The Limits of Using FrameNet Frames to Build a Legal Ontology

Anderson Bertoldi, Rove Luiza de Oliveira Chishman

Applied Linguistics Graduate Program
Universidade do Vale do Rio dos Sinos (UNISINOS) – São Leopoldo, RS – Brazil

`andersonbertoldi@yahoo.com, rove@unisinos.br`

Abstract. FrameNet frames have been used to develop lexical databases and annotated corpora for different languages. This paper analyses the use of FrameNet frames to build a legal ontology for the Brazilian Law. In order to discuss the problems of such approach to ontology development, the lexical units evoking the `Criminal_process` *frame were contrasted in English and Portuguese. Frame divergence between languages has consequences not only for legal ontology development but also for the development of legal lexical resources, such as lexical databases and corpora annotation.*

1. Introduction

Semantic Web technology for the legal domain has been an important topic in the last years. Semantic Web technologies involve both applications in corporate settings, such as knowledge management and intranet systems, and public information retrieval on internet (Benjamins at al., 2005). Semantic lexicons and legal ontologies have been developed to facilitate the access to legal information.

Lexicons and ontologies sometimes are considered as a similar resource. The parallel between word sense and ontological categories in one hand and lexical relations and ontological relations on the other hand suggests the similarity of these two resources (Hirst, 2003). Nevertheless, ontologies and lexicons are different resources. In the words of Gruber (1993, p.199), an ontology is an explicit specification of a conceptualization. While lexicons represent words senses in a natural language, ontologies are, by definition, an engineering artifact that represent the knowledge of a particular area in a formal language.

This paper analyses the use of FrameNet frames to build a legal ontology for the Brazilian Law[1]. It is a first attempt to construct a legal ontology lexically oriented. In this paper, the lexical units evoking the `Criminal_process` frame are contrasted in English and Portuguese. The aim of this contrastive study is discussing the conceptual structure evoked by lexical units and how this information is particular to each country. In social-oriented areas, such as Law, concepts may not be shared among countries. Considering countries that share the same language, a legal concept may evoke different

---

[1] The work presented here was developed in the scope of the of the project *Semantic Technologies and Legal Information Retrieval Systems*, a project supported by CAPES and CNJ (Conselho Nacional de Justiça) and coordinated by Professor Dr. Rove Luiza de Oliveira Chishman.

conceptual structures, because countries do not share the same set of laws and regulations.

FrameNet frames have been used to develop lexical databases and annotated corpora for different languages. Semantic frames are considered conceptual structures independent of language (Boas, 2005; Padó, 2007). As conceptual structures independent of language, semantic frames would have the characteristic of being universal. This is the principle that enables the transfer of semantic annotation from corpus of one language to another language (Padó, 2007) and the automatic development of lexicons expanding FrameNet frames to other languages other than English (Padó e Lapata, 2005).

This paper demonstrates that FrameNet frames are not always language-independent conceptual structures. In order to discuss the problems of using FrameNet frames to build legal ontologies, this paper is structured in seven sections. Section 2 presents legal ontologies and lexicons. Section 3 presents FrameNet methodology for frame creation. Section 4 presents the methodology for FrameNet creation other than English. Section 5 presents the FrameNet `Criminal_process` frame. Section 6 presents the mismatches between legal knowledge in USA and Brazil. Section 7 presents the conclusions of this work.

2. Legal Ontologies and Lexicons

Terminological lexicons and legal ontologies have been proposed for legal information retrieval purposes. *Core Legal Ontology* (CLO) (Gangemi et al., 2005) is an ontology developed by the Institute for Theory and Techniques for Legal Information (ITTIG-CNR). This ontology is used to structure legal concepts from the terminological lexicon *JurWordNet* (Gangemi et al., 2005). *LRI-Core* (Breuker et al., 2005) is a legal ontology developed by the Leibniz Center for Law, in the scope of the European project e-COURT (Breuker et al., 2005). The main purpose of LRI-Core is to support knowledge acquisition to legal domain ontologies and allow automatic indexing of legal documents.

Terminological wordnets like *Jur-WordNet* (Sagri et al., 2004) aim to improve legal information retrieval by connecting terms through semantic relations, mainly synonymy. LOIS (Lexical Ontologies for Legal Information Sharing) (Curtoni et al., 2005) was an investigation project supported by European Commission within the e-Content program. The aim of LOIS was to build a European legal wordnet for legal information retrieval. The semantic relations connect terms in different languages. The LOIS architecture was based on another European project, the *EuroWordNet* (Vossen, 1998). In LOIS the different language databases were connected through an interlingual index. This work differs from lexicons and ontologies presented in this section because it aims at using lexical databases to develop legal ontologies.

3. FrameNet

FrameNet is a lexical database that describes word meaning according to the principles of Frame Semantics. In FrameNet lexical items are conceived as lexical units. A lexical unit is the combination of a word form with a meaning. Every new meaning of a word represents a new lexical unit. Therefore, it is the lexical unit that evokes the frame, not

the word. According to Fillmore and Baker (2010), the method of lexical analysis in FrameNet follows five steps: (1) **Characterizing the frames**, (2) **Describing and naming frame elements**, (3) **Selecting lexical units**, (4) **Creating manual annotations of sample sentences** and (5) **Automatically generating lexical entries**.

4. Methodology for FrameNet Creation

In order to discuss the use of FrameNet frames for legal ontology development, it is necessary to present some points related to FrameNet and multilinguality. FrameNet for languages other than English has been created using the expansion methodology. Expansion methodology assumes that semantic frames stay the same and only the linguistic information is substituted to create new FrameNets. This is the methodology adopted by Spanish FrameNet (Subirats, 2009) and Japanese FrameNet (Ohara, 2009). According to Lönneker-Rodman (2007, p.5), expansion methodology risk to "(…) neglecting language-specific differences in lexicalization". Lönneker-Rodman (2007) presents four types of mismatches between frames in FrameNet construction: (1) **Semantic Frame**, (2) **Frame Elements**, (3) **Semantic Type and Frame Element Coreness** and (4) **Frame Relations**. The criteria for new frame creation presented in Lönneker-Rodman (2007) show that lexical changes between two languages will change the conceptual structure, in other words, lexical changes may affect the structure of a semantic frame. This work analyses how different the conceptual structure may be in a social-oriented field like Law.

5. `Criminal_process` Frame

In the FrameNet terminology, `Criminal_process` frame is a non-lexical frame. The function of non-lexical frames is to connect semantically related frames. Non-lexical frames do not present frame-evoking lexical units. They represent complex events divided in more specific frames. `Criminal_process` frame describes the different steps of a criminal process according to the American legal system. In case of complex frames, like `Criminal_process`, each sequence of events or states is described as a single frame, related to the complex frame through *Subframe* relations and to the other subframes through *Precedes* relation. `Criminal_process` frame is divided in four subframes temporally succeeded: `Arrest`, `Arraignment`, `Trial`, and `Sentencing`. `Arraignment` frame is divided in three subframes: `Notification_of_charges`, `Entering_a_plea`, and `Bail_decision`. `Trial` frame also presents three subframes: `Court_examination`, `Jury_deliberation` and `Verdict`. There is still the frame `Try_defendant`. In FrameNet terminology, the `Trial` and `Try_defendant` frames are in a *Perspective* relation. That relation describes frames that are similar and represent two sides of the same event. Therefore, the `Trial` frame describes the organization of the trial, while the `Try_defendant` frame describes the event of trying a defendant.

6. Mismatches between Legal Knowledge in USA and Brazil

In order to contrast the legal knowledge described by `Criminal_process` frame to Brazilian legal system, firstly it was necessary to create manually frames to represent the legal information about a criminal process according to the Brazilian legal system.

`Criminal_process` frame was contrasted with a Brazilian criminal process frame considering three levels of linguistic analysis: lexical units, frames, and frame elements. In the contrastive study, it is possible to perceive that semantic frames present different levels of equivalence. Some frames in FrameNet found an equivalent frame in Brazilian legal system, with lexical units presenting equivalents in Portuguese, correspondence between the American and Brazilian legal knowledge, and the same frame elements for both frames in English and Portuguese. Other FrameNet frames found equivalence only between lexical units in English and Portuguese, that is, the legal event represented in FrameNet frame did not exist in Brazilian Legal system.

The FrameNet frame `Try_defendant` represents a legal event in which a defendant is tried by a jury or a judge in a court (FrameNet definition). The core frame elements for `Try_defendant` frame are: CHARGES, DEFENDANT, GOVERNING_AUTHORITY, JUDGE, and JURY. The lexical unit that evokes this frame is *to try* that has as an equivalent in Portuguese the lexical unit *julgar*. The legal event represented by `Try_defendant` frame is comparable to the legal event of trying a defendant in Brazil. The lexical unit *julgar* in Portuguese evokes a legal knowledge comparable to the legal knowledge evoked by *to try*. It is possible to say that `Try_defendant` is equivalent to the Brazilian legal frame `Julgar_acusado`.

Other frames, like `Notification_of_charges`, present only lexical unit equivalence. `Notification_of_charges` represents a legal event in which the judge informs the accused of the charges against him/her (FrameNet definition). The core frame elements for `Notification_of_charges` are: ACCUSED, ARRAIGN_AUTHORITY, and CHARGES. The lexical units that evoke `Notification_of_charges` frame are: *to accuse*, *charge*, *to charge*, *to indict*, and *indictment*. These lexical units present equivalent in Portuguese: *to accuse/acusar*, *charge/acusação*, *to charge/acusar*, *to indict/pronunciar*, and *indictment/pronúncia*, but the legal knowledge evoked by these lexical units in English is not the same legal knowledge evoked by their equivalents in Portuguese.

There are still frames that do not present equivalence of any type. This is the case of `Arraignment` frame. `Arraignment` frame describes a legal event that is typical of the American system which is based on Common Law. Even the frame-evoking lexical units do not find an equivalent in Portuguese: *arraign* and *arraignment*.

7. Conclusion: Consequences of Frame Mismatches for Ontology Development

The conception of semantic frame as a language-independent conceptual level motivates proposals to use semantic frames as interlingual representation for multilingual databases (Boas, 2005). Other proposal for semantic frames is automatic generation of frame-based lexical databases (Padó, 2007). Semantic frames are still used for corpus annotation in languages other than English (Burchardt et al., 2009). Considering the use of FrameNet frames for lexical resource development, it is important to ask whether FrameNet could be a starting point even for ontology development. The assumption that semantic frames are conceptual structures language-independent could enable the use of semantic frames as ontological categories.

The contrastive study of `Criminal_process` frame brought important evidences against the use of FrameNet frames for legal ontology development. Semantic frames may represent a more language-independent conceptual level, but being language-independent does not mean being social or cultural-independent. Law is a social-oriented area, which means that laws are not equal for all countries. The frame evoked by a lexical unit will reflect the legal knowledge of a regulation from a specific country. In countries that speak the same language, a lexical unit may evoke a different legal knowledge in both countries.

The levels of mismatches between frames in different languages show that semantic frames cannot be considered a conceptual level language and cultural-independent. An example of semantic frame that does not exist in Brazilian legal system is the `Arraignment` frame. The arraignment is a hearing in which a suspect is asked to entering a plea. Therefore, the lexical units that evoke `Arraignment` frame in English do no find a translation equivalent in Portuguese. In other cases, some parts of a frame do not find a correspondence in Brazilian criminal process. This is the case of `Trial` frame. `Trial` frame is divided in three subframes: `Court_examination`, `Jury_deliberation`, and `Verdict`. According to the Brazilian legal system, the `Jury_deliberation` frame is not a step in a trial.

FrameNet semantic frames are not a good source of concepts for a legal ontology. Legal frames are social and culturally oriented. Using FrameNet frames would mean adopting a conception of legal system based in the United States legal organization. Different from other works in development of lexical resources, the social character of legal frames make them being specific for each society. Using FrameNet frames to develop an ontology to cover the concepts of the Brazilian legal system would entail the problem of adapting frames created for American legal system. A possible solution until this moment would be the manual modeling of a Brazilian legal ontology.

References

Benjamins, V. R., Casanovas, P., Breuker, J. and Gangemi, A. (2005). Law and the Semantic Web, an Introduction. *In*: Benjamins, V. R., Casanovas, P., Breuker, J. and Gangemi, A. (Eds.). *Law and the Semantic Web*: Legal ontologies, methodologies, legal information retrieval, and applications, LNAI (3369). Berlin/Heidelberg: Springer-Verlag, p.1-17.

Boas, H. C. (2005). Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. *International Journal of Lexicography.* Vol. 18, No.4, p. 445-478.

Breuker, J. A., Valente, A., and Winkels, R. (2005). Use and Reuse of Ontologies in Knowledge Engineering and Information Management. *In*: Benjamins, V. R., Casanovas, P., Breuker, J. and Gangemi, A. (Eds.). *Law and the Semantic Web:* legal ontologies, methodologies, legal information retrieval, and applications. Lecture Notes on Artificial Intelligence 3369. Springer-Verlag, Berlin/Heidelberg, p.36-64.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2009). Using FrameNet for the semantic analysis of German: annotation, representation, and automation. In: Boas, H. (Ed.) *Multilingual FrameNets* - Methods and Applications. Berlin/New York: Mouton de Gruyter, p.209-244.

Curtoni, P., Dini, L., Di Tomaso, V., Mommers, L., Peters, W., Quaresma, P., Schweighofer, E. and Tiscornia, D. (2005). Semantic access to multilingual legal information. *In*: Schweighofer, E. (Ed.). Proceedings of EU Info Workshop "Free EU Information on the Web: the future beyond the new EUR-LEX" of JURIX 2005 – The 18 th Annual Conference on Legal Knowledge and Information Systems. Brussels: Vrije Universiteit, pp.1-11. Available at: www.di.uevora.pt/~pq/papers/eu-ws-lois.pdf.

Fillmore, C.J., and Baker, C. (2010). A Frames Approach to Semantic Analysis. *In*: B. Heine, B. and Narrog, H. *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, p. 313-339.

Gangemi, A., Sagri, M. T., and Tiscornia, D. (2005). A Constructive Framework for Legal Ontologies. *In*: Benjamins, V. R., Casanovas, P., Breuker, J. and Gangemi, A. (Eds.) *Law and the Semantic Web:* legal ontologies, methodologies, legal information retrieval, and applications. Lecture Notes on Artificial Intelligence 3369. Springer-Verlag, Berlin/Heidelberg, p.97-124.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*. v.5, n.2, p.199-220.

Hirst, G. (2004). Ontology and the lexicon. *In*: Staab, S.; Studer, R. (Eds). *Handbook on Ontologies*. Springer: Berlin, 2004, p. 209-229.

Lönneker-Rodman, B. *Multilinguality and FramNet*. Technical Report. TR-07-001. Berkeley: ICSI, 2007.

Ohara, K. H. (2009). Frame-based contrastive lexical semantics in Japanese FrameNet: The case of risk and kakeru. In: Boas, H. C. (Ed.) *Multilingual FrameNets in computational lexicography:* Methods and applications. Berlin/New York: Mouton de Gruyter, p.163-182.

Padó, S. (2007). *Cross-lingual Annotation Projection Models for Role-Semantic Information*. PhD Thesis. Saarbrücken: Universität des Saarlandes.

Padó, S. and Lapata, M. (2005). Cross-lingual projection of role-semantic information. In: *Proceedings of HLT/EMNLP-05*, Vancouver: Association for Computational Linguistics, 2005, p.859-866.

Sagri, M. T., Tiscornia, D. and Bertagna, F. (2003). Jur-WorNet. In: P. Sojka et al. (Eds.) Second International Wordnet Conference - GWC 2004. Brno: Masaryk University, p.305-310.

Subirats, C. (2009). Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In: Boas, H. C. (Ed.) *Multilingual FrameNets in computational lexicography:* Methods and applications. Berlin/New York: Mouton de Gruyter, p.136-162.

Vossen, P. (1998). Introduction to EuroWordNet. *Computers and the Humanities*. Vol. 32, Nº. 2-3, p.73-89.