction Grammar (BCG) developed, the analyses of constructions, from the fully lexical to the highly schematic, were proposed in terms of complex feature structures that neither derived from assumed deep structures, nor presented empty categories. Unification of such features acts as the main formal operation in this model – as demonstrated in Figures 1 and 2 – and no transformation rules are predicted. Hence, as one of the most famous slogans of Construction Grammar would state it: "What you see is what you get" (Fillmore, 2013, p. 111–113).

Together with feature structures and unification, valence descriptions are also an important part of BCG analyses. According to Fillmore (2013, p. 118–119), in BCG, valences include both obligatory and optional valents, that is, both arguments and adjuncts, in syntactic terms. There is, nevertheless, a distinction between core and peripheral valents. Fillmore adds that the complete valence of a lexical construction includes specification about the semantic roles, grammatical functions and phrase types for each valent.

The resemblance between such a proposal and the valence descriptions found in framenets is neither coincidental, nor due to the fact that Fillmore wrote his 2013 chapter on Berkeley Construction Grammar after Berkeley FrameNet was already a mature research initiative. The idea of matching semantic information to syntactic behavior is present in Fillmore's work at least since 1968 – in *The Case for Case* – and more prominently in his 1977 paper *The Case for Case Reopened*. FrameNet lexicographic annotation is, thus, inspired by the same kind of epistemological background that led to the development of BCG.

Nevertheless, advocating in favor of the continuity between grammar and the lexicon, by assuming that lexical items, as well as morphemes, are constructions themselves does not mean to abandon the differentiation among those, idioms and clause-level constructions, or, as Langacker (2008, p. 6–7) phrases it:

> Overlap among lexicon, morphology, and syntax does not prevent us from defining them and drawing useful distinctions, any more than the absence of a precise boundary between green and blue condemns us to seeing only grue – a gradation does not imply undifferentiated homogeneity.

In a similar direction, Goldberg (1995, p. 7) also points out that by not positing a strict division between grammar and the lexicon, constructionists do not deny the existence of distinctly lexical and syntactic constructions.

Those theoretical claims of Construction Grammar set the ground for the development of a computational resource that, at one time, is capable of bridging the gap between a frame-based lexicon and a constructicon, while still maintaining some methodological boundaries when it comes to deciding whether a piece of language should be accounted for in the lexicon, in the constructicon or in both (see Torrent, Lage, Sampaio, Tavares & Matos, 2014a for a discussion of such a methodology).

Given the theoretical background briefly presented in the paragraphs above, modeling the continuity between grammar and the lexicon presents itself as a feasible task in a framenet, provided that some minor aspects of the database structure be incremented in order to promote a gradual interconnection among lexical units, constructions and their semantic import: the frames. We will, in the next two sections of this chapter, (1) present the requirements for such a model and (2) show how the original FrameNet database structure was changed so as to accommodate those requirements in one possible implementation.

3.1 Requirements for modeling the continuity between grammar and the lexicon

The first step in defining the requirements for a database featuring deeper integration between the constructicon and the lexicon was to analyze the properties of the three main entities in such a database: frames, lexical units and constructions.

Frames are defined by Fillmore (1982, p. 111) 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 this seminal concept was "translated" into the idea of FrameNet (Fillmore & Atkins, 1992; Fillmore, Petruck, Ruppenhofer & Wright, 2003) and interconnected frames became the basis for a lexical resource, they were modeled as having a name, a definition usually followed by a set of examples, a list of participants and props (the FEs) involved in the scene being described, and a set of both internal and external relations. As for the FEs, they also feature a definition, sometimes followed by examples, a semantic type and a set of relations as well. Although the frame reports only show the internal relations in which FEs take part, every frame-to-frame relation is also a FE-to-FE relation. As an example, consider the `Travel` frame in Figure 5.

Travel

Lexical Unit Index

Definition:

In this frame a Traveler goes on a journey, an activity, generally planned in advance, in which the Traveler moves from a Source location to a Goal along a Path or within an Area. The journey can be accompanied by Co-participants and Baggage. The Duration or Distance of the journey, both generally long, may also be described as may be the Mode_of_transportation. Words in this frame emphasize the whole process of getting from one place to another, rather than profiling merely the beginning or the end of the journey.

Ellen JOURNEYED to Europe with five suitcases.

Samantha JOURNEYED 2500 miles with her family by sea to China.

The Osbournes *took* a TRIP from Beverly Hills to London on the Concorde.

FEs:

Core:

Area [Area] **Semantic Type:** Location	This is the Area in which the traveling takes place. This frame element describes the enclosed area inside which travelling, of unspecified Source, Path or Goal takes place. We TRAVELLED in Europe.
Direction [dir] **Excludes:** Area	The direction in which the Traveler goes. They began their ODYSSEY north.
Goal [Goal] **Semantic Type:** Goal **Excludes:** Area	The Goal is the location where the travelers end up.
Mode_of_transportation [MoT]	The Mode_of_transportation expresses how the motion of the Traveler is effected, by their body or by a vehicle which holds and conveys the Traveler. Vehicles can move in any way and in any medium. They are usually expressed obliquely with 'in' or 'by'. Barney used to TRAVEL by bus a lot Strom TRAVELLED on foot to see the Pope
Path [Path] **Semantic Type:** Path **Excludes:** Area	The Path is the route along which the travel takes place.
Source [Src] **Semantic Type:** Source **Excludes:** Area	The Source is the starting point of the trip.
Traveler [Trav] **Semantic Type:** Sentient	This is the living being which travels. Normally, the Traveler is expressed as an external argument.

Non-Core:

Baggage [Bag]	The Baggage are the items necessary for travel that accompany the Traveler. Ellen JOURNEYED to Europe with five suitcases.
Co-participant [co-p] **Semantic Type:** Sentient	The Co-participant is the person or persons who accompany the Traveler on the journey.
Depictive [Depict] **Semantic Type:** State	The state of the Traveler during the journey. We TRAVELED around unencumbered.
Descriptor []	A characterisitic of the traveling event.

Figure 5. The `Travel` frame in Berkeley FrameNet

Definitions and examples, as well as the background colors used in the reports are mostly meant to help human users in their experience with framenet data. We will thus focus on the status and semantic type of FEs and on the relations.

Each FE may be assigned a Semantic Type, whose function is to capture commonalities among the frame-specific semantic roles. The Sentient Semantic Type, for example, applied to the FE TRAVELER in the `Travel` frame, indicates that the entity instantiating this FE must be in control of its will to carry out actions. Types may also be assigned to frames, indicating whether they are evoked by lexical items or not, and to LUs, indicating their semantic polarity, for example.

Because the annotation of non-core – that is, peripheral and extra-thematic – FEs is not mandatory, they do not take part in frame-internal relations, which are meant to model the fact that FEs that are necessary for the frame to be instantiated may be necessary in three different ways. The first kind of frame-internal relation is Excludes. In such a relation, the instantiation of a given FE precludes the others related to it from being instantiated. In the `Travel` frame, the FE AREA excludes GOAL, PATH and SOURCE. Requires is the opposite of the Excludes relation and holds between two FEs that must be instantiated together (Ruppenhofer et al. 2016).

The third kind of frame internal relation is the Core set. Core sets occur when the presence of a given core FE makes the instantiation of the other FEs in the set optional (Ruppenhofer et al. 2016). In the `Travel` frame, the GOAL, the PATH and the SOURCE are in a Core set, meaning that one can mention one, two or all of those FEs, as shown in (6)–(8), respectively.

(6) [I_{TRAVELER}] **traveled**^{Target} [to Carmel$_{\text{GOAL}}$].

(7) [I_{TRAVELER}] **traveled**^{Target} [from San Francisco$_{\text{SOURCE}}$] [to Carmel$_{\text{GOAL}}$].

(8) [I_{TRAVELER}] **traveled**^{Target} [from San Francisco$_{\text{SOURCE}}$] [to Carmel$_{\text{GOAL}}$] [along Hwy 1$_{\text{PATH}}$].

In regards to frame-to-frame relations, Berkeley FrameNet defines eight of them: Inheritance, Using, Perspective_on, Subframe, Precedes, Causative_of, Incohative_of and See_also. The `Travel` frame inherits from `Self_motion`, meaning that traveling is a kind of self-propelled motion. In computational terms, it means that every core and peripheral FE in the `Self_motion` frame must be mapped to a FE in the `Travel` frame. `Travel` also has a subframe, `Setting_out`, meaning that the latter is a separate event that happens inside the first. For the Subframe relation, as well as for the other ones, there's no such strict mapping requirements as the one needed for Inheritance relations. A diagram with frame-to-frame relations involving the `Travel` frame is presented in Figure 6. The FE-to-FE relations sustaining the Inheritance between `Travel` and `Self_motion` are also shown.

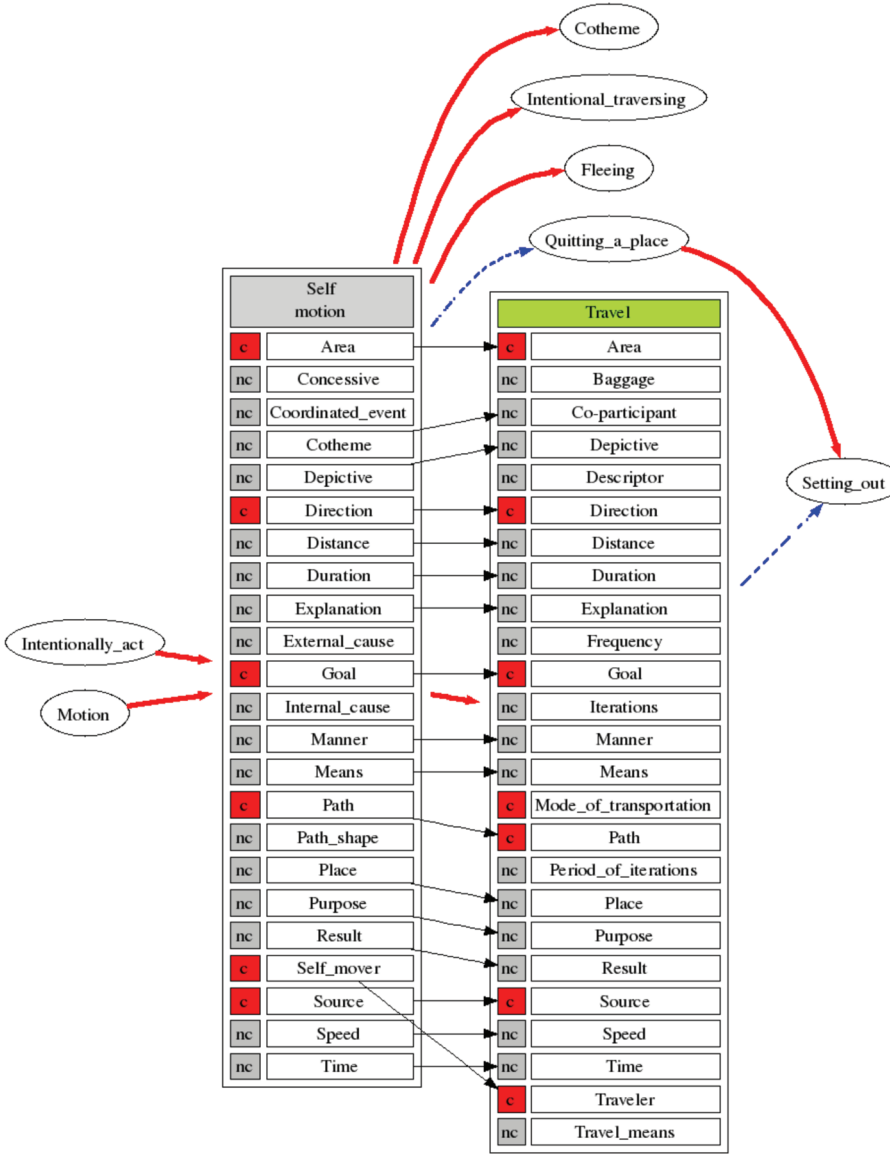

Figure 6. The frame-to-frame relations involving the `Travel` frame as shown in Berkeley FrameNet's FrameGrapher.

The LUs evoking the `Travel` frame, according to the FrameNet database for English are: *commute.v, excursion.n, expedition.n, getaway.n, jaunt.n, journey.n, journey.v, junket.n, odyssey.n, peregrination.n, pilgrimage.n, safari.n, tour.n, tour.v, travel.n, travel.v, traveler.n, trip.n, voyage.n, voyage.v*. For each one of these LUs, there will be (a) an associated lemma, which, in turn, will have one or more associated lexemes with their inflection possibilities – that is, their word forms; (b) a part of speech; (c) a definition; (d) annotation sets exemplifying the instantiation patterns of the FEs in the local syntactic context of the LU; and (e) valence patterns derived from annotation.

Last but not least, constructions share similarities with both frames and lexical units. On the frame side, constructions also have an elaborate internal structure of constituents, the CEs, although they also have an external syntax that does not correlate to any property of frames. They also are related to each other in a network of construction-to-construction relations, at the same time that their daughter signs may be related to each other in different ways. On the LU side, constructions may also evoke frames, and may require the presence of some specific lexical material.

Given the properties of frames, LUs and constructions just presented, plus the fact that frames can be evoked by both LUs and constructions, as shown in Figures 1–2, the requirements for modeling those three entities in one single integrated framenet database may be summarized as shown in Figure 7.

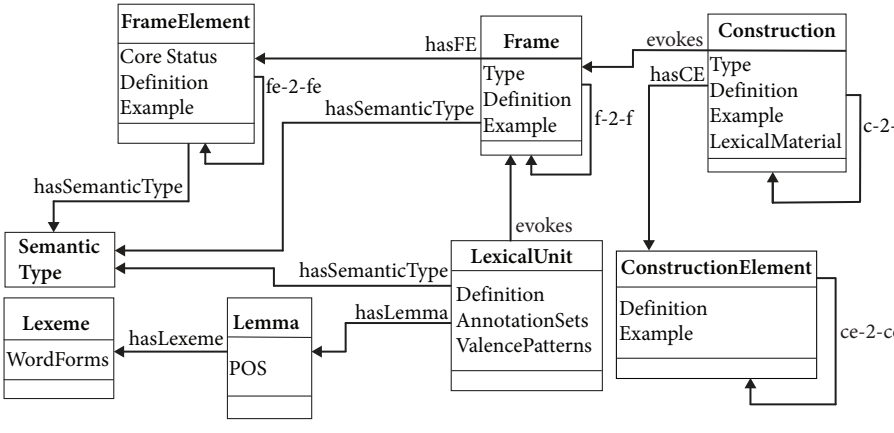

Figure 7. Requirements specification for modeling frames, LUs and constructions in a framenet

Since the requirements specification is outlined, we now move to presenting the implementation effort carried out to model the theoretical issues presented so far.

3.2 FN-Br 2.0

FN-Br 2.0 is the computational implementation developed to model the theoretical issues presented so far. The relational model, the same as the one used in Berkeley FrameNet (Baker et al., 2003), was preserved for the database so as to make FN-Br 2.0 easier to align with other framenets. Also following Berkeley FrameNet, the Relational Database Management System (RDBMS) used is MySQL.

The database structure was modeled based on four premises:

1. Enhance data consistency and integrity;
2. Support multilinguality;
3. Facilitate the creation of relations between the entities in FrameNet Brasil;
4. Reduce the number of auxiliary tables.

Data consistency is incremented by the use of Foreign Keys (FKs), a resource present in most RDBMSs (Elmasri & Navathe, 2010). FKs help maintain the integrity of the relations between tables in the database because a given record can only refer to records in other tables if the Primary Key (PK) that is referenced actually exists. Such a feature precludes records being referenced to by other entries to be removed from the database by some user's mistake or a flaw in the system. In FN-Br 2.0, FKs are also indexed, reducing the time necessary to access the data in join operations between tables.

Multilinguality support was restructured as well. In the first version of the database supporting the FrameNet Brasil World Cup Dictionary (Torrent, Salomão, Campos et al., 2014), each language-specific representation of a frame was a separate record in the database. Hence, working with three different languages demanded the creation of three records in the Frame table. A new frame-to-frame relation (Translation) had to be created to associate the three frames to one another. Although this is a fully functional solution – it was used for the FrameNet Brasil World Cup Dictionary –, it generates two problems: the complexity in information retrieval (through SQL queries) is augmented, and the occurrence of inconsistencies becomes more frequent, because the FE-to-FE relations must also be repeated every time a Translation relation is posited.

In FN-Br 2.0 a new approach was adopted. Fields requiring translations – such as Name, Description and so on – are shared by many components of the framenet model. Those fields were grouped in the Entry table together with the fields Entry and Language. The Entry table is shared by every component that might need a multilingual representation. In the case of frames, for example, one single record is created for each frame. This record has an Entry field referencing the Entry table. For each language, only one join operation is needed between the Frame and Entry

tables so as to generate a language-specific representation for that frame. Beyond initiatives that are multilingual from the very beginning – such as the m.knob project –, this feature allows for a more controlled expansion of the Berkeley FrameNet database into Brazilian Portuguese, since all the correspondences – them being partial or complete – between English and Brazilian frames are managed through the Entry table. Figure 8 illustrates this process, with the `Sufficiency` frame in the FN-Br 2.0 database.

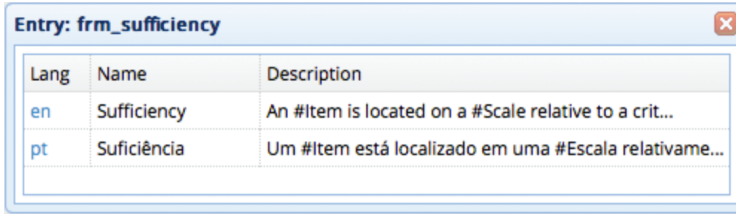

Figure 8. Entries for the `Sufficiency` frame in FN-Br 2.0

As shown in Section 3.1, modeling the continuity between grammar and the lexicon implies the establishment of relations (and self-relations) among the components in the Lexicon and in the Constructicon. In the relational model, such relations are implemented as associations between tables – more specifically, between the records in the tables –, through the use of FKs. Nevertheless, there are also cases in which a given record in Table A, for example, may be associated to many records in Table B, and vice-versa, in a many-to-many relation. In those cases, relational models require the creation of additional tables to represent associations. In a framenet, the typical case is that of frame-internal relations. Take the `Travel` frame depicted in Figure 5, for example. As it may be seen, the FE Area excludes the FEs Source, Path and Goal. To model these relations, an additional table is created every time a relation is posited between Area and the other three FEs, as well as additional tables are created for every FE-to-FE relation structuring the core set. A greater number of tables augments the complexity of the model and demands constant maintenance of the programs accessing the database.

FN-Br 2.0 adopted a different strategy. An *Entity* table was created, representing the components in the model at a higher level of abstraction. Each component that may be involved in a many-to-many relation with another component is considered to be a type of *Entity* – in fact, such components inherit from an *Entity*. In other words, each table representing a component is associated to the *Entity* table. Because each component is an entity, a relation between components can be abstracted as a relation between entities. The extant relations between entities are

stored in the *EntityRelation* table. Each relation has a specific type, stored in the *RelationType* table. The relation types are also clustered in groups, allowing for the distinction between, for instance, frame-to-frame relations and FE-to-FE relations.

This strategy fits well with the discussion presented in Section 3: there is a gain in terms of generality – blurring the limits between lexicon and grammar – at the same time that the specialization of each entity is allowed through the definition of different and specific types of relations. Moreover, the creation of new types of relations between components becomes more flexible, without the need to change the database structure. Currently, the following components are treated as entities: *Construction, ConstructionElement, Frame, FrameElement, GenericLabel, Label, LayerType, LU, SemanticType, Property, SubCorpus, Template, TypeInstance* and *POS*.

Finally, in relational models, *Status* and *Type* tables are common. *Status* tables store records indicating the statuses of other records in the database. For instance, an *AnnotationSetStatus* table may store the possible statuses for an *AnnotationSet*. *Type* tables, on the other hand, store records indicating the association of a given type to another record. For example, the *InstantiationType* table can store the types of null instantiations of a FE in a sentence. Usually, such tables feature a small fixed number of records, which must be related to the main table. The creation of new tables like these two leads to a structural change in the model.

FN-Br 2.0 does not use tables for specific types and statuses. Two other tables – *Type* and *TypeInstance* – are used for this function. Each record in the *Type* table corresponds to a specific type (*CoreType, InstantiationType, FramalType, LexiacalType, StatusType, AnnotationStatusType*). Values for each specific type are stored in the *TypeInstance* table. Each type is associated with its values through the *hasType* and *hasStatus* relations. Therefore, new types can be easily created (as well as new values can be added to the existing types) without the need to change the structural model of the database.

Because of the implementations presented in this section, FN-Br 2.0 provides the computational environment needed for the development of analyses that embrace the continuity between grammar and the lexicon. In Section 4 we will present two sample analyses that demonstrate the potential of FN-Br 2.0 for accounting for the relation between frames and constructions.

4. Sample analyses

In this section, we analyze two constructions in Brazilian Portuguese, the `dative_with_infinitive` cxn, deploying the features made available by FN-Br 2.0, and the `inceptive_aspect` cxn, which presents new challenges that must be addressed by the model.

4.1 The dative with infinitive construction

Two aspects of the model proposed for the `dative_with_infinitive` cxn will be analyzed in the following subsections: first, we show how frame evocation is treated in FN-Br 2.0; second, we demonstrate how inheritance relations between constructions are accounted for.

4.1.1 *Constructions may evoke frames*

As shown in Section 2.2, the `dative_with_infinitive` is a frame-bearing construction, evoking the `Sufficiency` frame. Moreover, there is a one-to-one relation between the CEs in the construction and the FEs in the relevant frame. The former FrameNet Brasil software apparatus and database structure, which were derived from those of Berkeley FrameNet, did not support the creation of construction-to-frame relations such as the one needed to properly account for the semantic import of the `dative_with_infinitive` and many more constructions in Brazilian Portuguese.

Since both frames and constructions are entities in FN-Br 2.0, a new relation type was created in the database: the Evoking relation. In this relation, a construction is mapped to the frame it evokes and, in case there is a CE-to-FE correspondence, this information is also stored in the database, allowing the annotation tool to automatically assign the relevant FE labels to the linguistic material instantiating the CEs once they are annotated. Figure 9 shows the Relation Editor tool in FN-Br 2.0. Note that the CE HEAD_NP (NP_NÚCLEO) in the `dative_with_infinitive` cxn maps to the FE ITEM in the `Sufficiency` frame, while the CE PARA_SINF maps to the FE ENABLED_SITUATION (SITUAÇÃO_HABILITADA). The existence of such a mapping allows for the automatic annotation of the FEs shown in Figure 3, Section 2.2.

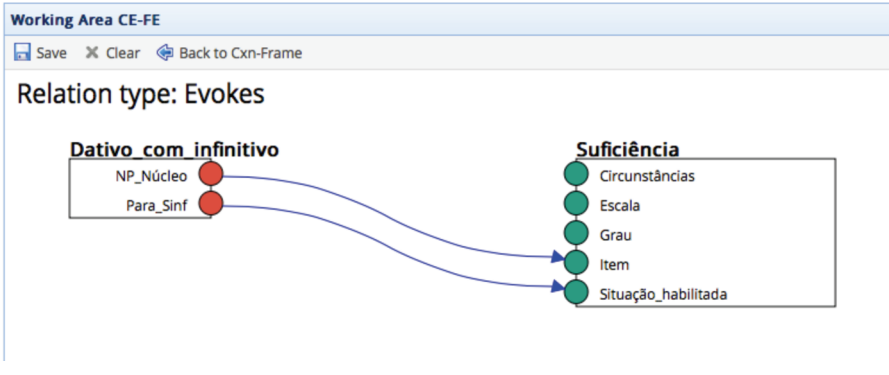

Figure 9. The evoking relation in FN-Br 2.0

Although having a similar name, the relation represented in Figure 9 is not the same as the *evokes* keyword in Embodied Construction Grammar (ECG). First, in ECG, the meaning pole of a construction is represented in terms of embodied schemas (Bergen & Chang, 2005, p. 151), which may be frame-like, but also include cognitive structures such as executing-schemas (x-schemas) and image schemas (Dodge, 2010, p. 43–44; Bergen & Chang, 2013, p. 177–178). Second, in ECG, the *evokes* keyword is not meant to relate a construction to a schema in terms of how the formal pole of each daughter sign of the construction maps to its meaning pole as defined in terms of a given frame, but, rather, to indicate the relation between one schema and the background schema(s) against which it is to be defined (Bergen & Chang, 2005, p. 152).

Dodge (2010, p. 47–50) provides an example of the use of the *evokes* keyword in ECG for the definition of the meaning import of the lexical constructions for the prepositions *in* and *out*. In the representation depicted in Figure 10, the form constraints for each construction specify the orthographic form of each preposition. As for the meaning, it is shown that both prepositions have their meaning defined in terms of the Trajector-Landmark (TL) schema, although perspectivized differently in each case. Such a difference in perspective is accounted for by stating that in both cases, the Bounded Object (BO) schema is evoked as a background against which the TL schema is to be defined: while in the case of *in*, the profiled area role of the TL schema is mapped to the interior role of the BO schema, in the case of *out*, the same role of TL is mapped to the exterior role of BO.

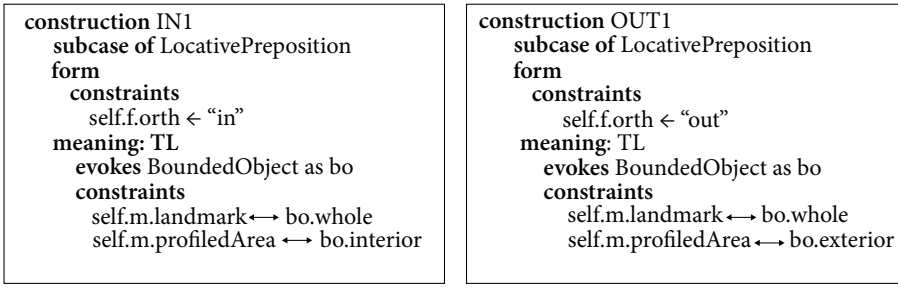

Figure 10. The representation of the constructions IN1 and OUT1 in ECG (Dodge, 2010, p. 49)

The kind of relation explicated above may also be useful for the FN-Br Constructicon, as it has proven to be so in a computational representation of ECG, the ECG Analyzer (Bryant, 2008). However, further studies are still needed in order to define the scope, constraints and application of such a relation. For now, the Evoking relation in the FN-Br Constructicon has the sole purpose of formalizing the mapping between the internal structure of constructions with that of frames.

More than allowing the partial automation of the annotation process, such a mapping may lead to a reassessment of some of the decisions made by Berkeley FrameNet in the process of frame creation. As an example, let us compare (9), (10) and (11):

(9) Eu recebi um dinheiro para pagar as contas.
I receive.PAST.1SG a money to pay.INF the bills
I received some money to pay the bills.

(10) Eu comprei um livro para estudar pra prova.
I buy.PAST.1SG a book to study.INF to test
I bought a book to study for the test.

(11) Eu tenho dinheiro para pagar as contas.
I have.PRES.1SG money to pay.INF the bills
I have the money to pay the bills.

If we consider the verbs *receber* 'to receive', *comprar* 'to buy' and *ter* 'to have' as lexical targets, (9), (10) and (11) can be annotated for the `Receiving`, `Commerce_buy` and `Possession` frames, respectively, as shown in (9a)–(11a):

(9a) [Eu$_{\text{Recipient}}$] *RECEBI*^{Target} [um dinheiro$_{\text{Theme}}$] [para pagar as contas$_{\text{Purpose_of_theme}}$]

(10a) [Eu$_{\text{Buyer}}$] *COMPREI*^{Target} [um livro$_{\text{Goods}}$] [para estudar pra prova$_{\text{Imposed_purpose}}$]

(11a) [Eu$_{\text{Owner}}$] *TENHO*^{Target} [dinheiro para pagar as contas$_{\text{Possession}}$]

Note, first, that there is an asymmetry in the annotation, with the infinitival sentence headed by *para* being annotated as a separate FE in (9a) and (10a), and as part of the direct object in (11a). In fact, the annotation proposed for (11a) does not account properly for the constituent structure of this sentence. Since *para pagar as contas* could be easily moved to left of the sentence, it is not likely to be a part of the NP headed by *dinheiro*. The reason why (11a) is annotated as such is due to the fact that there is no non-core FE that could be assigned to the *para* sentence.

One possible solution would be creating such an FE, maybe named PURPOSE_OF_POSSESSION. However, there is an alternative analysis that represents a gain in generality: instead of proposing frame-specific non-core FEs to deal with instances of the same kind of infinitival sentence, one could annotate (9)–(11) also for the `dative_with_infinitive` cxn, thus assigning the ENABLED_SITUATION FE of the `Sufficiency` frame automatically to all instances. Such an analysis would capture the general fact that the money and the book in sentences (9)–(11) are the ITEMS whose `Sufficiency` enables the first person to pay for the bills and study for the test.

When more frame-bearing constructions are added to FN-Br 2.0 and analytical generality reaches domains other than purpose, the role of non-core FEs

may be deeply revised, i.e., instead of creating non-core FEs for each frame whose LUs are attested in sentences featuring these circumstantial elements, one could create more general constructions evoking circumstantial frames, and restrict the lexical annotation to those FEs that are truly defined by the valence properties of the lexical item.

Importantly, no claim is being made towards using some kind of argument-adjunct distinction to draw a line between FEs that should be created and those that shouldn't. The claim being made here still respects the basic BCG assumption that the minimal valence of lexical items includes both core and peripheral valents. However, as Fillmore (2013, p. 132) points out:

> In addition to 'core' and 'periphery' (…), there are also constructions that introduce into clauses various 'extrathematic' subordinate structures that are not directly a part of the semantic frame of the syntactic head of the clause.

Such a distinction is also valid for Berkeley FrameNet (Ruppenhoffer et al., 2016), although it seems not to be consistently applied for the specific examples annotated in (9a)–(11a): Berkeley FrameNet classifies the PURPOSE_OF_THEME FE in the `Receiving` frame as peripheral, while the IMPOSED_PURPOSE FE in the `Commerce_buy` frame is defined as extra-thematic.

The integrated approached in favor of which we advocate, would also, thus, enhance the consistency of the FN-Br Lexicon, since a richer constructicon has the potential for providing an additional more reliable criterion for the differentiation between peripheral and extra-thematic FEs, removing the need for creating the latter in several different frames.

4.1.2 *Constructions may inherit from other constructions*

Besides being related to the frames they evoke, constructions are also related to other constructions in a network (or lattice) of inheritance relations. Inheritance is usually approached in two different ways in Construction Grammar. As Kay (2005) points out, cognitively inspired approaches, such as Goldberg's (1995, 2006), adopt multiple inheritance link types, such as Polysemy, Instance, Subpart and Metaphor (Goldberg, 1995, p. 75–81), while monotonic approaches, such as Kay & Fillmore's (1999), adopt only one type of inheritance link.

In the first approach – the so-called normal mode of inheritance (Goldberg, 1995, p. 73–74) –, links between constructions are defined as cognitive objects and are meant to capture how the mother construction motivates the daughter. No strict constraints on how much of the information in the mother construction is transferred to the daughter are posited, as long as the daughter construction does not conflict with the mother. Hence, subregularities, exceptions and partial generalizations are allowed.

In the latter approach – the complete mode of inheritance –, links between constructions are meant to account for the generalities observable across the network of constructions. All the information in the mother construction must be equally or more specifically present in the daughter (Kay & Fillmore, 1999, p. 7).

Beyond the differences in regards to the kinds of constraints that must be satisfied when positing a link, the adoption of either the normal or the complete mode of inheritance may lead to proposing a completely different set of constructions for a language. Kay (2005) demonstrates this fact, by proposing an alternative monotonic approach to Goldberg's (1995) account for the ditransitive cxn. In her analysis, Goldberg (1995, p. 75–77) proposes five Polysemy links connecting the central sense of the ditransitive cxn – 'X causes Y to receive Z' – to five extensions of this sense, which include X enabling Y to receive Z, X causing Y not to receive Z or X intending to cause Y to receive Z. Kay (2005) argues that, instead of positing the existence of Polysemy links connecting six different ditransitive constructions, the grammar of English should feature one `abstract_recipient` cxn, which adds a recipient argument to the minimal valence of a predicator, and three maximal subconstructions: the `direct_recipient` cxn, the `intended_recipient` cxn and the `modal_recipient` cxn. By exclusively using complete inheritance and the same kind of unification processes between frames used for the `dative_with_infinitive` construction in Figures 1 and 2, Kay (2005) shows the variation in the senses of a construction can be accounted for by the interaction of frames and their elements inside the construction.

Building on that and similar analyses (Kay & Fillmore, 1999; Fillmore, 1999), our first attempt to model construction-to-construction relations in the FN-Br Constructicon adopts the complete mode of inheritance. Hence, when a daughter construction inherits from its mother in the resource, all the CEs in the mother must map to an equal or more specific CE in the daughter. Multiple inheritance is allowed, meaning that a construction may inherit structure from more than one mother.

The approach we adopted to inheritance is very similar to the one used in ECG with the *subcase of* keyword (Dodge, 2010, p. 51). In ECG, this keyword is used for modeling both construction and schema inheritance relations. In FN-Br 2.0, inheritance also holds between constructions and between frames, with almost the same kinds of constraints. Let us now return to the `dative_with_infinitive` cxn and see how its network of inheritances is formed.

As pointed out by Laviola (2015), the combination of a HEAD_NP showing an augmented valence that requires a PARA_SINF may yield two different readings in Brazilian Portuguese: that of enablement/`Sufficiency` already discussed in Section 2, and one of obligation, as exemplified in (12).

(12) Eu tenho provas pra corrigir.
I have.PRES.1SG tests to correct.INF
I have some tests to grade.

Laviola (2015) has also shown that, for the obligation reading to be available, the HEAD_NP, *provas* in the case of (12), must instantiate what would be a core FE of the frame evoked by the infinitival verb. Moreover, such a core FE is preferably an undergoer-like function. In (12), the frame evoked by *corrigir* is `Assessing`, and *provas* would be assigned the FE for the PHENOMENON being assessed by the first person ASSESSOR.

Therefore, two `dative_with_infinitive` constructions seem to exist in Brazilian Portuguese, since both the meaning and the unification constraints of the constructions licensing (5) and (12), the enablement and the obligation readings, respectively, are different. Nevertheless, both constructions are a more specific type of an `infinitival_relative` cxn whose infinitival sentence is headed by *para*. Since Brazilian Portuguese also admits other prepositions acting as heads of `infinitival_relative` constructions, such as *de* 'of' for example, the `para_infinitival_relative` cxn would be a more specific type of the former, which in turn, is a more specific type of a general `relative` cxn.

Because the presence of *para* brings to the `para_infinitival_relative` cxn a subjacent purpose reading that is not present in the `de_infinitival_relative` cxn, the former also inherits from the `purpose_adjunctive_clause` cxn. The inheritance network just described is shown in Figure 11, which depicts the working area of the FN-Br 2.0 web tool where such relations are modeled.

Note that there are two inheritance paths leading to the `dative_with_infinitive` constructions: the one in the right shows that both are a specific type of

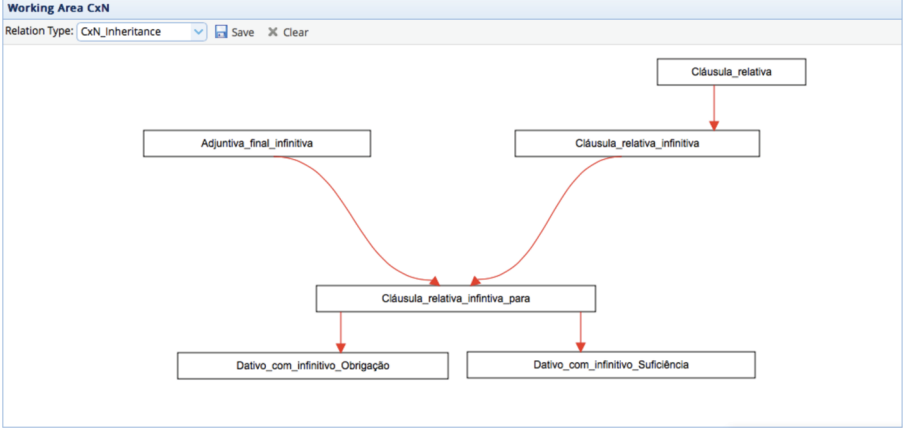

Figure 11. The inheritance network of the `dative_with_infinitive` constructions

relative_clause (cláusula_relativa), while the one in the left shows that they also incorporate structure from a purpose_infinitival_adjunctive clause (adjuntiva_final_infinitiva). Both inheritance paths converge to a para_infinitival_relative_clause (cláusula_relativa_final_infinitiva), an abstract construction whose structure is shared by the two dative_with_infinitive constructions. The relative_clause inheritance path has one additional level, the one featuring an abstract infinitival_relative_clause (cláusula_relativa_infinitiva), whose structure is shared by both the *para*-headed and the *de*-headed infinitival relatives.

Inheritance relations modeled by FN-Br 2.0 also take into consideration the internal constituency of the constructions involved, meaning that the daughter signs of the constructions are mapped to each other. Such a mapping may occur in either a one-to-one or a many-to-one fashion.

The first case, observable in the relative_clause inheritance chain, is straightforward: because the dative_with_infinitive constructions are ultimately a type of relative clause, the HEAD_NP CEs in the daughter constructions of this chain are mapped to the HEAD_NP CEs in the mothers all the way up. In turn, the PARA_SINF CEs in the bottom two levels of the chain are mapped to a SINF CE in the infinitival_relative cxn – in which the head preposition is left unspecified – and to a SREL CE in the relative_clause cxn, in which the kind of VP in the relative sentence is left unspecified.

The second case requires deeper explanation. The reason for positing an inheritance link between an adjunctive clause and (ultimately) a relative clause may seem obtuse. However, Torrent (2009, 2015) has shown that the PARA_SINF CE in the dative_with_infinitive cxn inherits the structure and constraints of the purpose_adjunctive cxn, specially the fact that, unlike most typical relative clauses, Para-infinitival ones can be fronted. Hence, an inheritance link connecting these constructions is posited. Since the purpose_adjunctive cxn does not share the same internal structure of a relative_clause, this inheritance link is of a many-to-one kind: the two CEs of the mother construction – the preposition *para* and the infinitival sentence – are both mapped to the PARA_SINF CE, as shown in Figure 12.

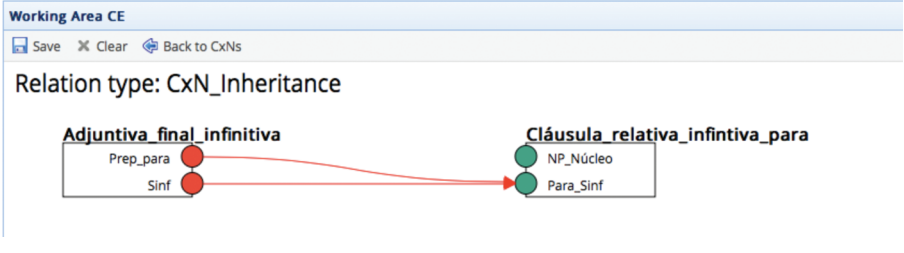

Figure 12. The CE-to-CE mapping sustaining the inheritance relation between the purpose_adjunctive cxn and the para_infinitival_relative cxn in FN-Br 2.0

4.2 The inceptive aspect construction

The analyses presented in the previous section demonstrate the advances already achieved by FN-Br 2.0. However, other important aspects of constructions and their relations with frames are still to be modeled. In this section, we present another construction in Brazilian Portuguese, the `inceptive_aspect` cxn (Sigiliano, 2011).

This construction has two daughter signs, one aspectual marker in the finite form followed by an infinitival VP, which may or may not be headed by a preposition. The first sign may be instantiated by several different verbal stems, some of which are canonical inceptive aspectual markers, such as *começar* and *iniciar* 'start', while others are not typically aspectual, such as *danar* 'harm', *desatar* 'untie', *entrar* 'enter' and *romper* 'break'. Sentences (13)–(17) are constructs licensed by this construction.

(13) Maria começou a estudar logo cedo.
Maria start.PAST.3SG to study.INF soon early
Maria started studying early in the morning.

(14) Maria danou a reclamar do irmão.
Maria harm.PAST.3SG to complain.INF of the brother
Maria started to (iteratively) complain about her brother.

(15) Maria desatou a falar mal do emprego.
Maria untie.PAST.3SG to talk.INF badly of the job
Maria started to (iteratively) complain about her job.

(16) Maria rompeu a chorar.
Maria break.PAST.3SG to cry.INF
Maria burst into tears.

(17) Entrou a chover.
enter.PAST.3SG to rain.INF
It started raining.

As it may be seen from the examples, the different aspectual markers yield slightly different inceptive readings.

As for the infinitival verb, Sigiliano (2011, p. 131) points out that 13 different semantic types of verbs can occur in the constructs licensed by this construction, as shown in Table 1.

Sigiliano (2011) shows that the combination of the aspectual marker with the VINF is not free of constraints, and, also, that metaphors are key in the definition of such restrictions. We will look into each of these two aspects more in detail in the next two sections.

Table 1. Semantic types of VINF adapted from Sigiliano (2011)

Semantic type of VINF	Examples
Requesting	*ordenar* 'give orders', *persuadir* 'persuade'
Desiring	*querer* 'want', *desejar* 'desire'
Perception	*ver* 'see', *ouvir* 'listen'
Cognitive process	*saber* 'know', *entender* 'understand'
Believing	*achar* 'think', *acreditar* 'believe'
Communication	*dizer* 'say', *falar* 'talk'
Motion	*ir* 'go', *vir* 'come'
Action	*fazer* 'do', *pegar* 'take'
State	*ser* 'be', *permanecer* 'remain'
Natural phenomenon	*chover* 'rain', *nevar* 'snow'
Change of state	*secar* 'become dry', *passar* 'pass'
Feeling	*amar* 'love', *odiar* 'hate'
Emotion expression	*chorar* 'cry', *soluçar* 'hiccup'

4.2.1 *Constructions specify slot-filling constraints*

Sigiliano (2011, p. 132), in a corpus-based diachronic study, demonstrates that, while some of the non-canonical aspectual markers take almost any type of VINF, such as *entrar* 'enter', attested with 12 of the 13 semantic types listed in Table 1, others, such as *romper* 'break', take only four types. Also, for the case of *romper*, 68% of the licensed constructs involve an infinitival verb of emotion expression, 18% a verb of action, 7,8% one of motion and 6,2% a verb of communication.

In addition, the combination of *romper* with other semantic types of VINF sounds rather infelicitous in Brazilian Portuguese, as it is shown, for instance, in (18a)–(18d).

(18) a. *Maria rompeu a persuadir os colegas.
Maria break.PAST.3SG to persuade.INF the colleagues
**Maria burst into persuading the colleagues.*

b. *Maria rompeu a querer um carro.
Maria break.PAST.3SG to want.INF a car
**Maria burst into wanting a car.*

c. *Maria rompeu a ouvir os colegas.
Maria break.PAST.3SG to listen.INF the colleagues
**Maria burst into listening to her colleagues.*

d. *Maria rompeu a saber a verdade.
Maria break.PAST.3SG to know.INF the truth
**Maria burst into knowing the truth.*

Hence, in some cases, the attested constructs indicate a strong correlation between the aspectual marker and the semantic type of the VINF, indicating the existence of constraints regarding the filling of the auxiliary and VINF slots of the construction. In terms of modeling, this and other problems must be addressed.

First, the `inceptive_aspect` cxn may evoke either the `Activity_start` or `Process_start` frames, depending on the valence affordances of the VINF: when the VINF valence requirements involve an agentive external argument, the frame evoked is `Activity_start`; when it doesn't, the frame evoked is `Process_start`, provided that the difference between these two frames is precisely the lack of an AGENT FE in the latter. Such a distinction in the `X_start` frames in Berkeley FrameNet is related to the fact that, in the lattice of frames, `Activity_start` inherits from `Process_start`, adding the AGENT FE and turning the EVENT FE in `Process_start` into a more specific ACTIVITY FE.

This first problem could be solved, in principle, by following the same solution adopted for the two readings of the `dative_with_infinitive`, that is, by creating two inceptive constructions in the Constructicon. An alternative solution would be to create only one construction and to model the constraints according to which the existence or absence of an agentive FE in the frame evoked by the VINF would define whether the frame evoked by the construction is `Process_start` or `Activity_start`.

Despite the fact that the first solution is already feasible in FN-Br 2.0, we claim that the infrastructure for the kind of constraint modeling proposed in the second solution needs to be included in FN-Br 2.0. If dealing with this difference in the frames evoked by the construction is not a good enough reason for that, it would still be needed to allow the model to account for the collocational restrictions that characterize the second problem to be addressed in regards to the `inceptive_aspect` cxn: how to tell the system that sentences like (16) are possible, while sentences like (18a)–(18d) are not?

Again, the frame evoked by the VINF plays a key role in the definition of this constraint: almost 75% of the attested examples presented by Sigiliano (2011) featuring *romper* as the aspectual marker have a VINF evoking either the `Communication` or the `Communication_noise` frames, which inherits from the former. Such a high percentage is not coincidental: according to Sigiliano (2011), the reason behind such a correlation is grounded on the diachronic principle of persistence (Hopper, 1991), because some residual semantics of *romper,* still present in its aspectual use, poses metaphorically grounded constraints to the types of infinitival verbs that may co-occur with it.

Such a claim leads us to approach the second challenge to FN-Br 2.0, which we will discuss in the next section.

4.2.2 *Constructions may instantiate metaphors*

In order to explain why the `inceptive_aspect` cxn featuring *romper* as the aspectual marker licenses mostly constructs with verbs of communication and emotion expression, while also blocking most of other semantic types, Sigiliano (2011) adopts Talmy's (2001) Force Dynamics. According to her analysis, aspectual markers such as *romper* 'break' and *desatar* 'untie' preserve the residual semantics of their use as main verbs, which can be represented by the force-dynamics pattern presented in Figure 13.

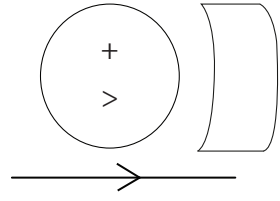

Figure 13. Force-dynamics pattern (Talmy, 2001 apud Sigiliano, 2011, p. 137)

In this pattern, the Agonist – represented by the circle – has an intrinsic force tendency to go towards (>) the Antagonist – represented by the concave figure. Since the Agonist is the strongest entity in this pattern (+), the resultant of the force interaction – represented by the arrow in the bottom of the diagram – is motion towards – and actually through – the Antagonist.

Hence, while, as a main verb, *romper* indicates that a force – the Agonist – moves through a barrier – the Antagonist – by destroying it, in sentences like (19); as an aspectual marker, *romper* maintains an image-schematic version of this meaning, indicating that an activity metaphorically goes through a barrier and begins. The ACTIONS ARE SELF-PROPELLED MOVEMENTS metaphor (Lakoff, 1979, p. 220) provides the basis for this semantic extension.

(19) A força da água rompeu a barragem.
The force of the water break.PAST.3SG the dam
The force of the water destroyed the dam.

Sigiliano (2011) moves on to explicate that the reason why verbs of emotion expression tend to occur in instances of the `inceptive_aspect` cxn featuring *romper* is due to the fact that aspectual meaning provided by the combination of this marker with the construction is that of an abrupt start. According to the author, another metaphor – EMOTIONS ARE LIQUIDS IN A CONTAINER – is also brought into play for those cases: since emotions are conceived as liquids, when the container is broken these emotions are no longer contained and tend to spread (Sigiliano, 2011, p. 139).

In a nutshell, in order to model the slot-filling constraints for the `inceptive_aspect` cxn it would be necessary not only to provide a means to automatically

evaluate whether the frame evoked by the VINF features or not an agentive FE – and, hence, define whether the construction evokes `Activity_start` or `Process_start` –, but, also, to model a metaphorically based constraint that relates the frame that would be evoked by the aspectual marker – if it was used as a main verb – and the one evoked by the VINF.

FN-Br 2.0 is unable to approach these issues for now, specially because it does not contain a metaphor repository. In the future, when metaphor repositories are made available, we can assess the possibility of incorporating them to our database.

5. Conclusions

In this chapter we showed that the continuity between grammar and the lexicon, a core principle of Construction Grammar, can be approached computationally by integrating framenet-based lexicons and constructicons in one single relational database. We demonstrated the feasibility of this task by providing exemplar analyses of the `dative_with_infinitive` constructions involving (1) the unification of the information associated with the various annotation layers, (2) the evoking relation between a construction and a frame, and (3) the inheritance relation between constructions. By showing an analysis of the `inceptive_aspect` cxn, we also pointed to new challenges that must be faced in the expansion of the analytical capacity of FN-Br 2.0, specially the modeling of frame-based and metaphor-based constraints.

Acknowledgments

The FrameNet Brasil Constructicon is funded by CAPES/STINT grant # 88887.125411/2016-00.

References

Baker, C. F., Fillmore, C. J., & Cronin, B. (2003). The structure of the FrameNet database. *International Journal of Lexicography*, 16(3), 281–296. doi:10.1093/ijl/16.3.281

Bergen, B. K. & Chang, N. (2005). Embodied Construction Grammar in Simulation-Based Language Understanding. In J. -O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions* (pp. 147–190). Amsterdam: John Benjamins. doi:10.1075/cal.3.08ber

Bergen, B. K. & Chang, N. (2013). Embodied Construction Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 168–190). New York: Oxford University Press.

Bryant, J. E. (2008). *Best-fit Constructional Analysis*. Ph.D. Dissertation in Computer Sciences. University of California Berkeley. Berkeley.

Costa, A. D., & Torrent, T. T. (2017). A Modelagem Computacional do Domínio dos Esportes na FrameNet Brasil [The Computational Modeling of the Sports Domain in FrameNet Brasil)]. In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology* (pp. 201–208). Uberlândia: Universidade Federal de Uberlândia.

Dodge, E. K. (2010). *Constructional and Conceptual Composition*. Ph.D. Dissertation in Linguistics. University of California Berkeley. Berkeley.

Elmasri, R. & Navathe, S. (2010). *Fundamentals of Database systems*. Boston: Addison-Wesley.

Fillmore, C. J. (1968). The Case for Case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1–89). New York: Holt, Rinehart and Winston.

Fillmore, C. J. (1977). The Case for Case Reopened. In P. Cole & J. Sadock (Eds.), *Syntax and Semantics 8: Grammatical Relations* (pp. 59–82). New York: Academic Press.

Fillmore, C. J. (1982). Frame Semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm* (pp. 111–137). Seoul: Hanshin Publishing Company.

Fillmore, C. J. (1999). Inversion and Constructional Inheritance. In G. Webelhuth, J. P. Koenig & A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (pp. 113–128). Stanford: CSLI.

Fillmore, C. J. (2007). Valency Issues in FrameNet. In T. Herbst & K. Götz-Votteler (Eds.), *Valency: Theoretical, Descriptive and Cognitive Issues* (pp. 129–160). Berlin and New York: Mouton de Gruyter.

Fillmore, C. J. (2013). Berkeley Construction Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 111–132). New York: Oxford University Press.

Fillmore, C. J. & Atkins, B. T. (1992). Toward a Frame-Based Lexicon: the Semantics of RISK and its Neighbors. In A. Lehrer & E. F. Kittay (Eds.), *Frames, Fields and Contrasts* (pp. 75–102). New York: Routledge.

Fillmore, C. J., Kay, P. & O'Connor, M. C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of Let al.ne. *Language* 64(3), 501–538. doi:10.2307/414531

Fillmore, C. J., Lee-Goldman, R. & Rhomieux, R. (2012). The FrameNet Constructicon. In H. C. Boas & I. A. Sag (Eds.), *Sign-Based Construction Grammar* (pp. 309–372). Stanford: CSLI Publications.

Fillmore, C. J., Petruck, M. R. L., Ruppenhofer, J. & Wright, A. (2003). FrameNet in Action: The Case of Attaching. *International Journal of Lexicography* 16(3), 297–332. doi:10.1093/ijl/16.3.297

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. New York: Oxford University Press.

Hopper, P. (1991). On some principles on grammaticalization. In E. Traugott & B. Heine (Eds.), *Approaches to Grammaticalization*, vol. I (pp. 17–36). Amsterdam: John Benjamins. doi:10.1075/tsl.19.1.04hop

Kay, P. (2005). Argument Structure Constructions and The Argument-Adjunct Distinction. In M. Fried & H. Boas (Eds.), *Grammatical Constructions: Back to the Roots* (pp. 71–98). Amsterdam: Benjamins. doi:10.1075/cal.4.05kay

Kay, P. (2013). The Limits of (Construction) Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 32–48). New York: Oxford University Press.

Kay, P. & Fillmore, C. J. (1999). Grammatical Constructions and Linguistic Generalizations: The What's X Doing Y? Construction. *Language* 75, 1–33. doi:10.2307/417472

Lage, L. M. (2013). *Frames e construções: A implementação do constructicon na FrameNet Brasil*. [Frames and constructions: The implementation of the constructicon in FrameNet Brasil]. M.A. Thesis in Linguistics. Universidade Federal de Juiz de Fora, Juiz de Fora.

Lakoff, G. (1979). The Contemporary Theory of Metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 202–251). New York: Cambridge University Press.

Langacker, R. W. (2008). *Cognitive Grammar: a basic introduction*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780195331967.001.0001

Laviola, A. B. (2015). Construções em Contraste: uma Análise Comparativa Português-Inglês no Tangente à Implementação de Constructicons. [Constructions in Contrast: a Portuguese-English Comparative Analysis Regarding the Implementation of Constructicons]. M.A. Thesis in Linguistics. Universidade Federal de Juiz de Fora. Juiz de Fora.

Miranda, N. S. & Salomão, M. M. M. (2008). *Construções do Português do Brasil*. Belo Horizonte: Editora da UFMG.

Paiva, V. M. R. L., & Torrent, T. T. (2017). Uma Proposta Metodológica para a Categorização Automatizada de Atrações Turísticas a partir de Comentários de Usuários em Plataformas Online [A Methodological Proposition for the Automatic Categorization of Touristic Attractions from User Comments in Online Platforms]. In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology* (pp. 232–239). Uberlândia: Universidade Federal de Uberlândia.

Ruppenhoffer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R. & Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. Berkeley: ICSI, available at: http://framenet.icsi.berkeley.edu.

Sigiliano, N. S. (2011). Persistência e Extensão na Construção Inceptiva em [V1fin (PREP) V2inf] do Português [Persistence and Extension in the [V1fin (PREP) V2 inf] Inceptive Construction in Portuguese]. *Letras & Letras*, 27(1).

Talmy, L. (2001). *Toward a Cognitive Semantics*, vol. I. Cambridge: The MIT Press.

Torrent, T. T. (2009). *A Rede de Construções em Para (SN) Infinitivo: Uma Abordagem Centrada no Uso para As Relações de Herança e Mudança Construcionais*. [The Para (NP) Infinitive Construction Network: a Usage-Based Approach to Construction Inheritance and Change]. Ph. D. Dissertation in Linguistics. Universidade Federal do Rio de Janeiro. Rio de Janeiro.

Torrent, T. T. (2015). On The Relation Between Inheritance and Change: The Constructional Convergence and The Construction Network Reconfiguration Hypotheses. In J. Bardđal, E. Smirnova, L. Sommerer & S. Gildea (Eds.), *Diachronic Construction Grammar* (pp. 175–214). Amsterdam: John Benjamins. doi:10.1075/cal.18.06tor

Torrent, T. T. & Ellsworth, M. J. (2013). Behind the Labels: Criteria for Defining Analytical Categories in FrameNet Brasil. *Veredas* 17(1), 44–65.

Torrent, T. T., Lage, L. M., Sampaio, T. F., Tavares, T. S., & Matos, E. E. S. (2014). Revisiting border conflicts between FrameNet and Construction Grammar: Annotation policies for the Brazilian Portuguese Constructicon. *Constructions and Frames*, 6(1), 34–51. doi:10.1075/cf.6.1.03tor

Torrent, T. T., Salomão, M. M. M., Matos, E. E. S., Gamonal, M. A., Gonçalves, J., Souza, B. C., Gomes, D. S. & Peron-Correa, S. R. (2014). Multilingual Lexicographic Annotation for Domain-Specific Electronic Dictionaries. *Constructions and Frames* 6(1), 73–91. doi:10.1075/cf.6.1.05tor

Torrent, T. T., Salomão, M. M. M., Campos, F. A., Braga, R. M., Matos, E. E. S., Gamonal, M. A., Gonçalves, J., Souza, B. C., Gomes, D. S. & Peron-Correa, S. R. (2014). Copa 2014 FrameNet Brasil. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations (pp. 10–14). Dublin: Dublin City University.