

Discovering frames in specialized domains

Marie-Claude L'Homme(1), Benoît Robichaud(1), Carlos Subirats Rüggeberg(2)

(1) Observatoire de linguistique Sens-Texte (OLST)

Université de Montréal
C. P. 6128, succ. Centre-ville
Montréal H3C 3J7 Canada

(2) Universidad Autónoma de Barcelona

Facultad de Letras, Departamento de Filología Española
Bellaterra, Barcelona, Spain

mc.lhomme@umontreal.ca, benoit.robichaud@umontreal.ca, carlos.subirats@gmail.com

Abstract

This paper proposes a method for discovering semantic frames (Fillmore, 1982, 1985; Fillmore et al., 2003) in specialized domains. It is assumed that frames are especially relevant for capturing the lexical structure in specialized domains and that they complement structures such as ontologies that appear better suited to represent specific relationships between entities. The method we devised is based on existing lexical entries recorded in a specialized database related to the field of the environment (*erode, impact, melt, recycling, warming*). The frames and the data encoded in FrameNet are used as a reference. Selected information was extracted automatically from the database on the environment (and, when possible, compared to FrameNet), and presented to a linguist who analyzed this information to discover potential frames. Several different frames were discovered with this method. About half of them correspond to frames already described in FrameNet; some new frames were also defined and part of these might be specific to the field of the environment.

Keywords: terminology, predicative lexical unit, Frame Semantics, actants, frame elements, environment

1. Introduction

Descriptive work on the specialized lexicon (carried out mainly by terminologists) has traditionally focused on terms that denote entities (e.g., *carbon dioxide, methane, organics, turbine*); these terms lend themselves to taxonomic representations (*carbon dioxide => greenhouse gas => gas*). Similarly, most concepts encoded in ontologies represent entities. For instance, while some properties appear in the environmental ontology *Envo* (“environmental condition”, “arid”), most concepts represented are concrete (“biome”, “habitat”). Processes and events are important components of knowledge in many specialized domains (including the environment). However, their representation and description raise a number of questions that differ from those raised by entities. Processes and events are designated by predicative terms (e.g., *decline, degradation, impact, pollute, prediction, recycling*) whose semantics cannot be captured completely with ontological or taxonomic representations. Predicative terms evoke frames comprising frame elements realized in specific grammatical constructions (*decline of X, X impacts Y*).

This paper proposes a method for discovering semantic frames (Fillmore, 1982, 1985; Fillmore et al., 2003) in specialized domains. We applied the method using linguistic descriptions extracted from a database related to the field of the environment. It is assumed that frames are especially relevant for capturing the lexical structure of specialized domains and that they complement other

structures such as ontologies that are considered better suited to represent specific relationships between entities. This assumption has already been made for the organization of knowledge in a specialized field (Faber et al., 2006); we extend it here to terms viewed as lexical units (Cruse, 1986).

The method we devised is based on existing lexical entries recorded in a specialized database (described in Section 4.1). The frames, lexical units and annotations encoded in FrameNet (2014) were used as a reference. Selected information was extracted automatically from a terminological database on the environment and from FrameNet; it was then presented to a linguist who analyzed this information in order to discover potential frames. Whenever possible, LUs were associated to frames. In other cases, new frames were proposed.

This paper is structured as follows. Section 2 presents previous work that applied Frame Semantics or compatible frameworks to the analysis of terminological data. The hypotheses on which our own work is based are also discussed. Section 3 comments on those aspects of Frame Semantics and FrameNet that are relevant for this research. Section 4 presents our methodology, along with a description of the data that was used. Section 5 discusses our results. We conclude with a summary of the results and mention a few directions for future work.

2. Why frames in specialized fields?

Recently, researchers have investigated new models for taking into account terms that refer to events and processes in specialized domains. One framework that

seems to attract the interest of terminologists is that of frames (Faber, 2012; Pimentel, 2013; Schmidt, 2009; Wandji et al. 2013). Frames in this context can be directly based on Frame Semantics (Fillmore, 1982, 1985; Fillmore et al., 2003) and the methodology developed within the FrameNet (2014) project or be loose adaptations of the original framework. Among other things, frames are particularly useful to analyze terms that designate processes and events (in other words, predicative terms) and the constructions in which their participants (frame elements) are realized.

Pimentel (2013) and Schmidt (2009) refer directly to Frame Semantics: Pimentel applied the framework to establish equivalence relationships between English and Portuguese verbs in the field of law; Schmidt introduced some adaptations to account for multilingual data (English, French and German) in the field of soccer. Wandji et al. (2013) attempt to discover frames in the field of medicine using natural language processing techniques and external resources (a medical terminology is used to identify potential frame elements automatically in a French corpus). Finally, Faber (2012) refers to Frame Semantics in order to account for processes in the field of the environment and proposes a very general frame to account for the different processes observed in the field. However, it seems that the general frame was defined for the purposes of this specific project rather than according to the principles of Frame Semantics and the methodology designed in FrameNet (cf. Section 3).

In this work, we hypothesize that terms sharing argument structures in specific subject fields (along with other lexico-semantic properties) evoke semantic frames. However, contrary to the previous work cited above, here semantic frames are discovered after terms are described rather than postulated prior to the descriptive work. The data used in this analysis is extracted from a database that was not compiled with a view to describing terms as lexical units that evoke frames; however, we believe that these can be discovered afterwards based on the lexico-semantic properties of lexical units. Hence, we assume that our descriptions can lend themselves to a frame-like analysis.

Our method should allow us to:

1. Unveil frames in specific subject fields based on predicative terms such as verbs and nouns (more specifically, frames in the field of the environment) that share lexico-semantic properties.
2. Define new frames specific to the field of the environment: Although some frames might correspond to frames already defined in FrameNet, it is likely that new frames will be discovered in the data analyzed.
3. Validate the notion of “frame” on data that was not previously encoded for that purpose.

3. Frame Semantics and FrameNet

Frame Semantics (FS) is based on the assumption that the meanings of lexical units (LUs) are constructed in relation to background knowledge, whose structure can be analyzed in terms of semantic frames. Frames are schemas or knowledge structures emerging from everyday experience (Fillmore, 1985; Fillmore and Baker, 2010). More precisely, a frame can be defined as the schematic representation of a situation that includes participants, “props”, and other conceptual elements, which constitute its frame elements (FEs).

According to Frame Semantics, lexical units are understood relative to frames. Thus the meaning associated with a particular unit cannot be understood independently of the frame it evokes. Based on this theory, FrameNet (FN) is investigating the various patterns of conceptualization that are involved in the English lexicon.

Since it is the semantic frame evoked by an LU (whether it belongs to the general or specialized domains) that enables the understanding of its meaning, the meaning of specialized LUs studied in this paper is analyzed with reference to the semantic frames they evoke. The goal of this project is to apply Frame Semantics and the methodology of the FN project to the study of a specialized domain, since it also is the frame evoked by a specialized lexical unit that enables its production and understanding.

Frames are defined following certain basic procedures. First of all, we need to determine if a lexical unit is a state, an event, a process, or an entity, since this basic distinction determines the primary division of the possible frames that a lexical unit may evoke. If an LU is an event, we have to determine its semantic domain, i.e. the most basic semantic frame to which it might be related.

For instance, if we want to determine the frame that is evoked by the verb *leave*, we have to consider that it belongs to the motion domain, which evokes a scenario where there are certain semantic roles, like a Theme, a Source, a Goal, and a Path, among others. First, we must determine what aspects of this general scenario are profiled by *leave*, which implies determining what are the frame elements that are profiled by this verb. In this case, *leave* evokes a frame where an object—a Theme—moves away from the point where the motion started—the Source—. Thus, we can create the frame **Departing** for *leave*, and for all the other lexical units that evoke the same frame, like *depart* (v), *departure* (n), *exit* (v), *exodus* (n), etc. The final decisions about the creation of new frames depend on the degree of granularity and precision of the frame semantic analysis we want to establish. Moreover, we can take advantage of the analysis that has been made in English FrameNet, and other FrameNets, like German, Japanese, Spanish FN, etc. (Boas 2009).

Frames are described in FrameNet (2014) that presents the following data:

- A description of the frame along with the list of frame elements FEs, part of which are core

(obligatory participants); the other part corresponds to non-core FEs (non-obligatory) (Figure 1). LUs that evoke this frame are also listed;

- A description of each LU that belongs to this frame (Figure 2): this description presents syntactic patterns and the valency of the LU;
- Annotated sentences that show how LUs and FEs are realized linguistically (Figure 3).

Frames can also share relationships with other frames (Figure 1). However, these relationships were not exploited directly in this analysis.

Departing ; An object (the **Theme**) moves away from a **Source**. The **Source** may be expressed or it may be understood from context, but its existence is always implied by the departing word itself.

FEs	
Core	
Source (Src)	All the verbs in this frame express some change of location, away from one place and to another. Any constituent that expresses the initial position of the Theme , before the change of location, is tagged with Source . Often the Source is understood from context.
Theme (Thm)	This is the object which moves. It may be an entity which moves under its own power, but it need not be.
Non-Core (partial list)	
Circumstances (Cir)	Circumstances describe the state of the world (at a particular time and place) which is specifically independent of the event itself and any of its participants.
Containing_event (Con)	This FE denotes an event that occurs or state of affairs that holds at a time that includes the time during which the event or state of affairs reported by the target occurs and of which it is taken to be a part.
(...)	
Frame-frame relations (partial list)	
Inherits from	Eventive_affecting
Uses	Motion
(...)	
Lexical units	
<i>decamp.v, depart.v, departure.n, disappear.v, disappearance.n, emerge.v, escape.n, escape.v, exit.n, exit.v, exodus.n, leave.v, skedaddle.v, vamoose.v, vanish.v</i>	

Figure 1: Partial description of the frame **Departing** (FrameNet 2014)

leave.v.

Frame: Departing

Definition: COD: go away from

Syntactic realizations (of core FEs) :

Source (98)	DNI.-- (29) NP.Obj (55) NP.Dep (3) PP[from].Dep (7) AVP.Dep (4)
Theme (98)	NP.Ext (95) NP.Obj (1) CNI.-- (2) INC.-- (1)

Valence Patterns:

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns			
1 TOTAL (1)	Cotheme	Source	Theme	
	PP[with] Dep	DNI --	NP Ext	
1 TOTAL (1)	Cotheme	Source	Theme	Time
	PP[with] Dep	DNI --	NP Ext	NP Dep
1 TOTAL (1)	Depictive	Depictive	Source	Theme
	AVP Dep	PP[unlike] Dep	DNI --	NP Ext

Figure 2. Partial description of the LU *leave* (FrameNet 2014)

*[ThemeThe two men] LEFT^{target} [Source^tthe office] .
Remember that [ThemeI] had LEFT^{target} [SourceEngland]
[Timeless than three days before this] [Timeafter an
exceptionally grey and dull winter] and the bright colours
which typify so much Australian knitting were a delight .
[TimeFollowing his marriage in 1892] [Themehe] LEFT^{target}
[SourceEngland] [Journeyon his third and final East African
tour] .*

Figure 3. A sample of annotated sentences with the target *leave* (FrameNet 2014)

4. Methodology

In this section, we give a description of the data used to carry out our analysis, the method developed, and the different tools designed to extract data in order to help us capture similarities between terms. The differences between FN and our own data, and the way we handled them are also discussed.

4.1 A specialized lexical database on the environment: DiCoEnviro

The analysis is based on data extracted from a specialized database called *DiCoEnviro*. The database (under construction) contains terms in English, French, Portuguese and Spanish. Entries provide a description of

impact ₂, vt

impact: *change* ~ *biodiversity_1, resource_1*

Status : 2

Synonym(s) : impact (vi, impact on something)

Contexts

Lexical relations

Explanation - Typical term	Related term
Related Meanings	
R	affect <i>influence_2</i>
Other Parts of Speech and Derivatives	
Noun	<i>impact_1</i>
<i>Biodiversity</i> or a <i>resource</i> that has been i.	impacted

French : *avoir un impact_1*

Figure 4: Entry *impact* (DiCoEnviro 2014)

the lexico-semantic properties of terms (Figure 4): actantial (i.e. argument) structure, linguistic realizations of actants (i.e. arguments), annotated sentences (based on the methodology devised within the FN project, Ruppenhofer et al. 2010), and lexical relationships (including paradigmatic relationships and collocations). The terms taken into account in this study are extracted from the English and the French versions of the database. We considered verbs (e.g., *impact*, *recycle*, *warm*, *perturber*, *réchauffer*) and nouns that refer to processes or events (*flood*, *decline*, *perturbation*, *fonte*, *incinération*). Up to now, 107 LUs in English and 159 in French were analyzed. All LUs analyzed come with annotated contexts. The database also contains LUs that have not been annotated yet: they were not considered in this specific analysis.

In annotations (Figure 5), the predicative unit appears in capital letters and in bold. Participants are divided into two categories: actants (in bold) correspond to obligatory participants (roughly equivalent to core frame elements); circumstants correspond to non-obligatory participants. Participants appear in different colors according to their role (Cause, Patient, etc.). A table summarizes the different patterns found in annotations and provides information on the syntactic function and groups of participants. More details about our annotation method are given in L'Homme (2012).

The energy sector is also very sensitive to weather and its variations, and hence will **BE significantly IMPACTED by climate change**. Coastal settlements in, for example, the Gulf of Guinea, Senegal, Gambia, Egypt, and along the East - Southern African coast would **BE adversely IMPACTED by sea-level rise through inundation and coastal erosion** (high confidence 6). Loss and retreat of glaciers would **adversely IMPACT runoff**

and water supply of peoples around the world in areas where glacier melt is an important water source (high confidence 6).

Actants		
Cause	Subject (SN) (1) Complement (PP-by) (4)	change erosion inundation loss retreat
Patient	Object (SN) (1) Subject (SN) (2)	runoff sector settlements supply
Others		
Location	Complement (PP-in)	area
Manner	Modifier (AdvP) (2)	adversely
Degree	Modifier (AdvP)	significantly

Figure 5: A sample of annotations for *impact* (DiCoEnviro 2014)

4.2 Analysis of data for discovering frames

The methodology for discovering frames consists basically in: 1) extracting relevant data from the DiCoEnviro, and 2) using FrameNet data (in English) as a reference to identify a first set of existing frames that the terms in our database could evoke. A set of tools were devised to help us carry out our analysis.

During the first stages of the analysis, the English and French data were considered separately. The following data was extracted from each version of the DiCoEnviro (the data was presented in tables that could be sorted according to various criteria, cf. Figure 6):

- Semantic roles of actants and the order in which they appear (columns 4, 5, 6, and 7);
- Semantic roles of circumstants (i.e. non-obligatory participants) (column 8);

2	vi	freeze.1.en freeze: Patient ~	Patient	-	-	-	Cause Expense Time	Change_of_phase	-
2	vi	melt.1a.en melt: Patient ~	Patient	-	-	-	Degree Location Time	Change_of_phase	IncepPredMinus : glacier.1.en IncepPredMinus : ice cap.1.en IncepPredMinus : ice.1.en
2	vi	thaw.1a.en thaw: Patient ~	Patient	-	-	-	-	Change_of_phase	IncepPredMinus : permafrost.1.en
2	vi	cool.1a.en cool: Patient ~	Patient	-	-	-	Manner Time	Change_of_temperature	IncepPredPlus[MAN.froid] : climate.1.en

Figure 6: Sample of extracted data from the English version of the DiCoEnviro (2014)

DiCoEnviro	FrameNet																										
<p>melt.1a (vi) SA: melt: Patient{ ice } ~ Change_of_phase (FN tel quel) :</p> <table border="1"> <tr> <td>DiCoEnviro</td> <td>↔</td> <td>FrameNet</td> </tr> <tr> <td>Patient</td> <td></td> <td>Undergoer</td> </tr> </table> <table border="1"> <tr> <td>en</td> <td>fr</td> </tr> <tr> <td>freeze.1</td> <td>dégel.1</td> </tr> <tr> <td>melt.1a</td> <td>fondre.1</td> </tr> <tr> <td>thaw.1a</td> <td>fonte.1</td> </tr> <tr> <td></td> <td>gel.1</td> </tr> <tr> <td></td> <td>geler.1</td> </tr> </table> <p>11 contextes...</p>	DiCoEnviro	↔	FrameNet	Patient		Undergoer	en	fr	freeze.1	dégel.1	melt.1a	fondre.1	thaw.1a	fonte.1		gel.1		geler.1	<p>melt (V) ID: 5954 COD: make or become liquefied by heating. Change_of_phase: In this frame an <i>Undergoer</i> undergoes a change of phase. Note that this frame contrasts with <i>Change_of_consistency</i> in that this frame describes a change of an <i>Undergoer</i> between different phases (i.e. solid to liquid or frozen to "unfrozen"). 5 Examples Annotated Contexts</p> <table border="1"> <thead> <tr> <th>Type</th> <th>FE</th> </tr> </thead> <tbody> <tr> <td>Core</td> <td>Undergoer</td> </tr> <tr> <td>Extra-Thematic</td> <td>Circumstances Result Subregion</td> </tr> <tr> <td>Peripheral</td> <td>Degree Initial_state Manner Place Speed Time</td> </tr> </tbody> </table> 	Type	FE	Core	Undergoer	Extra-Thematic	Circumstances Result Subregion	Peripheral	Degree Initial_state Manner Place Speed Time
DiCoEnviro	↔	FrameNet																									
Patient		Undergoer																									
en	fr																										
freeze.1	dégel.1																										
melt.1a	fondre.1																										
thaw.1a	fonte.1																										
	gel.1																										
	geler.1																										
Type	FE																										
Core	Undergoer																										
Extra-Thematic	Circumstances Result Subregion																										
Peripheral	Degree Initial_state Manner Place Speed Time																										

Figure 7: Comparison of DiCoEnviro data with FN data for corresponding lexical items

- Verbs and nouns associated with specific lexical functions (LFs) (column 10). LFs are used to describe lexical relationship between lexical units (in this work all relations encoded are those observed between the components of a collocation. Mel'čuk et al. 1995).¹

In addition to the previous tables, a comparison with the English FN was carried out to analyze the English data. All LUs contained in the English DiCoEnviro were searched in the XML supplied by the FrameNet team. The tool searches the DiCoEnviro for terms that also appear in FN and presents side by side relevant data from both databases (Figure 7). More specifically, from FrameNet, it retrieves the frames that the lexical items evoke along with their definition, its core FEs, and the relationships shared by frames with other frames. This information is presented on the right hand side of the table. From the DiCoEnviro, it retrieves the actantial structure and the actants. The information appears on the left hand side of the table.

In addition when a correspondence between a term in the

DiCoEnviro can be established, the table is updated with the comparative information (this is dealt with in Section 4.4).

During the analysis, we noticed some differences between the XML file and the online version of FN. When such differences existed, we referred to the online version.

4.3 Handling specific differences between FrameNet and the DiCoEnviro

When carrying out our comparison of the English version of the DiCoEnviro with FN data, specific differences between them needed to be taken into account and handled.

The most important differences were related to the encoding of participants in DiCoInfo compared to the description of frame elements in FN. Here is how we handled the differences.

First, in FrameNet, FEs are defined at the level of frames (cf. Figure 1) while in DiCoEnviro, actants (and circumstants) are stated at the level of LUs (cf. Figure 4). We established that LUs in DiCoEnviro can belong to a frame if a relationship could be established between the set of core FEs and the actants and if the FEs and actants were represented with comparable labels. This is a necessary condition but not a sufficient one in the sense that LUs in the DiCoEnviro can have the same set of roles

¹ The relationship between lexical functions and Frame Semantics was first explored in Alonso Ramos et al. (2008).

and still belong to different frames (as will be seen below, labels used in the DiCoEnviro are much more general than those defined in FrameNet). Circumstants could differ from one LU to the other.

Secondly, due to the objectives of each resource, the number of core FEs in a frame could differ in comparison with the number of actants represented for a LU in the DiCoEnviro. Often, the number of core FEs was higher than the number of actants in the DiCoEnviro. In some cases, the DiCoEnviro defines a participant as being a circumstant and a correspondence could be established with FrameNet. In other cases, the specificity of the specialized domain needed to be taken into consideration. For example, the **Cause_temperature_change** frame in FN states four core FEs (Agent or Cause, Item, and Hot_cold_source). The latter FE (Hot_cold_source) is realized in structures such as *He chilled the drinks on ice*. There were no such instantiations of this participant in the DiCoEnviro data.

Thirdly, labels used for most FEs are very specific since they are defined within a frame. In the DiCoEnviro, labels are general and defined for the entire set of terms that are included in the database. We considered that the labels were equivalent if the meaning of the lexical unit was the same. For example, labels such as Entity, Item, Theme, and Undergoer in FrameNet were assumed to correspond to Patient in DiCoEnviro. Similarly, labels such as Agent and Protagonist were considered to correspond to Agent in the DiCoEnviro.

Fourthly, in FrameNet, different labels can account for an FE that would be realized in the same syntactic function. In the DiCoEnviro, actants can be split (Agent or Cause for instance). In both cases, we considered these as being instantiations of the same argument position. For instance, the frame **Objective_influence** is defined as follows in FN:

Objective_influence: An Influencing_variable, an Influencing_situation, or an Influencing_entity has an influence on a Dependent_entity, Dependent_variable, or a Dependent_situation.

Then, core FEs are listed without reference to their position relative to the lexical units that evoke the **Objective_influence** frame. The syntactic positions in which they can be realized appear in the annotations of sentences linked to LUs.

In the DiCoEnviro, the actants are represented as follows (for *affect*):

affect: Cause ~ Patient

In order to compare the data from FrameNet and DiCoEnviro, FEs and actants were represented as follows:

Frame element 1: Influencing_variable |
Influencing_situation | influencing_entity
Frame element 2: Dependent_entity |
Dependent_variable | Dependent_situation
Actant 1: Cause
Actant 2: Patient

4.4 Defining frames with the data

The data from the two resources and contained in the different tables (cf. Figures 6 and 7) was carefully examined. Then, LUs from DiCoEnviro were assigned manually to a relevant frame in an XML file). As shown in Figure 8, we indicate:

- The name of the frame (**Change_of_phase**)
- The comparison with FrameNet (in this case, the description in the DiCoEnviro is entirely compatible with the frame as defined in FrameNet (*FN tel quel*)); but there are other cases (further discussed in Section 5) where the correspondence is partial or a new frame needed to be created;
- The list of core Frame elements (Undergoer)
- The list of Actants (Patient)

```

<Frame identificateur="Change_of_phase" note="FN tel quel">
  <Core-FEs>
    <Core-FE no="1">Undergoer</Core-FE>
  </Core-FEs>
  <Actants>
    <Actant no="1">Patient</Actant>
  </Actants>
  <Lexies>
    <Lexie dico="dicoenviro" langue="en" identificateur="thaw" no_acceptation="1a"/>
    <Lexie dico="dicoenviro" langue="en" identificateur="melt" no_acceptation="1a"/>
    <Lexie dico="dicoenviro" langue="en" identificateur="freeze" no_acceptation="1"/>
    <Lexie dico="dicoenviro" langue="fr" identificateur="fondre" no_acceptation="1"/>
    <Lexie dico="dicoenviro" langue="fr" identificateur="fonte" no_acceptation="1"/>
    <Lexie dico="dicoenviro" langue="fr" identificateur="geler" no_acceptation="1"/>
    <Lexie dico="dicoenviro" langue="fr" identificateur="dégel" no_acceptation="1"/>
    <Lexie dico="dicoenviro" langue="fr" identificateur="gel" no_acceptation="1"/>
  </Lexies>
</Frame>
```

Figure 8 Definition of a frame with the LUs in the DiCoEnviro

- The list of LUs in the DiCoEnviro that evoke this specific frame. (*thaw, melt, freeze, fondre, fonte, geler, dégel, gel*).

Once a DiCoEnviro LU is assigned to a frame, this is taken into account in our tables (cf. Figures 6 and 7).

5. Results

The analysis revealed a number of interesting results we now discuss. We divided them into five categories that we will present separately (Table 1):

1. The actantial structures of LUs in the DiCoEnviro and the list of FEs in frames recorded in FrameNet were entirely compatible (despite the fact that labels used are different, cf. 4.3);

2. The actantial structures of LUs in DiCoEnviro are partly compatible with the list of FEs in a frame;
3. A new frame was created;
4. The argument structures of sets of LUs in DiCoEnviro displayed an alternation; they are recorded in the DiCoEnviro in two separate entries but may appear in the same frame in Framenet;
5. Some cases are problematic and still have to be investigated.

Category	Number of Frames	Number of English LUs	Number of French LUs
Entirely compatible	12	27	36
Partly compatible	11	34	44
Alternation	2	8	8
New	19	29	55
Not defined	8	9	16
TOTAL	52	107	159

Table 1: Results of the analysis

As can be seen in Table 1, 69 LUs in English and 88 LUs in French contained in the data analyzed evoke 25 frames that appear in FrameNet. These include frames under the first three categories (entirely compatible, partly compatible and alternation). In 12 frames, the actantial structures in DiCoEnviro are entirely compatible with the description of the frame provided in FrameNet. This is the case with the LUs *damage* (n) and *damage* (v). In FrameNet, they appear in the **Damaging** frame.

FrameNet: **Damaging**. An Agent affects a Patient in such a way that the Patient (or some Subregion of the Patient) ends up in a non-canonical state. (The list of core FEs includes a Cause.)

In DiCoEnviro, *damage* (v) and (n) also have two arguments and evoke the **Damaging** frame.

DiCoEnviro: *damage*, v.: Cause ~ Patient; *damage*, n.: ~ to Patient by Cause

In addition, although *degrade* and *degradation* are not recorded in FrameNet, they evoke the same frame based on our analysis and, thus, were added to its list of lexical units. The French LUs *dégradation*, *dégrader*, *endommager* also evoke this frame and were assigned to it.

In 34 cases in English and 44 in French, the description of the argument structures of LUs in the DiCoEnviro and the number of frame elements in FrameNet are not exactly the same even though the frame evoked appears to be the same. For instance, the **Cause_temperature_change** frame in FrameNet states three core FEs (Agent, Cause, Hot_cold_source), while the relevant LUs in DiCoEnviro refer to two arguments only. We considered these descriptions as partly compatible, since the overall frame

appears to be the same.

FrameNet: **Cause_temperature_change**: In this frame, an Agent changes the temperature of an Item. (The list of FEs includes a Hot_cold_source.)

DiCoEnviro: *cool* 1b, v.: Agent or Cause ~ Patient;
warm 1b: Agent or Cause ~ Patient

The French LUs *réchauffement*, *réchauffer* 1b, *refroidissement*, *refroidir* 1b evoke the same frame and were assigned to it.

Although they share many similarities with existing frames in FrameNet, it must be kept in mind that the frames discovered in the DiCoEnviro are much more restricted and the LUs that evoke them probably correspond to subsenses (Cruse, 2011). For instance, the **Cause_temperature_change** frame (that includes LUs such as *cool* and *warm* in the DiCoEnviro) only applies to actants realized with lexical units such as *climate*, *Earth*, *atmosphere*).

In some cases (alternations), the DiCoEnviro makes systematic distinctions that would not necessarily be made in frames defined in FN. We identified two such cases in our data. For instance, the verb *predict* can be found in the two following constructions:

... *models*, which *predict carbon fluxes*
... *our models will better enable us to predict the consequences...*

Since the alternation is regular in the data related to the environment, we defined two different actantial structures in the DiCoEnviro to account for them:

Predict 1a: Method ~ Patient
Predict 1b: Agent ~ Patient with Method

However, the two LUs were placed in the same frame already defined in FrameNet, i.e. **Predicting**:

Predicting: A **Speaker** states or makes known a future **Eventuality** on the basis of some **Evidence**. (The list of core FEs, include Speaker, Eventuality, Medium, and Topic.)

A total of 19 new frames (frames that do not appear in FrameNet) accounting for 19 English LUs and 55 French LUs were discovered. These can correspond to frames that simply do not yet appear in FrameNet (for instance, a **Flood** frame) or frames that are specific to the field (a **Change_natural_feature** was created to include LUs such as *retreat*, *thinning*, *erode*). A particularly interesting set of frames were created to account for LUs evoking waste management:

Managing_waste: *manage*, *management*
Recover: *recupérer*, *valoriser*, *valorization*
Dispose_of: *disposal*, *éliminer*, *elimination*
Cause_change_into_reusable_material: *compost*, *composting*, *compostage*, *composter*, *transformer*

Finally, in 8 cases in English and 16 cases in French, frames could only be defined provisionally for the

following reasons. Some frames currently contain only one LU and need to be further validated with additional data. In other cases, the relationship between LUs in the DiCoEnviro and those appearing in FrameNet need to be further investigated. This is the case with a set of LUs in the DiCoEnviro, i.e. *change* (v), *alter*, *alteration*, *perturbation*. At first sight, these seem to evoke the **Cause_change** frame. However, this frame includes LUs such as *convert* and *conversion* that behave quite differently in our data, as shown below:

A turbine converts this movement to electrical energy. Our research focuses on the conversion of biomass to energy and fuels ...

... human activities change the composition of the atmosphere

These aerosols , in addition to directly reflecting or absorbing sunlight, can alter cloud processes ...

6. Conclusion

This work consisted in evaluating the extent to which existing terminological descriptions could be used to discover semantic frames in a specific subject field, that of the environment. A set of 107 predicative terms in English and 159 in French were analyzed and compared to the contents of FrameNet. The results show that predicative LUs can be assigned to existing frames in FN; for others, new frames are proposed. Finally, a small set of LUs need to be further investigated. These results indicate that our data, although it was not initially defined for that purpose, lends itself to a description compatible with Frame Semantics.

The results also show a number of differences between the data provided in FN and that encoded in the DiCoEnviro. Some are due to conceptualizations that may differ in specialized subject fields (more specific entities to which a frame applies, for instance); other have a methodological cause (number of actants vs. number of core frame elements, labelling of actants vs. frame elements).

Terminologists are currently adding new LUs to the DiCoEnviro and this work will lead to the discovery of additional frames (some of which will correspond to existing frames; some of which will be new). In addition, relations between frames already defined can be perceived intuitively (e.g., between the different frames related to waste management or between those evoking types of changes). We have started establishing these relations.

7. Acknowledgements

This work was supported by the Social Sciences and Humanities Research Council (SSHRC) of Canada. We would like to thank the three anonymous reviewers for their useful comments on a previous version of this paper.

8. References

Alonso Ramos, M. O. Rambow and L. Wanner. 2008. 'Using Semantically Annotated Corpora to Build

Collocation Resources'. In *Proceedings of the International Conference on Linguistic Resources and Evaluation (LREC)*, Marrakesh.

Boas, H. C. (ed.) 2009. *Multilingual FrameNets in Computational Lexicography*. The Hague: Mouton.

Cruse, D.A. 1986. *Lexical Semantics*, Cambridge: Cambridge University Press.

Cruse, A.. 2011. *Meaning in Language*. Oxford: Oxford University Press.

DiCoEnviro. Dictionnaire fondamental de environnement L'Homme, M.-C.) 2014. *DiCoEnviro. Dictionnaire fondamental de l'environnement*

[http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search_enviro.cgi].

Envo. Environmental Ontology. 2013. Ontology Lookup Service, The European Bioinformatics Institute [<http://environmentontology.org/home>]

Faber, P. et al. 2006. Process-oriented terminology management in the domain of Coastal Engineering. *Terminology* 12(2): 189-213.

Faber, P. (ed.). 2012. *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin/Boston: De Gruyter.

Fillmore, C.J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2): 222-254.

Fillmore, C. J. & C. Baker. 2010. A frames approach to semantic analysis. In *The Oxford Handbook of Linguistic Analysis*, Bernd Heine and Haiko Narrog (eds), 313-339. Oxford: OUP.

Fillmore, C.J., C.R. Johnson & M. Petruck. 2003. Background to FrameNet. In *International Journal of Lexicography* 16(3): 235-250.

FrameNet. [<https://framenet.icsi.berkeley.edu/fndrupal/>]. Accessed March 15, 2014.

L'Homme, M.C. 2012. Adding syntactico-semantic information to specialized dictionaries: an application of the FrameNet methodology. In Gouws, R. et al. (eds.). *Lexicographica* 28: 233-252.

Mel'cuk, I., A. Clas & A. Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*, Bruxelles : Duculot.

Pimentel, J. 2013. Methodological bases for assigning terminological equivalents. A contribution. In *Terminology* 19(2): 237-257.

Ruppenhofer, J. M. Ellsworth, M. Petruck, C. Johnson & J. Scheffczyk. 2010. FrameNet II: Extended Theory and Practice. 2010. [<http://framenet.icsi.berkeley.edu>].

Schmidt, T. 2009. The Kicktionary – A Multilingual Lexical Resources of Football Language. In Boas, H.C. (ed). *Multilingual FrameNets in Computational Lexicography. Methods and Applications*, Berlin/New York: Mouton de Gruyter: 101-134.

Wandji, O., N. Grabar and M.C. L'Homme. 2013. Discovery of semantic frames for a contrastive study of verbs in medical corpora. in *Terminology and Artificial Intelligence*, Proceedings. TIA 2013, Paris.