

The Brazilian Portuguese Constructicon: Modeling Constructional Inheritance, Frame Evocation and Constraints in FrameNet Brasil

Adrieli Laviola, Ludmila Lage, Natália Marção, Tatiane Tavares,
Vânia Almeida, Ely Matos, Tiago Torrent

Federal University of Juiz de Fora – FrameNet Brasil, Rua José Lourenço Kelmer s/nº - Campus Universitário, 36036-900 – Juiz de Fora, Minas Gerais, Brazil

{adrielilaviola | lu.meireleslage | ndmarcao | tatyytavares | vavagalmeida}@gmail.com, {ely.matos | tiago.torrent}@ufjf.edu.br

Abstract

FrameNet Brasil comprises two major resources: a Lexicon, expanded from the Berkeley FrameNet, and a Constructicon, whose development led to considerable changes in the FrameNet standards for data structure. This paper reports on the development of the Brazilian Portuguese Constructicon, focusing on the changes made to the FrameNet Brasil database structure so as to model (a) Inheritance relations between constructions, (b) Evoking relations between constructions and frames, and (c) Constraints applying to the Construct Elements. As a result of such changes, FrameNet Brasil integrates the two resources into one single database, yielding annotations that are able to provide a comprehensive account of both formal and semantic aspects of sentences.

Introduction

FrameNet Brasil (FN-Br) is the Brazilian Portuguese branch of FrameNet, a global initiative for creating frame-based lexical resources started at the International Computer Science Institute in 1997 (Fillmore et al. 2003). Following the path of the mother project and also of other FrameNets – Japanese, Swedish and German –, FN-Br added a Constructicon to its existing Lexicon.

FrameNet resources are computational models based on the theories of Frame Semantics of Fillmore (1982) and Construction Grammar (Kay and Fillmore 1999).

Frame Semantics claims knowledge should not be seen as a collection of simple and disconnected fragments, but as complex structures based on socially shared expectations, named frames. Therefore, a frame can be understood as a kind of scene, “a 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” (Fillmore 1982:111).

Complementarily, Construction Grammar posits that the grammar of a language is the set of its **grammatical constructions**, the rules that unite formal and semantic information into various kinds of linguistic objects, together with the principles that constrain and connect them (Fillmore 2013:112).

Such “various kinds of linguistic objects” include morphemes, words, idioms, argument structures and so on. In this sense, constructions, from the fully lexical to the highly schematic, are proposed in terms of feature structures following a continuum of complexity. Moreover, Construction Grammar is a non-derivational approach, that is: there are no assumed deep structures or empty categories.

Given the main purpose of FrameNet, i.e. that of creating a computational representation of frames, defined in terms of their participants and props, connected to each other via a network of relations, the FrameNet Brasil Lexicon adopts the same aims of Berkeley FrameNet, which comprise (i) defining LUs (Lexical Units), pairings of a lemma and a frame, and (ii) annotating sentences that exemplify the syntactic and semantic valence patterns in which the LUs occur. Valence patterns in FrameNet can be regarded as the grammar associated to lexical constructions, in the sense that they are licensed by the LU.

The Constructicon, in turn, stores grammatical information that goes beyond the grammar of the LU, both in terms of generality – i.e. it comprises more abstract argument structure constructions (cxn) evoking general non-lexical frames – and in terms of scope – it includes constructions that cannot be linked to a lexical head. Both resources are connected to each other via the frame database.

Foundations for the Constructicon

For developing the Constructicon according to the theoretical principles of Frame Semantics and Construction Grammar, it was necessary to (i) provide a means of linking constructions to frames in the same way LUs are connected to frames, (ii) model relations between constructions, so that the resulting recourse has the form of a lattice, instead of a list of entries. To achieve this goal, changes to the original structure of the FN-Br database – which was the same of the Berkeley FrameNet database – were necessary.

In the Berkeley FrameNet database, only relations between frames and between a LU and a frame are modeled. In the English Constructicon (Fillmore et al. 2012), the connection between constructions and frames – in those cases where a construction evokes a frame – is noted in the textual description of the construction, but not modeled in the database. The same holds for Inheritance relations between constructions: when an English construction is defined as a more specific subtype of another construction, this information is provided only in the description of the entry, not in the model.

For a Constructicon to represent frame evocation and constructional inheritance properly, the model has to address two aspects of constructions: (a) their similarity to LUs, to the extent that they are a pairing of one form and one meaning, and (b) their similarity to frames, since they also present internal structure, i.e. constituent parts related to each other. To do so, FN-Br included the notion of Entity in its database. Entities are the main objects in the model, and can be of various types: word forms, lexemes, lemmas, LUs, frame elements, frames, construction elements, constructions and relations. Moreover, the system allows any kind of Entity to relate to any other kind. Thus, in FN-Br, any conceptual representation can be connected to any other type of conceptual representation, provided a relation type is created for it.

Modeling Constructions: Constituency, Unification and Inheritance

Berkeley Construction Grammar (BCG) relies on three fundamental concepts for defining constructions: constituency, unification and inheritance (Kay and Fillmore 1999). In this section, we will discuss how these aspects of grammatical constructions have been addressed in the FN-Br database.

For BCG, the relations among linguistic items must be formalized so as to explain both the internal constituency of the constructs, and its external – general – properties. Thus, in the Constructicon, each construction is defined in terms of its constituent parts, the Construct Elements (CEs)

(Fillmore et al. 2012). In the FN-Br Constructicon, the daughter signs of the construction – the CEs – are created considering their formal aspects, while the semantic information is assigned through the unification of the construction with a frame. Such aspects include the name of the CE, a definition and constraints applying to it. As an example, consider the Active_Transitive Cxn in Figure 1.

Definition	
[NP[V NP]] construction type. This construction exhibits an external Subject argument and a Predicate. The internal argument is a Direct Object.	
Example(s)	
Construction Elements	
Predicate [ce_predicate_act_trans_dir]	The Predicate is composed of a verb and its direct object.
Subject [ce_subject_act_trans_dir]	The external argument of the complement taking VP.

Figure 1: The Active_Transitive Construction

As shown in Figure 1, the Active_Transitive Cxn is a subtype of the Subject_Predicate Cxn [NP [V NP]], and comprises two constituents. To make the descriptions presented in prose in Figure 1 readable to machines, FN-Br included a Constraint Editor in the database management tool. Through the Constraint Editor, it is possible to model that a CE of a given construction is a construction of its own. Figure 2 shows the constraints applied to Active_Transitive Cxn.

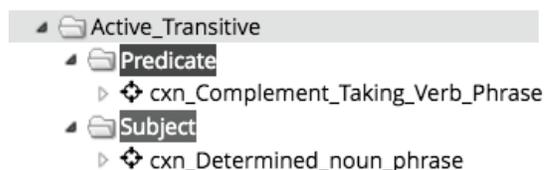


Figure 2: Constraints applied to the Active_Transitive construction

The Constraint Editor is, thus, a means of modeling both constituency and some aspects of unification, to the extent that it unifies a given CE to the set of properties defined for the constructions constraining it, also avoiding the need to model the same properties more than once.

Another aspect of unification, that comprising the mappings between form and meaning, is addressed via a

relation between constructions and frames: the Evoking relation. In this relation, the construction is mapped to the frame it evokes, and such mapping may include CE to FE links. For the Active_Transitive Cxn, the Evoking relation maps the CE Subject to the FEs Agent and Cause in the Transitive action frame, and the CE Predicate, more precisely, the Object NP to the Patient FE. Figure 3 shows a graphic representation of the Active_Transitive Cxn, featuring its CEs, their constraints, the frame it evokes, their FEs and the relations between them and the respective CEs.

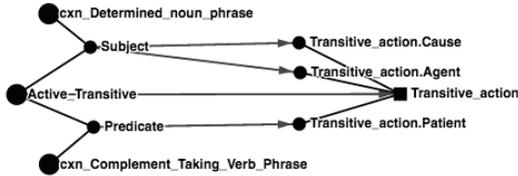


Figure 3: Graphic representation of the Active_Transitive Cxn

Finally, with the aim of implementing the Constructicon as a network, full Inheritance, employed in BCG, was used. In this relation (Kay and Fillmore 1999), all the specific information for each node dominating directly or indirectly a given node is inherited. Therefore, each daughter construction is a type of the mother construction. Inheritance is modeled in the FN-Br Constructicon as a relation between the CEs of the mother and daughter constructions. For the Active_Transitive Cxn, the inheritance relation informs it with the general properties of Subjects and Predicates, regardless of the valence of the predicate head.

One important gain of using the aforementioned relation – besides the linkage among constructions itself – is, thus, that more specific constructions can automatically inherit CEs – and also their constraints – already described in more general constructions. For instance, the General_NP defines a Noun CE, which is constrained to be a noun. The constructions Det_NP, NP and NP_Comp inherit from General_NP and, instantly, their heads must also be nouns.

Once constructions are modeled, sentences instantiating them can be annotated. If a construction evokes a frame and there’s a CE-FE correspondence modeled by the Evoking relation, the FE is automatically applied to linguistic material instantiated the CE. Moreover, additional information may be added to the annotation, such as the phrase type of the construct being annotated, for example. Figure 4 shows the annotation of (1), a sentence licensed by the Active_Transitive Cxn.

(1) O menino cortou o pano.
 The boy cut.PST.3SG the cloth
The boy cut the cloth.

Annotation in FN-Br – as in other FrameNets – is carried out in layers. The first layer indicates the CEs of the construction: the Subject – *o menino* – and the Predicate –

cortou o pano. The second layer corresponds to the frame the construction evokes: *Transitive action*. The FEs are automatically assigned in this layer through the unification between CEs and FEs. The last layer annotated represents the syntagmatic type of the construct. As mentioned earlier, the description is made only in formal terms as the semantic aspect is captured by the unification with the frame already present in FN-Br database.

[116009]	AST_MS_APP	NI	O	menino	cortou	o	pano.
CE			Sujeito		Predicado		
Transitive_action.FE			Agent		Patient		
CstrPT			Sfin			Sfin	

Figure 4: Annotation of a construct licensed by the Active_Transitive Cxn

So far, 110 constructions were modeled in the FN-Br Constructicon, comprising:

- 13 constructions from the *Para Infinitive* family (Lage 2013)
- 1 noun quantification construction (Tavares 2014)
- 65 constructions that are somehow equivalent to the ones in the Berkeley Constructicon (Laviola 2015)
- 31 constructions in the “core” grammar of Brazilian Portuguese, that is, those representing POS, phrase types and Subject-Predicate argument structures (Almeida 2016).

The constructions in (a-b) were the first to be included in the resource and follow the Berkeley Constructicon tradition, that is, they are computational models of well-studied constructions in the language. Constructions in (c), on the other hand, were included as part of an ongoing attempt to align constructicons in Brazilian Portuguese, English and Swedish. In all those cases, however, the subjacent notion that entities in the Constructicon should model non-core aspects of the language, being the core ones modeled as the valence affordances of LUs. However, as FN-Br started moving towards the development of a constructional parser, core constructions – (d) – are now included in the resource.

Those constructions do not substitute the valence patterns of LUs, but interact with them in a continuum, as it will be discussed in the following section.

Lexicon-Constructicon Integration

Assuming that lexicon and grammar are not strictly separated modules, but part of a *continuum* of constructions that ranges from more specific elements to more abstract patterns (Fillmore 2008), FN-Br is a resource in which these two components are connected, as already presented. Therefore, Lexicon and Constructicon correspond to a large network.

Since both LUs and constructions are entities in the FN-Br database that can be connected to frames – modeling the semantic import of sentences – FN-Br annotation can provide a means of unifying the different formal and functional aspects of sentences. If one considers (1) again, but, this time, instead of only annotating it for the Active_Transitive Cxn, annotates it also for the LUs in the sentence, the result of the annotation would yield Figure 5.

[116009]	AST_MS_APP	NI	o	menino	cortou	o	pano.
CE			Sujeito		Predicado		
Transitive_action.FE			Agent		Patient		
CstrPT			Sfin			Sfin	
People_by_age#.menino.n			menino				
FE	INC		Person				
GF							
PT							
Cutting.cortar.v				cortou			
FE			Agent		Item		
GF			Ext		ObjD		
PT			NP		NP		
Artifact.pano.n					pano		
FE	INC						
GF							
PT							

Figure 5: Complete annotation of *O menino cortou o pano.*

Note that layers and columns have a complementary role in FN-Br annotation. The first provide information about: (a) the CEs, FEs and Construct Phrase Type of the Active_Transitive Cxn, and (b) the FEs of the People_by_age, Cutting and Artifact frames and how they manifest in the grammar of the LUs *menino* ‘boy’, *cortar* ‘cut’ and *pano* ‘cloth’, respectively. On the other hand, the alignment of different tags related to same sequence of characters, unifies different kinds of information about each part of the sentence. For instance, from the annotation shown in Figure 5 it is possible to infer that the Subject of the sentence licensed by the Active_Transitive Cxn is a Person, defined in terms of its age, and also that the action this person executes is that of cutting an Item, which, in turn, is an Artifact.

Because the FN-Br database allows for the integration of both constructions, frames and LUs, on the structure side, and LU-based and construction-based annotation, on the other side, it provides a computational model of Brazilian Portuguese that can represent how form and meaning interact in the interpretation of sentences as a whole.

Conclusion

This paper presented the FN-Br Construction, a repertoire of Brazilian Portuguese constructions integrated to a frame-based lexicon. We showed how three key aspects of

Construction Grammar theory – constituency, unification and inheritance – are modeled in the database, and also demonstrated how the integration of the two resources enrich both the formal and semantic representation of sentences.

Acknowledgements

The FN-Br Construction project is funded by CAPES/STINT grant # 88887.125411/2016-00.

References

- Almeida, V. G. 2016. Identificação Automática de Construções de Estrutura Argumental. M.A. thesis, Department of Linguistics, Federal University of Juiz de Fora, Juiz de Fora, Brazil.
- Fillmore, C. J. 1982. Frame Semantics. In *Linguistics in The Morning Calm*, 111-137. Seoul, South Korea: Hanshin.
- Fillmore, C. J. 2008. Border Conflicts: FrameNet meets Construction Grammar. In *Proceedings of XIII EURALEX*, 49-68. Barcelona, Spain: IULA.
- Fillmore, C. J. 2013. Berkeley Construction Grammar. In Hoffmann, T. and Trousdale, G. *The Oxford Handbook of Construction Grammar*, 111-132. New York: Oxford University Press.
- Fillmore, C. J., Johnson, C. R., and Petruck, M. R. 2003. Background to FrameNet. *International Journal of Lexicography* 16(3):235-250.
- Fillmore, C. J.; Lee-Goldman, R. R. & Rhomieux, R. 2012. *The FrameNet Constructicon*. In Boas, H. C. Boas and Sag, I. A. (eds.). *Sign-Based Construction Grammar*, 309-372. Stanford: CSLI Publications.
- Golberg, A. 2006. *Constructions at Work: The nature of generalization in language*. Oxford: Oxford University Press.
- Kay, P. and Fillmore, C. J. 1999. Grammatical constructions and linguistic generalizations: The what’s X doing Y construction. *Language* 75:1-34.
- Lage, L. M. 2013. Frames e Construções: A implementação do construcion na FrameNet Brasil. M. A. thesis. Department of Linguistics. Federal University of Juiz de Fora, Juiz de Fora, Brazil.
- Laviola, A. B. 2015. Construções em Contraste: uma Análise Comparativa Português-Inglês no Tangente à Implementação de Constructicons. M. A. thesis. Department of Linguistics. Federal University of Juiz de Fora, Juiz de Fora, Brazil.
- Tavares, T. S. 2014. Construção Binominal de Quantificação Indefinita no Português – Uma abordagem construcionista. M. A. thesis. Department of Linguistics. Federal University of Juiz de Fora, Juiz de Fora, Brazil.
- 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. *Constructions and Frames* 6(1):34-51.