

The Multilingual FrameNet Shared Annotation Task: a Preliminary Report

Tiago Timponi Torrent¹, Michael Ellsworth², Collin Baker², Ely Edison da Silva Matos¹

¹Federal University of Juiz de Fora - FrameNet Brasil, Juiz de Fora, MG - Brazil

²International Computer Science Institute, Berkeley, CA - USA

{tiago.torrent, ely.matos}@ufjf.edu.br, {infinity, collinb}@icsi.berkeley.edu

Abstract

This paper presents the shared annotation task devised by the Multilingual FrameNet project together with partner projects. The shared framenet annotation task intends to probe how comparable frames are across languages by annotating translated and comparable texts using the same semantic standards in multiple languages. This paper reports on the initial work of agreeing on annotation standards, building annotation tools, and the results from the first joint frame annotations, from a TED talk and its translation into Brazilian Portuguese. The results indicate that the joint annotation task is feasible with existing FrameNet frames: over 80% of frame-bearing words in the Brazilian Portuguese translation of the TED talk fit precisely in frames found in Berkeley FrameNet's release 1.7. However, even languages as typologically similar as English and Brazilian Portuguese show some differences in density of frame-bearing words and the frequency of frame-bearing words by part-of-speech.

Keywords: Multilingual FrameNet, Shared Annotation, Interlingual Comparison

1. Multilingual FrameNet

Since 1997, the FrameNet Project at the International Computer Science Institute, in Berkeley, California, has been building a richly detailed lexical database of the core vocabulary of contemporary English, implementing the theory of Frame Semantics, developed by the late Prof. Charles Fillmore and colleagues (Fillmore 1976, 1982, Fillmore & Baker 2010). The Berkeley project has defined semantic frames, frame elements (roles) in these frames, and lexical units (word senses) which evoke the frames, extracted text from corpora and annotated the instances of these lexical units in the texts. The Berkeley FrameNet lexical database (browsable at <http://framenet.icsi.berkeley.edu>) currently contains 1,224 semantic frames, each of which has an average of 9.7 frame elements (FEs), and comprises 13,639 lexical units (LUs). There are 202,229 manually annotated instances of these lexical units, each containing annotation of the FEs that appear in the sentence.

All of this research has been done on English, but the researchers have frequently considered the obvious question: to what extent are the semantic frames created for English appropriate for analyzing other languages. Fortunately, inspired by the work at ICSI, a number of related projects have been developing frame semantic lexical databases for roughly a dozen languages, which vary in size, methodology, and availability. In all cases, the new projects have taken the Berkeley (English) frames as a starting point, although some have adhered more closely to the example of English. In general, these projects have found that a large proportion of the target-language words fit comfortably in those frames.

The FrameNet team has now embarked on a Multilingual FrameNet project, developing alignments across many of these FrameNets, seeking a better understanding of cross-linguistic similarities and differences in frame structure. Alignment on the frame level is often quite easy, as many projects have kept names or ID numbers which refer to the Berkeley frames. Going beyond frame connections, other techniques are being used to cluster and align lexical units across languages. One of these is using multilingual word vectors (Hermann & Blunsom 2014) which can be computed for a large range of languages from a wide variety of texts, and (unlike, e.g. bilingual dictionaries) lend themselves to quantitative measures of goodness of fit. We are currently testing these, but also considering techniques based on other curated resources, such as Open Multilingual WordNet (Bond & Foster 2013) and BabelNet (Navigli and Ponzetto 2012).

2. The Shared Annotation Task

The shared annotation task was devised in part as a means to evaluate the complexity of the work required to align the FrameNets developed for different languages during the past decade and more. By annotating either translations of a given text or comparable texts from the same genre and on the same topic, we aim to assess what kinds of differences must exist between FrameNets for different languages in order to provide an adequate analysis of the lexicon of each language. Moreover, the shared annotation task will generate a collection of texts annotated with frames and LUs for several languages, which can be used in the future, for instance, as training data in a variety of applications.

In the shared annotation task, annotators were limited to using the frames and frame elements from the 1.7 release of the Berkeley FrameNet data (BFN 1.7), so that everyone would annotate on the same basis. We anticipated that in many cases, a BFN 1.7 frame would be the best-fitting frame (BFF) for a word in another language, but in other cases, it might not be, suggesting that different languages might require different adaptations to those frames. In the latter case, annotators are instructed add the LU to the nearest BFN 1.7 frame, but also to indicate why that is not the best-fit frame for the LU. They could choose among the following predefined categories, or “other”:

- **Different Perspective:** the LU imposes a perspective that is different from the one in the original frame.
- **Different Causative Alternation:** the LU requires a causative interpretation that is not present in the original frame, which may be either inchoative or stative.
- **Different Inchoative Alternation:** the LU requires an inchoative interpretation that is not present in the original frame, which may be either causative or stative.
- **Different Stative Alternation:** the LU requires a stative interpretation that is not present in the original frame, which may be either causative or inchoative.
- **Too Specific:** the LU requires a frame more generic than the one available in the original database.
- **Too Generic:** the LU requires a frame more specific than the one available in the original database.
- **Different Entailment:** the LU has different entailments than the ones afforded by the original frame.
- **Different Coreness Status:** some non-core FE should be core in the target language.
- **Missing FE:** there should be a FE in the original frame that is missing.
- **Other:** all other non-listed cases.

Each annotation must include, at least, the Frame Element, Grammatical Function and Phrase Type layers. Labels in each layer can be tailored to the specific needs of each language, and, new layers can be added to the annotation.

These policies on the shared task were then carried out in a first round of shared annotation on a translated text, described in Section 2.1, using a web annotation tool developed by FrameNet Brasil, described in Section 2.2.

2.1. The Text

The first text to be annotated in the shared annotation task is the transcription of the TED Talk "Do Schools Kill Creativity?" (Robinson 2006). This is currently the most frequently viewed TED Talk, with more than 49 million views. The transcription of the 20-minute talk in English contains 267 sentences. This transcription has been translated to 61 languages by TED community members; the Brazilian Portuguese version, which will be discussed below, has 271 sentences.

2.2. The Annotation Tool

The shared annotation task is carried out with the FrameNet Brasil WebTool 3.0: a web-based database management and annotation tool, designed to allow easy customization of layers and labels from a multilingual perspective (Matos & Torrent 2016).

Because it is web-based, the tool does not require the annotation teams to install any software. Moreover, it allows teams to create language-specific annotation labels for Grammatical Functions, Phrase Types and other information. Annotators can even add new layers to the annotation system if necessary, directly in the tool interface, without having administrator privileges. This flexibility enables teams to create the analytical categories they need to address the specifics of their languages.

3. Preliminary Report

So far, consistent annotations of the TED Talk have been made for English (2 annotators) and Brazilian Portuguese (7 annotators). In this paper, we offer a preliminary contrastive report on those annotations, based, on the first 30 sentences of text in both languages, which we will refer to those sentences as the **sample text**.

The sample text comprises a total of 425 words for English and 322 for Brazilian Portuguese. Among those words, 89 different LUs were identified for English, yielding 132 annotation sets. (Each instance of each LU constitutes a separate annotation set.) For Brazilian Portuguese, 107 different LUs were identified, yielding 146 annotation sets. The annotation set/word ratio is then 0.31 for English, and 0.45 for Brazilian Portuguese. The density of annotation in the English sample text compares to 0.17 for all the full text annotation in Berkeley FrameNet; this may be due in part to a more complete annotation of the sample text and in part to a greater density of frames in the spoken genre. The difference in the density across languages is shown in more detail in Table 1, which gives the distribution of annotation sets by POS in each language.

Some of the differences, especially for conjunctions, stem from differences between the projects as to which parts of the semantics should be represented by FrameNet lexical annotation and which parts should be represented by constructions. Note that there is very little difference between the languages w.r.t. the density of annotation of verbs; we suspect that there may be two reasons for this:

1. Verbs tend to be the main predicates in sentences, evoking the central eventive frames, so translations might tend to keep the same number of central eventive frames.
2. Because semantic frames are arguably better models for events than for entities, FrameNet may simply have better, more robust models for events, which tend to be expressed more often by verbs in both languages.

POS	English	Br-Portuguese
Adjective	16	26
Adverb	6	11
Conjunction	8	20
Noun	48	51
Number	4	3
Preposition	9	5
Pronoun	-	2
Verb	41	40
TOTAL	132	148

Table 1: Distribution of annotation sets in the TED Talk sample text by part of speech of LU in each language.

In order to gauge the similarity between the annotations for English and for Brazilian Portuguese, a similarity score was calculated for each aligned pair of sentences, based on the frames evoked by the LUs in each language.

First we found the total number of frames evoked in each sentence. (When the total was different between the two languages, we used whichever number was higher.) Then the number of frames that were the same in both languages was counted and that number was divided by the total. For example, there were a total of 9 frames in sentence 7, and 4 of them were the same across languages, so the similarity score is 4/9, or 0.44. Table 2 presents the similarity scores for each of the 30 sentence pairs in the sample text and the average for all of them.

In Table 2, sentence pairs 1, 2 and 13 are marked with "N/A" because no frames have been assigned to these sentences in either language. They consist of two greetings (pairs 1 and 2) and one tag question (13). There are a number of cases where the similarity score is low

because both annotation teams added an LU that was not a perfect fit to a frame from BFN 1.7, but they each chose a different best-fit frame. We have treated these like "normal" cross-linguistic differences, but some other treatment might be appropriate.

Pair	Total Frames	Equal Frames	Score
1	N/A	N/A	N/A
2	N/A	N/A	N/A
3	1	1	1.00
4	4	1	0.25
5	1	1	1.00
6	7	4	0.57
7	9	4	0.44
8	2	1	0.50
9	10	5	0.50
10	3	1	0.33
11	2	1	0.50
12	4	2	0.50
13	N/A	N/A	N/A
14	3	3	1.00
15	6	5	0.83
16	7	2	0.29
17	4	1	0.25
18	5	0	0.00
19	2	1	0.50
20	18	5	0.28
21	5	1	0.20
22	5	3	0.60
23	11	5	0.46
24	7	4	0.57
25	11	5	0.46
26	8	8	1.00
27	13	6	0.46
28	5	2	0.40
29	3	1	0.33
30	11	5	0.45
Average Frame Similarity Score			0.51

Table 2: Frame Similarity Score between Languages per sentence pair.

In the following two sections, we discuss the main issues that emerged during the annotation for each language. Section 3.3 provides some cross-linguistic comparison of annotated sentences.

3.1. The Annotation for English

Annotating the TED talk has been challenging for Berkeley FrameNet, since it is a spoken genre, with a large number of conversation-specific LUs and constructions, such as *you know*, and *I mean*.... However, for the rest of the lexical items in the text, it has been possible to use the frames of BFN 1.7 without modification in the vast majority of instances. Out of 132 total LU instances, 125 (95%) fit their frame perfectly, 5 (e.g., *creativity.n*, *blood.n*) were in only found in frames that were too generic for the use in this text, 1 (*curiously.adv*) was in a

frame belonged to a different perspective, and 1 (*interest.n*) should actually be a MWE (*vested interest.n*), evoking a frame that does not exist in BFN 1.7.

However, these numbers hide a policy difference in the annotation of the English text compared with the Brazilian Portuguese. Until now, Berkeley FrameNet has considered pure conjunctions (e.g., *and.c*) and conversationally-grounded items like *actually.adv* and *you know.v* to be outside the scope of BFN annotation, since they are so entangled with interactional frames that FrameNet has not yet defined and with non-lexical constructions. There are 10 instances of *and.c* in the sample text, and 11 conversational particles, all of which would belong to very poorly fitting frames. If these are considered, then only 82% of LU instances belong to an appropriate frame in the annotation of the English text, which is remarkably similar to the ratio for Brazilian Portuguese, as we will see in the next section.

3.2. The Annotation for Brazilian Portuguese

Besides issues related to the fact that the TED Talk is a spoken genre, as pointed out in 3.1, the annotation of the sample text for Brazilian Portuguese was expected to pose additional challenges due to the way the shared task was designed. Since no changes could be made to BFN 1.7, we anticipated that there would be many cases in which an LU appearing in the text would have to be created in a non-BFF frame, and we provided means for annotators to do this, and save an explanation of why the frame chosen is not ideal, as a suggestion for someone about how to define the proper frame later.

Reason	Count
Different Perspective	1
Too Generic	5
Different Entailment	1
Different Coreness Status	1
Missing FE	4
Other	8
TOTAL	20

Table 3: Reasons for creating LUs with non-BFF status in Brazilian Portuguese

However, this turned out to be not very common. Among the 107 different LUs in the Brazilian Portuguese text, only 20 (18.7%) were created in non-BFF frames, meaning that Berkeley FrameNet frames provided an adequate model for more than 80% of the Brazilian Portuguese LUs. Moreover, if one considers the reasons behind the non-BFF status (shown in Table 3), BFN 1.7 frames seem to be even more easily expandable into Brazilian Portuguese.

The "Too Generic" cases, representing one fourth of the LUs created in non-BFFs, indicate that the usage would require a new, more specific frame not yet available in BFN 1.7; this proposed new frame would inherit from the non-BFF frame in which the LU in the text was created. Examples are LUs like *deus.n* 'god', in the Entity frame, and *e.c* 'and' in the Relation frame. The "Missing FE" cases all refer to non-core FEs which could be easily added to the frames, even in English, such as a Condition FE in the Concessive frame, and a Degree FE in the Causation frame. Some of the "Other" cases, however, refer to more complex (and interesting) cases, which will be discussed in the next section.

3.3. Some cross-linguistic examples

As it can be seen from Table 2, cross-linguistic frame similarity scores vary considerably from sentence pair to sentence pair. In this section, we provide examples covering three different parts in this range: sentence pairs with a 1.000 similarity score, sentence pairs with low similarity scores due to the occurrence of non-BFF frames, and sentence pairs with similarity scores close to the average, which are due to differences in translation and/or language structure.

The high end of the range is exemplified by sentence pair 26, in which sentences (1) and (2) were annotated for the same 8 frames in each language:

- (1) If you think of it, children starting school this year will be retiring in 2065.
- (2) Se formos pensar, as crianças entrando na escola esse ano estarão se aposentando em 2065. retire.PTCP in 2065

Table 4 presents the 8 frames selected for annotation and the LUs evoking each of them in English and Brazilian Portuguese.

As it can be seen from Table 4, LUs evoking the frames in both languages have the same POS. Also, none of them was assigned the non-BFF type. Although there are structural differences in the translation of (1) into (2) - e.g. the fact that *think.v* takes a second person subject in English, while *pensar.v* takes a first person plural subject in Brazilian Portuguese - such differences do not concern frame evoking material. Three other sentence pairs

received a score of 1,000, two have one LU for each language, and the other has 3.

Frame	En LU	Br-Pt LU
Conditional occurrence	if.c	se.c
Cogitation	think.v	pensar.v
People by age	child.n	criança.n
Activity start	start.v	entrar.v
Locale by use	school.n	escola.n
Calendric unit	year.n	ano.n
Quitting	retire.v	aposentar.v
Temporal collocation	in.prep	em.prep

Table 4: Frames for which sentences (1-2) were annotated and LUs evoking them in each language.

On the low end of the similarity score scale, with a score of 0.00, we find sentences (3) and (4) in pair 18.

- (3) And you're never asked back, curiously.
 (4) E curiosamente ninguém te
and curiously no one you
 convida de novo.
invite.PRES.3SG again

Table 5 shows the LUs annotated in each language and the frames they evoke. Note that there are no corresponding frames between the two languages. A "---" indicates that the frame was not evoked in one of the languages.

Frame	En LU	Br-Pt LU
Frequency	never.adv	---
Request	ask.v	---
Locative relation	back.adv	---
Typicality	curiously.adv	---
Relation	---	e.c
Manner	---	curiosamente.adv
People	---	ninguém.n
Have visitor over	---	convidar.v
Event instance	---	de novo.adv

Table 5: Frames for which sentences (3-4) were annotated and LUs evoking them in each language.

The low score in this sentence pair illustrate how different choices for non-BFF frames impact the comparability between the original sentence and its translation in terms of semantic frames. The English sentence has one LU created in a non-BFF frame (*curiously.adv*), which should actually be handled as a sentence-level modifier; it ironically suggests that the hearer should understand why educators are seldom asked again by the same host. Such a frame, which invokes the full conversational context, has not yet been defined for either language. In the Brazilian Portuguese translation, three LUs were created in non-BFF frames. One of them, *curiosamente.adv* - which actually translates as *curiously.adv* - was created in the Manner

frame, which is too generic and includes LUs such as *manner.n* and *way.n*, but not adverbs actually indicating manner. This use of Portuguese *curiosamente.adv* should probably be handled like English *curiously.adv*.

The LU *convidar.v* was created in a non-BFF frame for two reasons: first, because there was a missing non-core FE, Particular_iteration, and, second, because the Have_visitor_over frame seems to be, in fact, preceded by the frame evoked by *convidar.v*.

The LU *e.c*, which translates as *and.c* was created in the Relation frame, a very generic frame not really used by BFN for conjunctions such as this, as pointed out in 3.1.

In the middle of the score continuum, sentence pair 25, with a score of 0.46, has 5 coincidental frames out of 11. The sentences of this pair are shown in (5) and (6).

- (5) We have a huge vested interest in it, partly because it's education that's meant to take us into this future that we can't grasp.
 (6) Nos interessamos tanto por
us-RFL be-interested.PRES.1PL so much for
 ela em parte porque é da
she in part because be.PRES.3SG of
 educação o papel de nos
education the role of us
 conduzir a esse futuro misterioso.
conduct.INF to this future mysterious

Table 6 shows the frames evoked by the LUs in this sentence pair.

Differences between the frames evoked by the LUs in each version of this sentence can be classified into two types: (a) structural differences in the predicates and (b) the cascade effect of those on their modifiers.

The first predicate in each sentence is that encoding the interest people have in education. While in English, such information is coded by a noun, in Brazilian Portuguese, it is a verb that has this function, although both *interest.n* and *interessar-se.v* were created as LUs evoking the Mental_stimulus_experiencer_focus frame. BFN 1.7 defines this frame as follows: "An Experiencer has an emotion as caused by a Stimulus or concerning a Topic".

In the case of English, *interest.n* was created with a non-BFF status in this frame because this noun should actually be part of the MWE *vested interest.n*, which would then have to be created in a frame that contains the entailment that the interest in something is triggered by the fact that such something is of major importance for the collectivity.

Such an entailment is completely lost in the translation of this sentence into Brazilian Portuguese, for which the *Mental_stimulus_experiencer_focus* frame fits the verb *interessar-se.v* nicely.

Frame	En LU	Br-Pt LU
Size	huge.a	---
Mental_stimulus_exp_focus	interest.n	interessar-se.v
Degree	---	tanto.adv
Degree	partly.adv	em parte.adv
Causation	because.c	porque.c
Education_teaching	education.n	educação.n
Purpose	mean.v	---
Performers_and_roles	---	papel.n
Bringing	take.v	conduzir.v
Goal	into.prep	---
Temporal_collocation	future.n	futuro.n
Certainty	---	misterioso.a
Capability	can.v	---
Grasp	grasp.v	---

Table 6: Frames for which sentences (5-6) were annotated and LUs evoking them in each language.

However, the difference in the POS of the two LUs has a cascade effect in the other LUs in the sentence. The adverb *tanto.adv* 'so much', in this sentence, modifies *interessar-se.v* 'to be interested in'. It was annotated in the Degree frame, since it indicates to what degree the speaker is interested in education. Note that the definition of the Degree frame states that "LUs in this frame modify a gradable attribute and describe intensities at the extreme positions on a scale", and, in BFN, the Gradable_attribute FE in this frame is always instantiated as an adjective; there is, however, no reason why gradable attributes cannot be expressed by nouns or verbs. On the other hand, the end-of-scale reading of *tanto.adv*, could not, in this context, be expressed by an adjective such as *huge.a*, which was annotated for the Size frame in the English sentence, generating a frame mismatch in the sentence pair.

This difference, however, does not entail some translation loss, since size and degree are metaphorically linked. On the contrary, they highlight the importance of the net-like configuration of FrameNet at the conceptual - and not only word - level (Fillmore, Baker & Sato 2004) for cross-lingual comparison. In other words, although no obvious word-to-word relation could link and adjective like *huge.a* in English to the adverb *tanto.adv* in Brazilian Portuguese, a metaphor relation connecting the Size and Degree frames could do so.

The same kind of phenomenon is seen in the mismatch between the frames evoked by the predicates indicating the purpose/role of education to take people into the future.

Finally, the other differences derive from an inversion, in the translation, of the perspective adopted when talking about the future. In the original English version, the speaker uses a relative clause to modify *future.n*, framing it as something that people do not have the capability to understand. In the translation, the adjective *misterioso* 'mysterious' is used to modify *futuro.n*, leaving people's cognitive capacity aside. Even so, the Grasp and the Certainty frame are connected to each other in BFN 1.7 via the Awareness frame. Grasp inherits Awareness, while Certainty uses it. Once again, the fact that FrameNet is a net at the conceptual level sheds some light on differences found in the annotation of a given sentence and its translation.

4. Conclusion

The shared annotation task so far has shown that the frames of Berkeley FrameNet data release 1.7 are complete enough to serve as a basis for cross-linguistic annotation. The initial efforts at annotating an English TED talk and its Brazilian Portuguese translation show that about half of the Brazilian Portuguese frame instances are identical to the frames in English, and about 80% of the Brazilian Portuguese Lexical Units fit without caveat into the frames of BFN 1.7. It also demonstrates that some of the frame mismatches can be better understood if one considers the conceptual-level network of FrameNet. Further research is needed into whether frame definitions based on lexicographic practices are adequate for these kinds of frame mismatches in translations and structural differences between LUs across languages.

5. Acknowledgements

This material is based in part upon work supported by the U.S. National Science Foundation under Grant No. (1629989). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The annotation of the TED Talk for Brazilian Portuguese, as well as the development of the FN-Br WebTool were made possible by the CAPES/STINT Grant No. 99999.009910/2014-00 and by the CNPq Grant No. 448990-2014/8.

6. Bibliographical References

- Bond, F. and Foster, R. (2013). Linking and Extending an Open Multilingual Wordnet. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 1352–1362, Sofia, Bulgaria, August. Association for Computational Linguistics (ACL).

- Fillmore, C. J. (1976). The need for a frame semantics in linguistics. In Karlgren, H. (Ed.) *Statistical Methods in Linguistics*. Stockholm: Scriptor, pp. 5--29.
- Fillmore, C. J. (1982). Frame semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co., pp. 111--137.
- Fillmore, C. J., Baker, C. F. and Sato, H. (2004). FrameNet as a "Net". In Maria Tereza Lino (Conference Chair), et al., editors, Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), pages 1091--1094, Lisbon, Portugal, may. European Language Resource Association (ELRA).
- Fillmore, C. J., and Baker, C. F. (2010). A Frames Approach to Semantic Analysis. In B. Heine and H. Narrog, (Eds.), *Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, pp. 313--341.
- Hermann, K. M. and Blunsom, P. (2014). Multilingual Models for Compositional Distributed Semantics. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 58--68, Baltimore, USA, june. Association for Computational Linguistics (ACL).
- Matos, E. and Torrent, T. (2016). A Flexible Tool for an Enriched FrameNet. In 9th International Conference on Construction Grammar (ICCG9), Juiz de Fora, Brazil, october. Federal University of Juiz de Fora (UFJF).
- Navigli, R. and Ponzetto, S. P. (2012). BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence*, 193: 217--250.
- Robinson, K. (2006). *Do Schools Kill Creativity?* TED Talk. Video (19 minutes) is available at https://www.ted.com/talks/ken_robinson_says_schools_kill_creativity.