

# Adding a new language version to a lexical resource. Is it possible to assign term equivalents semi-automatically?<sup>1</sup>

**JANINE PIMENTEL**

Pontifícia Universidade Católica  
do Rio de Janeiro (PUC-Rio)  
PNPD/CAPES - 2013  
janine.m.pimentel@gmail.com

## Janine Pimentel és màster en

Lingüística (Lexicografia i Terminologia) per la Universitat Nova de Lisboa (Portugal) i doctora en Traducció per la Universitat de Montreal (Canadà). En aquests moments és investigadora postdoctoral a la Universitat Catòlica Pontifícia-Rio de Janeiro (Brasil). Ha fet recerca en teoria i metodologia de la lingüística de corpus i en semàntica d'esquemes per elaborar un diccionari en línia de terminologia jurídica en tres llengües (portuguès, anglès i francès).



## Resum

### **Adició d'una nova llengua a un recurs lèxic. És possible assignar termes equivalents de manera semiautomàtica?**

L'article descriu el procés d'afegir la versió francesa a un recurs lèxic sobre terminologia jurídica. El recurs JuriDiCo pren com a base teòrica la semàntica d'esquemes i la metodologia desenvolupada pels lexicògrafs de FrameNet. En aquest treball s'analitza si els criteris proposats en treballs anteriors, que assignaven equivalents de manera semiautomàtica entre termes anglesos i portuguesos, es poden aplicar a altres parelles de llengües (francès-anglès i francès-portuguès).

PARAULES CLAU: recursos lèxics; equivalents; francès; portuguès; anglès

## Abstract

The paper describes the process of adding a French version to a lexical resource that describes legal terminology. The resource, called JuriDiCo, is based on the theory of Frame Semantics as well as on the methodology developed by FrameNet lexicographers. We test if the criteria proposed in previous work for semi-automatically assigning the equivalents of English and Portuguese terms apply to other language pairs (French-English and French-Portuguese).

KEYWORDS: lexical resource; equivalents; French; Portuguese; English

TERMINÀLIA 11 (2015): 20-29 · DOI: 10.2436/20.2503.01.74  
Data de recepció: 30/09/2014. Data d'acceptació: 5/11/2014  
ISSN: 2013-6692 (impresa); 2013-6706 (electrònica) · <http://terminalia.iec.cat>

## 1 Introduction

The creation of multilingual terminological resources, in general, and the creation of multilingual lexical resources covering the specialized field of the law, in particular, raises the problem of linking potentially anisomorphic terminologies. For this reason, the multilingual resources that describe legal terminology have not always included all the equivalents that translators need to produce translations of legal texts (Groot and Laer, 2008). What is more, the inclusion of new languages in existing term banks and other multilingual lexical resources can raise methodological issues which terminologists may not have anticipated when they first created the resource that they are developing. Although the literature in terminology and in lexicography has shown that the assignment of equivalents requires that the degree and nature of the equivalence relation be established (full equivalence, partial equivalence, zero equivalence; one-to-one equivalents, one-to-many equivalents, etc.), little has been debated about the best methodologies to identify equivalents, especially in several language pairs. Suppose that a term X in language A is a full equivalent of the term Y in language B, that these pairs of equivalents were validated against a rigorous set of criteria and that they were included in a term bank. Suppose that a new version of the database is created and that the term Z in language C is a full equivalent of the term X in language A. Does that mean that the term Z is also a full equivalent of the term Y? What about partial equivalents? Does the same reasoning apply?

In order to bridge one of these gaps, a methodology for identifying terminological equivalents based on the theory of Frame Semantics (Fillmore, 1977, 1982, 1985; Fillmore and Atkins, 1992) as well as on the methodology developed by FrameNet lexicographers (Ruppenhofer et al., 2010) was proposed in previous work (Pimentel, 2012). The tested hypothesis was that frames, i.e. conceptual scenarios in which words participate, can function as interlingual representations grouping together equivalent terms. As a result, 200 Portuguese and English specialized verbs selected from a comparable corpus of judgments were grouped together into 76 frames in a bilingual lexical resource named *JuriDiCo* and 165 pairs of equivalents (71% full equivalents and 29% partial equivalents) were identified. A list of criteria for validating equivalents of this kind of units was also drawn (Pimentel, 2013). As *JuriDiCo* is a work in progress, the inclusion of other languages such as French has been recently considered. For this reason, it is relevant to test if the criteria proposed in Pimentel (2013) apply to other language pairs (French-English and French-Portuguese) and if, by adding a third language, the identification of equivalents can be accelerated by applying the reasoning described above. This could not only lead to a more

complete view of all equivalence scenarios but also to a semi-automatic treatment of equivalence relations.

This paper is organized as follows. Section 2 briefly defines what equivalence means for lexicographers and terminologists. Section 3 summarizes some of the criteria that have been discussed in the literature for assigning equivalents. Section 4 presents *JuriDiCo*, a bilingual lexical resource that describes legal terminology. Section 5 illustrates the process of creating *JuriDiCo*'s French version with the inclusion of 20 specialized verbs in the database. This section specifically concentrates on the task of assigning the equivalents between the French specialized verbs and the English and Portuguese ones. Section 6 presents the results of the research obtained up to date and section 7 answers the research questions raised here by drawing some conclusions based on the interpretation of the results of the research.

## 2 Equivalence

So, what is equivalence? Adamska-Sałaciak (2010, p. 387) reminds us that “to be able to talk about equivalence, there must be (at least) two entities of some kind, a certain relationship between those entities, and a certain value of that relationship”. But what are these “entities”? What kind of “relationship” exists between these entities? And what is meant by the “value of that relationship”? These are some of the questions that have been raised and discussed by lexicographers, such as Hartmann (2007 [1990]), Werner (1999), Wiegand (2005) and recently Adamska-Sałaciak (2010). For them, while studying equivalence, the first important aspect that needs to be examined is that of the levels at which equivalence can be formulated. Using the Saussurean distinction between *langue* and *parole*, generally, in translation studies, equivalence is seen as a phenomenon belonging to the level of *parole* because it has to do with instantiations of language in texts. However, in lexicography, as well as in terminology, it is more difficult to say whether equivalence is a systemic or intertextual phenomenon because equivalence is a relationship that exists between units (lexical or terminological) that are given in a lexicographic or terminological product. As Adamska-Sałaciak (2010, p. 388) notes, with the use of corpora lexicographers have come to apply both types of equivalence, “the intertextual type appearing in those instances where the source language (henceforth, SL) unit to be provided with an equivalent is larger than a single word”. Wiegand (2002, p. 245) takes a more radical position by arguing that “the *langue*-related concept of equivalence of contrastive lexicology is inappropriate for bilingual lexicography, because bilingual dictionaries are not conceptualized as aids for contrastive studies of language systems”. He explains that dictionaries are rather meant, in the first place, as a means to under-

stand and produce foreign texts and to make translations in both directions.

In terminology, equivalence is generally formulated at the “conceptual” level, i.e. two terms are equivalents if their conceptual properties coincide against the background of a specialized field. This raises the question whether concepts are independent from language. If concepts are consolidated at the level of *langue* and not of *parole*, as Wüster’s four-word model (1959) suggests, then equivalence can be said to be equated at the systemic level. However, as it has been mentioned in Rogers (2007), terminologists are increasingly using running texts to extract lexical data and when they move between text and system they do not necessarily find the same equivalents. This suggests that, as in lexicography, the intertextual type of equivalence is also taken into account in some kinds of terminology work.

Another important question concerns the entities between which there is a relationship of equivalence. Here, it is a matter of examining whether the phenomenon of equivalence is established between word senses, concepts or designations. For Adamska-Sałaciak (2010) the entities are word senses: “when we talk about a lexical item X in the SL being equivalent to a lexical item Y in the TL, what we mean is that X in a particular sense is equivalent to Y in a particular sense”. For Wiegand (2005, p. 21), the entities are “lexikalsemantische Einheiten”, a notion which roughly corresponds to that of “lexical units” (Cruse, 1986), i.e. words or expressions taken in one of their senses.

In terminology, equivalence is often viewed as a relation between concepts (Felber and Budin. 1989) but, contradictorily, the International Organization for Standardization (henceforth, ISO) defines equivalence as “the relation between designations in different languages representing the same concept” (ISO 1087-1, 2000, p. 30). We can formulate two hypotheses on why sometimes equivalence is defined as a relationship between concepts and some other times as a relationship between designations. Firstly, ISO (ISO 1087-1, 2000) identifies three types of relations between designations of concepts, i.e. synonymy, antonymy and equivalence. Each of these is defined as a kind of relation between two or more designations. Secondly, for a very long time partisans of the classical approach to terminology believed that designations should be ‘transparent’ and reflect the concept and the place the concept occupies in the conceptual system (Felber and Budin, 1989, p. 123). Thus, if designations mapped concepts, it would not make a difference to formulate equivalence either as a relation obtained between designations or as a relation established between concepts.

Having presented some definitions of equivalence for lexicographers and terminologists, it is relevant now to examine the methodological proposals that have been designed to assign equivalents.

### 3 Assignment of equivalents

Despite the relative consensus concerning the definition of equivalence, Hartmann’s statement, according to which “there is no consistent methodology [for assigning equivalents], the work is often patchy, partial, intuitive, arbitrary and tentative, and thus of limited generalisability” (2007[1985], p. 15), is still valid to a great extent. Nevertheless, some contributions, mostly stemming from lexicographers, are worth mentioning.

For Svensén (2009), the search for equivalents is similar to the search for suitable paraphrases in monolingual lexicography. The procedure consists in determining the headword’s content, for instance, based on monolingual dictionaries in the source language, and then in working towards the word or words available in the target language which best represent that content. He adds that “normally, there are from the very outset one or more equivalent candidates available in the sources on which the work is based, whether these consist of bilingual dictionaries (in digital form or not) or authentic material in the form of, for instance, bilingual corpora” (Svensén, 2009, p. 266). However, as Bergenholtz and Tarp (1995), who consider introspection an unreliable approach that should never stand alone and that should thus be combined with the use of corpora, we believe that Svensén’s view may be possible for general lexicography but is insufficient for the elaboration of specialized lexical resources, especially if the work is carried out by non-experts of the domain that the resource aims to cover. Similarly, Atkins and Rundell (2008) suggest using corpora to search for equivalents. They explain how a target language corpus and parallel corpora can be used to identify equivalents, and how to search concordances and use the information found to put translations in a database. Although they mention the use of comparable corpora, unfortunately they do not specify how the same can be done with this kind of corpus.

Nevertheless, according to Atkins and Rundell (2008) there are several factors that should be taken into account in the assignment of equivalents: semantic content, collocational context, vocabulary type, message and function. We will comment on the first four of these factors because as the authors explain: “the last [factor] is principally of interest when you’re looking for equivalents of grammatical items” (Atkins and Rundell, 2008, p. 468). “Semantic content” refers to what other linguists call denotation, reference and cognitive meaning. Therefore, semantic content designates “the ‘literal’ meaning of an expression together with its ‘connotation’ or any figurative meaning that may be associated with it” (Atkins and Rundell, 2008, p. 469). Usually, two words denoting the same object such as *tiger* in English and *tigre* in French form an exact match of semantic content. “Collocational context” refers to the collocates in the SL that may produce dif-



ferent translations in the TL. For instance, when *bunch* collocates with *flowers* its French equivalent is *bouquet*, but when it collocates with *hair* its French equivalent is *touffe* or *houppes*. Other authors, such as Hartmann (2007[1985]) had also referred to the importance of taking collocates into account when choosing the right equivalents (see the example he provides of *kochen/cook*). When it comes to “vocabulary type”, Atkins and Rundell explain that it refers to register, style, attitude, etc., that can cause SL-TL mismatches. For instance, the equivalent of an informal expression should be an informal expression as well. “Message” refers to those cases in which the literal meaning of a phrase is different from its underlying meaning or “pragmatic force”. For example, the French equivalent of the English expression *birds of a feather flock together* is *qui se ressemblent s’assemblent*. From our point of view, these two last factors may be less relevant in terminology.

Another important contribution comes from Piotrowski (1994). According to him, the feature according to which the relationship of equivalence should be established needs to be equated as a third external entity or “tertium comparationis”. He explains that “in order to be able to compare two entities it is essential to have a third one against which both could be described, evaluated, etc.” (Piotrowski, 1994, p. 128). The notion of “tertium comparationis”, issued from contrastive linguistics, should be external to both entities and it should also be something practical. Briefly, according to the author, referents or concepts as described in the semiotic triangle are not appropriate “tertium comparationis” principles because “there are no pure references, i.e. the act of distinguishing a referent depends to a large degree on the relevant language” and concepts “are not suitably external to any language” (Piotrowski, 1994, p. 129 citing Haas, 1962/1968). Neither is meaning appropriate because he rightly argues there is little theoretical agreement on what meaning is. Instead, the “tertium comparationis” should consist of two dimensions: a “situational dimension” and a “formal dimension”. The “situational dimension” corresponds to the discursive or contextual dimension, whereas the “formal dimension” corresponds to the collocational patterns of lexical items to which we have already referred. In this respect, Piotrowski notes that only the most typical frequent collocability patterns, or “focal collocations”, should be taken into account.

Considering all the methodological aspects that have been reviewed so far, it seems that the assignment of equivalents of terms that occur in specialized texts could be based on: 1) corpora; 2) the analysis of the semantic content / conceptual characteristics of terms; 3) the analysis of the collocational context of terms; 4) encyclopedic knowledge; 5) a tertium comparationis or entity external to terms. In the following section we illustrate the implementation of a methodology to describe terms and assign their equiva-

lents in a resource called JuriDiCo. The methodology is believed to apply the aforementioned aspects either directly or indirectly in one way or another. For a more detailed description on the application of these principles the reader is referred to Pimentel (2013).

#### 4 JuriDiCo: a lexical resource that describes legal terminology

One of the objectives of the research carried out to compile the JuriDiCo was to test the possibility of using the theoretical and methodological frameworks of Frame Semantics (Fillmore, 1977, 1982, 1985; Fillmore and Atkins, 1992) and FrameNet (Ruppenhofer et al., 2010) to assign the equivalents of a specific kind of terminological units, i.e. the specialized verbs that occur in legal texts. The hypothesis was that a methodology for the description of terms based on a theory that links linguistic semantics to encyclopaedic knowledge, such as Frame Semantics, and that offers a reasonably language-independent apparatus (Boas, 2005; Baker, 2009), such as semantic frames, could be suitable for identifying the equivalents of terms. This theory has proven its lexicographic relevance in the FrameNet project (<https://framenet.icsi.berkeley.edu/fndrupal/>) and it has also been used to describe terminologies (e.g. Dolbey et al., 2006; Faber et al., 2006; Schmidt, 2009). The descriptions of the terms as well as their equivalents were encoded in a database that functions as an observatory of the equivalence phenomenon (<http://olst.ling.umontreal.ca/juridico>).

JuriDiCo aims at serving a decoding and an encoding function, i.e. it should be helpful for understanding legal terminologies as well as for the translation and production of legal texts, in that it allows users to search the information they need by introducing the knowledge they already have about a concept or by introducing a term so that they learn about its characteristics. This way, JuriDiCo provides users with rich detail about the grammatical and phraseological profile of terms (actantial or argument structure, combinatorial properties) as well as extralinguistic information about them (frames).

In this lexical resource, each term is said to “evoke” a frame. Frames are conceptual scenarios that contain mandatory meaning slots or core Frame Elements (hereafter FEs) as well as optional slots or non-core FEs. For instance, the Argumentation frame corresponds to a scenario in which parties in a case are called upon to explain why they are accusing someone, being accused by someone, appealing a decision, etc. It therefore includes two core FEs: Arguer and Argument. The FE Arguer can refer to the accuser or to the defendant (criminal suit), to the petitioner or to the respondent (civil suit), whereas the FE Argument refers to a statement, reason or fact for or against the matter under discussion. The terms *argue*, *assert*, *con-*

*tend<sub>i</sub>*, *invoke<sub>i</sub>*, *state<sub>i</sub>* and *submit<sub>i</sub>* all evoke the Argumentation frame. However, their linguistic behavior is not exactly the same: *argue<sub>i</sub>*, *assert<sub>i</sub>* and *contend<sub>i</sub>* are typically followed by that-clauses (*The respondents contend that the order being sought should not be granted*), whereas *invoke<sub>i</sub>* is typically followed by a noun phrase (*The respondents invoke the Business Judgment Rule*).

Frames can group together certain synonyms, near-synonyms, opposites and related terms within a given language. This is the way in which JuriDiCo accounts for some semantic relations between terms. Frames also function as interlingual representations grouping together equivalent terms. The descriptions of the terms and frames as well as the information on the equivalence status of the terms are encoded in an xml editor, stored in a server and published in a webpage called JuriDiCo. The database can be consulted in three ways: users can browse an alphabetically ordered list of English terms and an alphabetically ordered list of Portuguese terms; they can use a search engine; or browse an alphabetically ordered list of frames. The following section describes the way in which the French version of JuriDiCo was implemented.

## 5 Adding a third language

The creation of JuriDiCo's French version consisted of four main phases: the adjustment of the database; the constitution of a specialized French corpus; the selection and analysis of terms and frames, and the assignment of equivalents.

### 5.1 The database

JuriDiCo's model of encoding data is strongly inspired by the one used to compile DiCoInfo (L'Homme, 2008) which is a multilingual lexical resource (French, English, Spanish) that describes the terminology of computing and the Internet. The first step in the creation of JuriDiCo's French version corresponded to the addition,

in JuriDiCo's database repository, of a folder containing the same encoding model that was used to compile the English and Portuguese versions. No changes in the structure of the xml template were made.

### 5.2 The corpus

The English and Portuguese versions of JuriDiCo are based on data extracted from a comparable corpus of authentic judgments produced by the Supreme Court of Canada and by the Supremo Tribunal de Justiça de Portugal. They were first created three years ago in Canada as part of our doctoral dissertation. Back then, Canadian English was chosen because documentation on legal language as well as legal experts stemming from this linguistic community were more accessible. For this reason, the following step in the methodology to add a French version to the database consisted in the compilation of a Canadian French subcorpus that parallels with the Canadian English one. However, as JuriDiCo's strategy for development foresees the inclusion of other genres of legal texts, it was decided that the French corpus should be more varied and include judgments not only produced by the Supreme Court of Canada but also by other Canadian courts, such as appellate courts of the provinces and territories. The French corpus, similarly to the Portuguese and English ones, totals over one million word forms. More recently, a Brazilian corpus of judgments has been added to the Portuguese version of the comparable corpus so as to study the differences between legal terminology used in Portugal and in Brazil (Pimentel, 2014). Similar studies on the differences between Canadian French and other geographic variations of French are also envisioned.

### 5.3 Terms and frames

The selection of the specialized verbs, the identification of the frames (Appendix I) and their description were entirely based on the methodology proposed in Pimen-

|   |  |                    |   |
|---|--|--------------------|---|
| 1 | ... sérieux et fouillé, les défendeurs | <b>invoquent</b>   | plusieurs moyens de fait et de droit à l'appui de ...   |
| 2 | ... Les appelants                      | <b>invoquent</b>   | à l'appui de leur prétention l'art. 406 C.c.B.-C ...    |
| 3 | ... La défenderesse                    | <b>invoquait</b>   | sa loi habilitante, adoptée par la législature ...      |
| 4 | ... agir empêche les requérants d'     | <b>invoquer</b>    | les droits de tiers dans des réclamations ...           |
| 5 | ... mise en demeure, les demandeurs    | <b>invoquaient</b> | des illégalités dans le processus d'évaluation ...      |
| 6 | ... du temps — ne pouvaient jamais     | <b>invoquer</b>    | la fraude viciant le consentement aux relations ...     |
| 7 | ... version générique du Viagra. Elle  | <b>a invoqué</b>   | l'invalidité du brevet de Pfizer en raison du ...       |
| 8 | ... la requête introductive du Comité  | <b>invoque</b>     | à la fois la nullité du décret et de l'autorisation ... |

FIGURE 1. Concordances of the verb *invoquer*

tel (2013), to which the reader is referred for more details. This is a bottom-up approach, in which verbs are first selected, then analyzed by means of a concordance program and finally grouped into frames.

For instance, the candidate-term *invoquer* has two different specialized meanings (figure 1): in concordances [1-4], *invoquer*<sub>1</sub> is associated with the [Argumentation] frame, whereas in concordances [5-8] *invoquer*<sub>2</sub> is associated with the [Contesting] frame. Although they both have the same number of actants (also called arguments), the linguistic realizations of the actant which corresponds to the syntactic object of the verb have a different semantic nature: *illégalité* [5], *fraude* [6] and *invalidité* [7] have a more negative connotation than *moyen* [1], *article* [2] and *loi* [3]. Therefore, the actantial structure of *invoquer*<sub>1</sub> is Arguer ~ Argument whereas the actantial structure of *invoquer*<sub>2</sub> is Arguer ~ Irregularity. Also, in this stage, the labels attributed to each semantic frame (e.g. [Argumentation], [Contesting]) were used to group together part of the French terms, namely the synonyms *alléguer*<sub>1</sub> and *invoquer*<sub>1</sub> as well as the synonyms *alléguer*<sub>2</sub> and *invoquer*<sub>2</sub>.

#### 5.4 Assignment of equivalents

Once the description of the French terms was completed, the frame labels that had been attributed to each specialized meaning were used to link the French terms with the English and Portuguese ones (Appendix I). This was done automatically by querying the database. Then, each pair of equivalents was studied closely and the criteria proposed in Pimentel (2012) were applied.

The first criterion consists in examining the way the terms evoke a frame. There are some cases in which the verbs in one language correspond to the opposites of the verbs in the other language, this meaning that they profile the frame in a negative or positive way. For instance, the French term *conformer*<sub>1</sub> cannot be considered an equivalent of *violate*<sub>2</sub>, because the former profiles the frame [Compliance] in a positive way, whereas the latter profiles it in a negative way. If the verbs infringed this criterion, they were considered zero equivalents, but if they met it, the following criteria were applied.

The second criterion observes how many FEs are profiled by the verbs. This corresponds to counting the number of actants of the verbs. For example, the three core FEs Arguer, Irregularity and Evidence of the frame [Contesting] are always instantiated in the contexts of *rebut*<sub>1</sub> but only Arguer and Irregularity are instantiated in the contexts of *invoquer*<sub>2</sub>.

The third criterion compares the linguistic realizations of the FEs, so as to confirm or refute whether the nature of the actants of the verbs is entirely identical. Differences in the linguistic realizations of the FEs can be grouped into three categories. In the first one, metonymies are observed. Certain verbs allow for the metonymy of one of the actants. For instance, the frame [Compliance] contains two slots which corre-

spond to three core FEs: the core FE Protagonist and the core FE Act (the Protagonist's behavior) correspond to one of the slots; Law corresponds to another. Whereas *conformer*<sub>1</sub> always profiles the core FEs Protagonist and Law, *comply*<sub>1</sub> profiles all three. The second difference between the realizations of the actants is related to a semantic preference for certain types of collocates. For instance, the frame [Regulations] contains two slots: the core FE Law and the core FE Issue. Although *réglementer*<sub>1</sub> and *codify*<sub>1</sub> both profile these two core FEs, the linguistic realizations of the core FE Law in the contexts of *codify*<sub>1</sub> denote entities with a stronger legal power than those in the contexts of *réglementer*<sub>1</sub>. The third difference in the linguistic realizations of the FEs is a matter of semantic prosody. For example, although *alléguer*<sub>1</sub> and *impugn*<sub>1</sub> both profile the two core FEs of the [Contesting] frame, the linguistic realizations of the core FE Irregularity in the contexts of *alléguer*<sub>1</sub> denote entities with a more negative connotation than those in the contexts of *impugn*<sub>1</sub>.

Finally, the valence patterns of the verbs were compared. The terms *ordonner*<sub>1</sub> and *determinar*<sub>2</sub> evoke the frame [Order] which involves four core FEs: JUDGE, LAW, DUTY and PROTAGONIST. Whereas *ordonner*<sub>1</sub> admits only one valence pattern, its Portuguese candidate equivalent admits three:

*ordonner*<sub>1</sub>:

*Le premier juge a fait droit à la requête en injonction interlocutoire et ordonné à Gestion Lafrenière de cesser tout rejet dans le lac Heney.*

JUDGE / LAW (Subject, Noun Phrase) PROTAGONIST (Complement, Prepositional Phrase) DUTY (Complement, that clause)

*determinar*<sub>2</sub>:

*Entretanto, em 26 de Fevereiro de 2004, no âmbito de uma providência cautelar que instaurou contra a Ré, o Senhor Juiz determinou à Ré que: Atribuisse ao Autor reais tarefas no âmbito da sua categoria profissional de Primeiro Assistente de Direcção.*

JUDGE (Subject, Noun Phrase) PROTAGONIST (Complement, Prepositional Phrase) DUTY (Complement, that clause)

*O Sr. Juiz determinou que se abrisse vista ao M.<sup>o</sup>P.<sup>o</sup> para efeitos de suscitação de presente conflito.*

JUDGE (Subject, Noun Phrase) DUTY (Complement, that clause)

*O Tribunal determinou uma alteração da qualificação jurídica dos factos apontando até na direcção do homicídio qualificado.*

JUDGE (Subject, Noun Phrase) DUTY (Object, Noun Phrase)

As mentioned above, if the verbs infringed the first criterion, they were immediately considered zero



equivalents. Whenever the verbs infringed any of the criteria 2-4, the candidate equivalents were considered partial equivalents. If they met all criteria 1-4, they were considered full equivalents.

## 6 Results

We grouped together about 100 terms (20 of which were French, 47 were English and 30 were Portuguese) into 15 frames and identified 108 pairs of equivalents (65 pairs of French-English verbs and 43 pairs of French-Portuguese verbs). The number of terms that were added to JuriDiCo corresponds to an increase of 10% of the total number of terms described in (Pimentel, 2012). We did not identify any new frame. However, the number of equivalents that were included in the database corresponds to an increase of about 65% of the total number of JuriDiCo's equivalents. These can be divided into the following categories.

### 6.1 Zero equivalents

10% of the equivalent pairs were attributed the status of zero equivalence because they do not meet the criterion 1 described in section 5.4. For instance, *conformer<sub>1</sub>* profiles the [Compliance] frame in a positive way, whereas *exceed<sub>1</sub>* profiles it in a negative way. The same occurs in the French-Portuguese pair *conformer<sub>1</sub>* - *violar<sub>2</sub>*. *Conformer<sub>1</sub>* profiles the [Compliance] frame in a positive way, whereas *violar<sub>2</sub>* profiles it in a negative way. Other examples of a parallelism between French-English and French-Portuguese candidate equivalents that were considered “zero equivalents” are: *contrevenir<sub>1</sub>* - *comply<sub>1</sub>*, *contrevenir<sub>1</sub>* - *violar<sub>1</sub>*, *contrevenir<sub>2</sub>* - *comply<sub>2</sub>*, *contrevenir<sub>2</sub>* - *cumprir<sub>1</sub>*.

### 6.2 Partial equivalents

32% of the equivalent pairs were considered partial equivalents because they do not meet the criteria 2, 3 or 4 described in section 5.4. 5% of the equivalent pairs are partial equivalents due to differences in the actantial structures of the verbs, i.e. one (or several) core FEs are never instantiated in the contexts of one of the verbs, whereas they are very often instantiated in the contexts of the equivalent verb. This case was only observed in the French-English verb pairs (*alléguer<sub>1</sub>* - *rebut<sub>1</sub>*, *déterminer<sub>2</sub>* - *infer<sub>1</sub>*, *invoquer<sub>2</sub>* - *rebut<sub>1</sub>*, *procéder<sub>1a</sub>* - *proceed<sub>2a</sub>*, *procéder<sub>1b</sub>* - *proceed<sub>2b</sub>*) and never in the French-Portuguese ones. For instance, consider the contexts below that were taken from the comparable corpus:

C'est précisément à cette sorte d'examen factuel et juridique que les tribunaux **procèdent** quotidiennement.

The Crown can **proceed** summarily or by indictment.

O Tribunal da Relação **procedeu** à reapreciação das provas gravadas.

*Procéder<sub>1a</sub>* cannot be easily replaced by *proceed<sub>2a</sub>* and vice-versa because the linguistic behavior of the terms is different. If one translated the French context into English by using the verb *proceed*, one would have to omit an actant; and if one translated the English context into French, one would have to create an actant that was not expressed in the source context. Interestingly, however, *procéder<sub>1a</sub>* and *proceder<sub>2</sub>* are full equivalents, this meaning that the French and Portuguese terms are partial equivalents of the English *proceed<sub>2a</sub>*.

22% of the equivalent pairs display alternations in the semantic nature of a core FE. 6% of the differences are due to metonymies; 14% to semantic preference and 2% to semantic prosody. Most French-English pairs that received partial equivalence status for these reasons have French-Portuguese counterparts, i.e. if the French term is a partial equivalent of a given English term and this English term is a partial equivalent of a given Portuguese term, then the French term is also a partial equivalent of the Portuguese. The exceptions to this are *appliquer<sub>2</sub>* - *apply<sub>2</sub>* and *appliquer<sub>2</sub>* - *aplicar<sub>2</sub>*, *conformer<sub>1</sub>* - *comply<sub>2</sub>*. For instance, consider the contexts below:

While the Act applies mainly to public sector employers, it also **applies** to some private sector employers.

In McGuigan, the Court held that s. 21 of the Criminal Code **applies** to s. 85(1) (pp. 307-8).

La règle 39 des Règles de procédure **s'applique** à la preuve qui peut être utilisée sur une motion ou une requête.

A todos os contratos celebrados no âmbito das presentes condições **aplica-se** o Direito Holandês.

Whereas the English verb can be followed by terms that denote either somebody (Protagonist, e.g. *employers*) or a situation (Case, e.g. *section 85*), the French and Portuguese verbs are always followed by terms that denote a situation (Case, e.g. *preuve*, *Direito*).

Finally, 5% of the equivalent pairs display differences in their valence patterns. This is the case of the following verbs: *conférer<sub>1</sub>* - *conferir<sub>1</sub>*, *conférer<sub>1</sub>* - *grant<sub>1</sub>*, *conférer<sub>1</sub>* - *outorgar<sub>2</sub>*, *ordonner<sub>1</sub>* - *determinar<sub>2</sub>*, *ordonner<sub>1</sub>* - *order<sub>1</sub>*.

### 6.3 Full equivalents

58% of the equivalent pairs were considered full equivalents because they meet all the criteria described in section 5.4.

## 7 Conclusion

The paper described the process of adding a French version to JuriDiCo, a bilingual (English-Portuguese)

lexical resource describing legal terminology. It tested if the criteria to validate Portuguese-English candidate equivalents proposed in Pimentel (2013) applied to other language pairs (French-English and French-Portuguese) and if, by adding a third language, conclusions on the automatization of the task of assigning equivalents could be drawn. Firstly, the study revealed that the criteria applied were useful for the identification of candidate equivalents in a different language pair. For the first time, zero equivalents were counted and explanations for this equivalence degree were provided. Compared to previous work, the number of full equivalents that were identified for the French-English and French-Portuguese language pairs was lower than the number of full equi-

valents in the Portuguese-English language pair. Secondly, differences between the French-English and French-Portuguese scenarios of equivalence were identified. Although candidate equivalents in the French-English language pair are more often invalid (zero equivalents) than French-Portuguese ones, the amount of partial equivalents is similar. Finally, when a French term is a partial equivalent of a given English term and this English term is a partial equivalent of a given Portuguese term, then the French term is not necessarily a partial equivalent of the Portuguese term. For these reasons, the results of this study argue against the full automatization of the process of assigning equivalents. ✿

## References

- ADAMSKA-SAŁACIAK, Arleta (2010). "Examining Equivalence". *International Journal of Lexicography*, 23 (4), p. 387-409.
- ATKINS, Sue; RUNDELL, Michael (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- BAKER, Collin F. (2009). "La sémantique des cadres et le projet FrameNet: une approche différente de la notion de 'valence'". *Langages*, (4), p. 32-49.
- BERGENHOLTZ, Henning; TARP, Sven (1995). *Manual of Specialized Lexicography*. Amsterdam; Philadelphia: John Benjamins.
- BOAS, Hans (2005). "Semantic Frames as Interlingual Representations for Multilingual Lexical Databases". *International Journal of Lexicography*, 18 (4), p. 39-65.
- CRUSE, David Alan (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- DOLBEY, Andrew; ELLSWORTH, Michael; SCHEFFCZYK, Jan (2006). "BioFrameNet: A Domain-specific FrameNet Extension with Links to Biomedical Ontologies". In: BODENREIDER, Olivier (ed.). *Proceedings of KR-MED*. Baltimore, p. 87-94.
- FABER, Pamela; MONTERO MARTÍNEZ, Silvia; CASTRO PRIETO, María Rosa; SENSO RUIZ, José Antonio; PRIETO VELASCO, Juan Antonio; LEÓN ARAUZ, Pilar; MÁRQUEZ LINARES, Carlos; VEGA EXPÓSITO, Miguel (2006). "Process-Oriented Terminology Management in the Domain of Coastal Engineering". *Terminology*, 12 (2), p. 189-213.
- FELBER, Helmut; BUDIN, Gerhard (1989). *Terminologie in Theorie und Praxis*. Forum für Fachsprachenforschung. Tübingen: Günter Narr Verlag.
- FILLMORE, Charles (1977). "Scenes-and-Frames Semantics, Linguistic Structures Processing". In: ZAMPOLLI, Antonio (ed.). *Fundamental Studies in Computer Science*, 59, Amsterdam: North Holland Publishing, p. 55-88.
- FILLMORE, Charles (1982). "Frame Semantics". In: THE LINGUISTIC SOCIETY OF KOREA (ed.). *Linguistics in the Morning Calm*. Seoul: Hanshin, p. 111-137.
- FILLMORE, Charles (1985). "Frames and the Semantics of Understanding". *Quaderni di Semantica*, 6 (2), p. 222-254.
- FILLMORE, Charles; ATKINS, Sue (1992). "Towards a Frame-Based Lexicon: The Semantics of RISK and its Neighbors". In: LEHRER, Adrienne; KITTAY, Eva (ed.). *Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organization*. Hillsdale: Lawrence Erlbaum Associates, p. 75-102.
- GROOT, Gerard-René de; LAER, Conrad van (2008). "The Quality of Legal Dictionaries: an Assessment". [online] <<http://papers.ssrn.com/sol3/papers.cfm?abstractid=1287603>; <http://arno.unimaas.nl/show.cgi?fid=13383>>. [Last consulted: 20 September 2014]. [Maastricht Faculty of Law Working Paper No. 2008/6, 2008]
- HAAS, William (1962/1968). "The Theory of Translation". *Philosophy*, 37, p. 2208-228. [Reprinted in PARKINSON, G. (ed.). *The Theory of Meaning*. Oxford: Oxford University Press]



- HARTMANN, Rudolf K. (2007) [1985]. “Contrastive Text Analysis and the Search for Equivalence in the Bilingual Dictionary”. In: HYLDGAARD-JENSEN, Karl; ZETTERSTEN, Arne (ed.) *Symposium on Lexicography II*. Tübingen: Max Niemeyer, p. 121–132. [Reprinted in HARTMANN, Rudolf K. (2007). *Interlingual Lexicography Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary*. Tübingen: Max Niemeyer, p. 15–23]
- HARTMANN, Rudolf K. (2007) [1990]. “The not so harmless drudgery of finding translation equivalents”. In: ROSE, Marilyn (ed.). *Language and Communication. Special Issue on Translation*, 10 (1), p. 47–55. [Reprinted in HARTMANN, Rudolf K. (2007). *Interlingual Lexicography Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary*. Tübingen: Max Niemeyer, p. 30–37]
- ISO 1087-1 (2000). *Terminology work – Vocabulary – Part 1: Theory and application*.
- L’HOMME, Marie-Claude (2008). “Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés”. *Traduire*, 217, p. 78–103.
- PIMENTEL, Janine (2012). “Criteria for the Validation of Specialized Verb Equivalents: Applications in Bilingual Terminography”. Doctoral dissertation in Translation Studies, University of Montreal, Montreal.
- PIMENTEL, Janine (2013). “Methodological Bases for Assigning Equivalents: a Contribution”. *Terminology*, 19 (2), p. 237–257.
- PIMENTEL, Janine (2014). “First steps towards managing diatopic variation in a term base: the case of European and Brazilian Portuguese”. *Turjuman: Journal of Translation Studies*, 23 (2), p. 46–74.
- PIOTROWSKI, Tadeusz (1994). *Problems in Bilingual Lexicography*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- ROGERS, Margaret (2007). “Terminological Equivalence in Technical Translation: A problematic Concept? St. Jerome and Technical Translation”. *SYNAPS: Fagspråk, Kommunikasjon, Kulturkunnskap*, 20, p. 13–25.
- RUPPENHOFER, Josef; ELLSWORTH, Michael; PETRUCK, Miriam; JOHNSON, Christopher; SCHEFFCZYK, Jan (2010). “FrameNet II: Extended Theory and Practice. ICSI Technical Report”. [online]. <<http://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>>. [Last consulted: 20 September 2014].
- SCHMIDT, Thomas (2009). “The Kicktionary – A Multilingual Lexical Resource of Football Language”. In: BOAS, Hans C. (ed.). *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Berlin; New York: Mouton de Gruyter, p. 101–134.
- SVENSÉN, Bo (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.
- WERNER, Reinhold (1999). “Das Problem der Äquivalenz in zwei- und im mehrsprachigen Fachwörterbüchern”. In: HOFFMANN, Lothar; KALVERKÄMPER, Hartwig; WIEGAND, Herbert (ed.). *Fachsprachen: ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft = Languages for special purposes: an international handbook of special-language and terminology research*. Berlin; New York: W. de Gruyter, p. 1853–1884.
- WIEGAND, Herbert (2002). “Equivalence in Bilingual Lexicography: Criticism and Suggestions”. *Lexikos*, 12, p. 239–255. [AFRILEX-reeks/series 12: 2002]
- WIEGAND, Herbert (2005). “Äquivalenz, Äquivalentdifferenzierung und Äquivalentpräsentation in zweisprachigen Wörterbüchern: eine neue einheitliche Konzeption”. In: GOTTLIEB, Henrik; MOGENSEN, Jens Erik; ZETTERSTEN, Arne (ed.) *Symposium on Lexicography XI*. Tübingen: Max Niemeyer, p. 17–57.
- WÜSTER, Eugen (1959). “Das Wort der Welt, schaubildlich und terminologisch dargestellt”. *Sprachforum*, 3 (3), p. 183–203.

## Notes

1. We wish to thank Marie-Claude L’Homme for the support in the process of adding a third language to JuriDiCo, Benoît Robichaud for the addition of the French folder to the database, and Josée Poirier and Geneviève Landry for collecting part of the data included in this paper. The work described here was supported by the Fundação para a Ciência e Tecnologia, Ministério da Educação e do Ensino Superior de Portugal (FCT-MCTES), by the Social Sciences and Humanities Research Council of Canada (SSHRC) as well as by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PNPD 2013/ CAPES).

### APPENDIX 1. Data corpus

| French terms             | Frames                 | English candidate equivalents  | Portuguese candidate equivalents  |
|--------------------------|------------------------|--|---|
| alléguer <sub>1</sub>    | Contesting             | allege <sub>1</sub> ; impugn <sub>1</sub> ; rebut <sub>1</sub>   | arguir <sub>1</sub> ; impugnar <sub>2</sub> ; invocar <sub>1</sub>  |
| alléguer <sub>2</sub>    | Argumentation          | argue <sub>1</sub> ; assert <sub>1</sub> ; contend <sub>1</sub> ; invoke <sub>1</sub> ; state <sub>1</sub> ; submit <sub>1</sub>                                   | aduzir <sub>1</sub> ; alegar <sub>1</sub> ; declarar <sub>1</sub>   |
| appliquer <sub>1</sub>   | Apply law              | apply <sub>1</sub>   | aplicar <sub>1</sub>  |
| appliquer <sub>2</sub>   | Law applicability      | apply <sub>2</sub>   | aplicar <sub>2</sub> ; vincular <sub>1</sub>  |
| autoriser <sub>1</sub>   | Authorization          | authorize <sub>1</sub> ; permit <sub>1</sub> ; preclude <sub>1</sub> ; prohibit <sub>1</sub>   |   |
| conférer <sub>1</sub>    | Granting               | confer <sub>1</sub> ; deprive <sub>1</sub> ; entitle <sub>1</sub> ; grant <sub>1</sub>   | conceder <sub>1</sub> ; conferir <sub>1</sub> ; outorgar <sub>2</sub>   |
| conformer <sub>1</sub>   | Compliance             | comply <sub>2</sub> ; exceed <sub>1</sub> ; infringe <sub>2</sub> ; violate <sub>2</sub>   | cumprir <sub>1</sub> ; violar <sub>2</sub>  |
| contrevenir <sub>1</sub> | Constitutionality      | comply <sub>1</sub> ; infringe <sub>1</sub> ; satisfy <sub>1</sub> ; violate <sub>1</sub>  | violar <sub>1</sub>   |
| contrevenir <sub>2</sub> | Compliance             | comply <sub>2</sub> ; exceed <sub>1</sub> ; infringe <sub>2</sub> ; violate <sub>2</sub>   | cumprir <sub>1</sub> ; violar <sub>2</sub>  |
| démontrer <sub>1a</sub>  | Proof                  | establish <sub>2</sub> ; prove <sub>1</sub>  | demonstrar <sub>1</sub> ; provar <sub>1</sub>   |
| démontrer <sub>1b</sub>  | Proof                  | establish <sub>2</sub> ; prove <sub>1</sub>  | demonstrar <sub>1</sub> ; provar <sub>1</sub>   |
| déterminer <sub>1</sub>  | Assessing              | assess <sub>1</sub> ; determine <sub>1</sub>   | aferir <sub>1</sub> ; apreciar <sub>1</sub> ; ponderar <sub>1</sub>   |
| déterminer <sub>2</sub>  | Conclusion             | conclude <sub>1</sub> ; determine <sub>2</sub> ; infer <sub>1</sub>  | concluir <sub>1</sub>   |
| examiner <sub>1</sub>    | Assessing              | assess <sub>1</sub> ; determine <sub>1</sub>   | aferir <sub>1</sub> ; apreciar <sub>1</sub> ; ponderar <sub>1</sub>   |
| invoquer <sub>1</sub>    | Argumentation          | argue <sub>1</sub> ; assert <sub>1</sub> ; contend <sub>1</sub> ; invoke <sub>1</sub> ; state <sub>1</sub> ; submit <sub>1</sub>                                   | aduzir <sub>1</sub> ; alegar <sub>1</sub> ; declarar <sub>1</sub>   |
| invoquer <sub>2</sub>    | Contesting             | allege <sub>1</sub> ; impugn <sub>1</sub> ; rebut <sub>1</sub>   | arguir <sub>1</sub> ; impugnar <sub>2</sub> ; invocar <sub>1</sub>  |
| ordonner <sub>1</sub>    | Order                  | commit <sub>2</sub> ; impose <sub>1</sub> ; order <sub>1</sub> ; require <sub>1</sub>  | determinar <sub>2</sub>   |
| procéder <sub>1a</sub>   | Judge's Procedure      | proceed <sub>2a</sub>  | proceder <sub>2</sub>   |
| procéder <sub>1b</sub>   | Prosecutor's Procedure | proceed <sub>2b</sub>  |   |
| réglementer <sub>1</sub> | Regulations            | codify <sub>1</sub> ; establish <sub>1</sub> ; govern <sub>1</sub> ; mandate <sub>1</sub> ; prescribe <sub>1</sub> ; provide <sub>1</sub> ; recognize <sub>1</sub> | consignar <sub>1</sub> ; determinar <sub>1</sub> ; disciplinar <sub>1</sub> ; estabelecer <sub>1</sub> ; estatuir <sub>1</sub> ; estipular <sub>1</sub> ; prescrever <sub>1</sub> |
| <b>20</b>                | <b>15</b>              | <b>47</b>  | <b>30</b>   |