Evaluating an Interlingual Semantic Representation - application to FTRD's MTT-inspired system

Jungyeul Park, Estelle Maillebuau, Emilie Guimier De Neef, Jérôme Vinesse, Johannes Heinecke

TECH/EASY – France Télécom R&D
2, avenue Pierre Marzin, 22307 Lannion Cedex, France
{jungyeul.park | estelle.maillebuau | emilie.guimierdeneef | jerome.vinesse | johannes.heinecke}@orange-ftgroup.com

Abstract

This paper presents our semantic representation which has been introduced in France Télécom's third-generation rule-based interlingual MT system. After briefly describing our system, we compare MTT with our semantic representation. We propose to evaluate our semantic representation as an interlingual semantic representation in two ways. First, using our MT system we perform a robustness test which we refer to as a transparency test for our semantic representation. Second, we measure the rate of shared semantic representations between English and French: BLEU/NIST scores for a transparency test are relatively high and almost half of the corpus shares the same semantic representation.

Keywords

Interlingual semantic representation, Semantic graph, Semantic transfer rules, Transparency test, BLEU and NIST metrics.

1 Introduction

An interlingual semantic representation is designed to capture the meaning of a sentence that is common to each language (Mitamura et al. 2004). The main use of interlingual representation has been machine translation. Interlingual machine translation is an instance of rule-based machine translation approaches. In interlingual machine translation, the source language is transformed into an interlingual representation, that is source and target language independent representation, and the target language is then generated out of the interlingual representation[1].

[1] The text is adapted from Wikipedia. See *http://en.wikipedia.org/wiki/Machine_translation*. There were also recently UNL-based interlingual MT systems (e.g. Boitet (2005)).

Recently, the use of interlingual representations includes applications for question answering (Ogden et al. 1999), representing agent actions (Kipper and Palmer 2000) and knowledge acquisition from text (Nyberg et al. 2002)[2].

In this paper, we present the interlingual semantic representation of France Télécom R&D (FTRD). In the next section, we explore FTRD's system for semantic analysis. We compare MTT with our semantic representation in section 3 and then we present our semantic model as an interlingual semantic representation in section 4. We propose experimental evaluations of our semantic representation in English and French using BLEU and NIST metrics in section 5. We discuss conclusion and some future works in section 6.

2 System Overview

FTRD's system consists of segmentation, lexical analysis, dependency analysis and semantic analysis. Figure 1 shows results of each phase for *J'ai un peu de fièvre* ('I have a little fever ').

The dependency analysis in Figure 1b is used to build the structure of dependency using a segmented sentence in Figure 1a. For the semantic analysis in Figure 1c, the system uses morpho-syntactic lexicons which contain the actual morpho-syntactic information and a multilingual thesaurus which defines semantic units. The main purpose of the thesaurus is to accumulate information on different senses of words and relationships between words. The thesaurus has four levels, that are macro-domains (a set of domain) (e.g., *nature*), domains (e.g., mammals), themes (a set of synsets) (e.g., dogs), and synsets (e.g., *poodles* for English and *caniches* for French). This kind of organization facilitates the construction and the maintenance of semantic information. See Maillebuau (2002) for the detail of our semantic data.

3 Comparison between MTT and FTRD's semantic representation

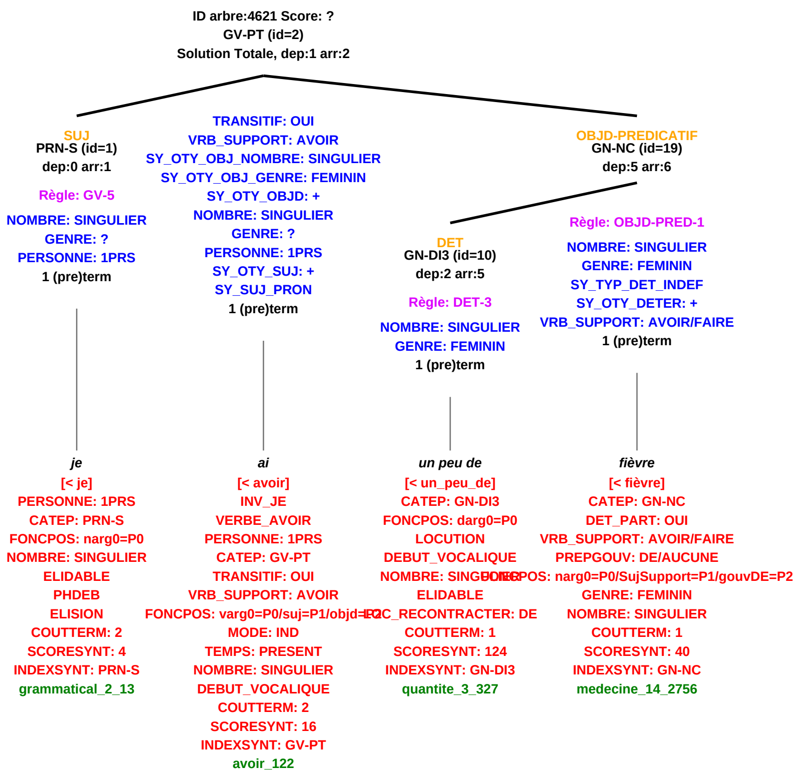
FTRD's semantic model is inspired by the Explanatory Combinatorial Dictionary (ECD) which is centred on the description of words (Mel'čnk, 1988). Lexical functions of ECD allow us to describe syntactic and phrasal cooccurrence of each word entry. For example, for *compliment*, the lexical function *Oper* allows to link it with light verbs like *faire*, *dire*, *exprimer*… That is, these light verbs would form collocations with the *compliment*. However, the sense of a word appears only if it benefits from its proper entry.

FTRD's semantic model describes different senses within concepts (which we name "*tribe*s"), which not only give the sense of a word, but also make relations with other words explicit. If *bark* is found in the tribe of *dog* and is linked to it by the lexicalization function *CRY*, the sense of *bark* is explicit and it may be easily calculated to manipulate its semantic representation. The position of *bark* in our semantic model (a tribe and a lexicalization function) is enough for defining it in its proper sense.

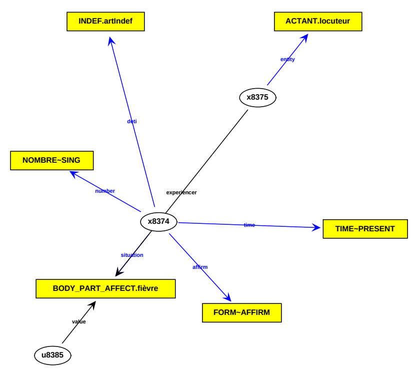
[2] The text is adapted from Mitamura et al. (2004).

je/je/PRN-S/Pp1s **ai**/avoir/GV-PT/Vmip1s **un peu de**/un_peu_de/GN-DI3/D[mf][sp] **fièvre**/fièvre/GN-NC/Ncfs

a. Segmentation and lexical analysis



b. Dependency analysis



c. Semantic analysis

Figure 1: FTRD's system overview, results of segmentation, lexical analysis, dependency analysis and semantic analysis

Every word described in our semantic model reflects a sense and every sense is described only once in order to avoid ambiguities. One of the constraints of FTRD's semantic model is to use (therefore, to describe) each lexicalization only once and this is different from the ECD.

The ECD becomes attached to cooccurrence (using a collocation) of a word even though FTRD's semantic model describes its place in the concept. The goal of our semantic model is not to describe links among words but to be able to calculate their senses according to their positions in a tribe. The beginning point of our semantic model is, therefore, a tribe, not a word.

Even though the starting purposes between our semantic model and ECD are different, certain links would be singled out. Both models describe relations between senses, that is, differential semantics which is opposed to componential semantics which is more applied to a sense itself (the decomposition of a sense).

Certain semantic relations described by the lexicalization function in FTRD's semantic model are very similar to those which are reflected by ECD. These relations are shown in Table 1.

<table>
<tr><th></th><th>MTT</th><th>FTRD's semantic model</th></tr>
<tr><td>Synonym</td><td>Lexical function Syn</td><td>Belonging to same sub-tribe with the same role and category</td></tr>
<tr><td>Generic/Specific Link</td><td>Lexical function Gener</td><td>Link between certain tribes of the same type</td></tr>
<tr><td>Syntactic derivation (nominalization, verbalization)</td><td>Lexical function S0, V0, ...</td><td>Change of category and of distribution of roles</td></tr>
<tr><td>Semantic derivation (cf. louer/locataire)</td><td>Lexical function</td><td>Semantic role or lexicalization function</td></tr>
<tr><td>Aktionsart (cf. commencer à, terminer de …)</td><td colspan="2">Described equally by functions, for example, Lexical function Incep = lexicalization function INCHOATIVE</td></tr>
<tr><td>Causative relations</td><td colspan="2">Described equally by functions, for example, Lexical function Caus = lexicalization function CAUSATIVE</td></tr>
<tr><td>More specific relations, for example, cry of animals</td><td colspan="2">Described equally by functions, for example, Lexical function Sound = lexicalization function CRY</td></tr>
</table>

Table 1 : Comparison summary between MTT and our semantic model

General principles (methods, purposes, applications …) of both semantic models are, therefore, different; certain number of relations described in the Table 1 would be linked. Some of lexicalization functions in our model are almost the same as lexical functions in ECD.

Semantic roles which are generally found in other literature, such as AGENT, PATIENT, THEME, LOCATION, INSTRUMENT, EXPERIENCER, etc. describe the distribution of arguments. As part of FTRD's semantic model, more variety in these roles is needed to describe types of tribes in our semantic model. Semantic roles defined before would be kept and other semantic roles might be added when it is needed to create types of tribe. For example, the semantic role SITUATION represents a situation itself, which is introduced in every predicate like *drink* = SITUATION AGENT PATIENT.

Principal information for creating and organizing types of tribe which describe verbs (or rather predicates) is the sense and the valence. Two predicates that belong to the same type of tribe have the same (i) distribution of arguments as well as corresponding semantic roles and (ii) applicable lexicalization functions. Verbs *eat* and *drink* belong to tribes for which the type is *OPERATION*. Their distribution of arguments is the same: SITUATION AGENT PATIENT and lexicalization functions used in tribes are equally similar (for example, *POSSIBLE* for *edible* and *drinkable*, *NEGATIVE* for *inedible* and *undrinkable*). Like verbs, a classification of adjectives using the function of its sense and of its syntactic behaviour, allows us to create and organize types of tribe of properties and relations. Tables 2 and 3 show respectively examples of tribe and lexicalization functions.

Type of tribe	Argument distribution	Example
CHANGE_STATE	SITUATION EXPERIENCER	*Melt*
COGNITIVE_ACTIVITY	SITUATION AGENT THEME	*conceive*
PERCEPTION	SITUATION EXPERIENCER THEME	*See*

Table 2 : Examples of tribes

Lexicalization function	Sense of the function
CAUSATIVE	Caused by something
ITERATIVE	Repeats

Table 3 : Examples of lexicalization functions

4 Interlingual semantic representation

Primitives of our semantic model are predicates and individuals. Predicates can be either lexical or grammatical. The former represents the semantic content of lexical words like [BODY_PART_AFFECT.mal] in Figure 2 which in turn represents the sense for *one part of the body is in pain*[3]. The latter conveys grammatical content with

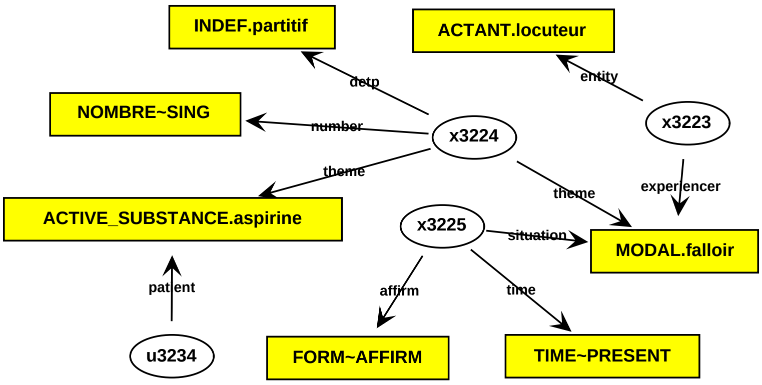
[3] Most predicates have a name in French, for example *mal* ('ill') in [BODY_PART_AFFECT.mal].

semantic counterpart like [NOMBRE~SING] in Figure 2 which represents *singular number*. Individuals are described through their connections with predicates like (x5699) in Figure 2.

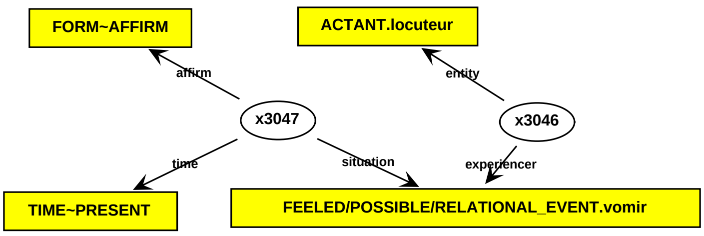
The internal data structure of our semantic representation uses a directed acyclic graphs (DAGs) that consists of a set of two types of nodes (predicates and individuals) and a set of edges. Thus we refer to our semantic representation as a semantic graph.

These graphs are well-adapted as an interlingual structure whenever language pairs, for example English and French, use isomorphic semantic structures. They deal with syntactically non-isomorphic cases like impersonal French verbs in (1) or light verb constructions in (2). Figure 2 shows semantic graphs for (1) and (2).

(1)	a. I need aspirin.	(2)	a. I feel sick.
(1)	b. Il me faudrait de l'aspirine.	(2)	b. J'ai des nausées.



a. *I need aspirin* (*Il me faudrait de l'aspirine*)

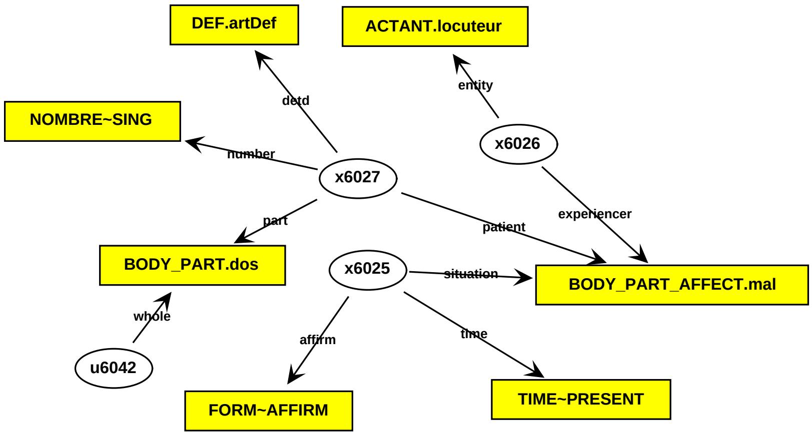


b. *I feel sick* (*J'ai des nausées*)

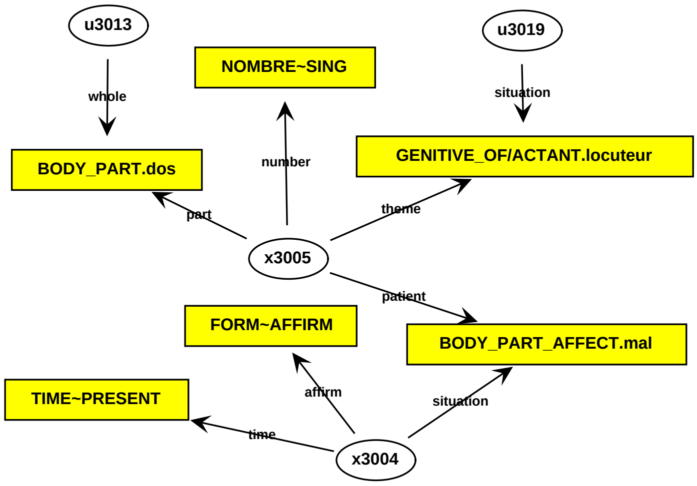
Figure 2 : Isomorphic semantic structure for a non-isomorphic syntactic structure

When language pairs are semantically non-isomorphic, we use semantic transfer rules. For instance, the proper English translation for the French sentence *J'ai mal au dos* is *My back hurts*. Semantic graphs of these sentences are given in Figure 3 which are semantically non-isomorphic. Figure 4 shows the semantic transfer rules from French into English for Figure 3 (that is, the semantic transfer rules for *J'ai mal au dos* into

My back hurts). The dot symbol (.) is used as a separation mark between source and target languages. See Park et al. (2007) for the detail of semantic transfer rules[4].



a. *J'ai mal au dos*



b. *My back hurts*

Figure 3 : Non-isomorphic semantic structure

```
# French
BODY_PART_AFFECT.mal(situation=x3,patient=x1,experiencer=x2) &
BODY_PART.*1(part=x1,whole=x4) &
DEF.artDef(detd=x1) &
ACTANT.*2(entity=x2) .
# English
BODY_PART_AFFECT.mal(situation=x3,patient=x1) &
BODY_PART.*1(part=x1,whole=x4) &
GENITIVE_OF/ACTANT.*2(situation=x5,theme=x1).
```

Figure 4 : Semantic transfer rules from French into English (from *avoir mal à X* to one's *X hurts*)

One of the advantages of our semantic representation is that it easily normalizes over conversives (e.g. *X bought a book from Y* versus *Y sold a book to X*), however it

[4] The use of transfer rules may suggest that our system is rather a transfer-based machine translation system at a language-dependent semantic level. We claim, rather, that the semantic level is language independent but that for some sentence pairs, the content expressed in different languages is semantically different.

sometimes fails to handle more complex linguistic phenomena such as pragmatics (e.g. *X started its business* versus *X opened its doors to customers*) [5].

5 Experimental Evaluation

In this section, we propose to evaluate our semantic representation as an interlingual semantic representation in two ways. First, using our MT system (Projet *Transat* 2006) we perform a robustness test which we refer to as a transparency test for our semantic representation. Second, we measure rate of the shared semantic representation between English and French.

A transparency test is that we configure the source and the target language identically in our MT system, for example a source language is English and a target language is also English to verify our semantic representation's robustness. If the generated target sentence is the same as the input source sentence, we refer to this sentence as 'semantically' transparent in our system. Table 4 shows scores of BLEU and NIST for English and French. The system generates 8065 and 1017 semantic graphs for a bilingual corpus which contains 200 test sentences in English and French respectively [6]. BLEU (Papineni et al. 2002) and NIST (Doddington 2002) are generally used for automatic Machine Translation evaluation.

From English to English		From French to French	
BLEU	NIST	BLEU	NIST
0.6826	6.6233	0.6214	5.7591

Table 4 : Transparency test for English and French

Non-transparent behaviour of the system may occur for different reasons, which are distinguished by these measures: (i) difficulty to build the expected semantic representation from text, (ii) weaknesses in the expressivity of the interlingual representation, and (iii) difficulty to produce the relevant text from the semantic representation. Further investigation should help clarify the proportion of each factor.

There are 30 and 35 silent sentences which have no output in English and French. Silent sentences are mostly caused by a lack of construction rules which generate the semantic graph from a dependency analysis, and are therefore symptomatic of difficulties in building the semantic representation rather than weaknesses in its expressivity. We actually working on this problem and considering a system that allows us to automatically build construction rules.

If we do not consider silent sentences, we increase BLEU/NIST scores to 0.8646/8.5728 for English and 0.8571/8.4734 for French. That is, if we resolve the problem of a lack of construction rules, we not only have a more robust semantic analysis system but also a more robust generating system from a semantic analysis.

[5] Examples and the text are adapted from Mitamura et al. (2004)

[6] 200 test sentences are randomly extracted from FTRD's bilingual corpus which contains several conversation scripts between a chemist and a customer. The original corpus contains about 1200 bilingual sentences

We are working on this problem and trying to build construction rules automatically for better performance.

Using a transparency test for our semantic representation, we measure the ability not only to analyse texts towards the interlingual semantic representation, but also to generate accurate surface texts from the semantic representation.

There is 48% of 200 sentences that share their semantic representation between English and French. It is important to obtain shared semantic graphs between French and English to keep the semantic consistency in our system. It also allows us to limit transfer operations when it comes to MT systems (which our system is originally designed for) and in consequence to try to achieve a proper interlingual system.

6 Conclusion

In this paper, we present our semantic representation as an interlingual semantic representation. We propose to evaluate our semantic representation in two ways: a transparency test for our semantic representation and the measurement of rate of the shared semantic representation between English and French. BLEU/NIST scores for a transparency test are relatively high and almost half of a corpus shares their semantic representation.

Among different reasons for non-transparent behaviour of the system, difficulty to build the expected semantic representation is mostly caused by a lack of construction rules and we working on this problem. On the other hand, difficulty to produce the relevant the text from the semantic representation is less serious since BLEU/NIST score is much increased if we do not consider silent sentences.

Our question is then, whether the semantic model is really an interlingual semantic representation. If we consider other language pairs, the number of the shared semantic representation would be decreased and this might reflect weakness in the expressivity of the interlingual representation. As it is mentioned in Mitamura et al. (2004), "it is difficult to design a perfect interlingual representation that covers all known languages, and there is no universally acceptable interlingual representation currently in existence". As we mentioned in section 4, when sentence pairs are semantically non-isomorphic between given language pair, we would use semantic transfer rules and we believe that this would be the best way to remedy weakness in the expressivity[7].

Acknowledgements

Authors would like to thank to four anonymous reviewers for their helpful remarks.

Bibliography

Christian Boitet. 2005. Gradable Quality Translations Through Mutualization of Human Translation and Revision, and UNL-Based MT and Coedition. In Jesús Cardeñosa, Alexander Gelbukh, Edmundo Tovar (Eds.) *Universal Networking*

[7] Using transfer rules, we increase BLEU and NIST scores for machine translation from French to English by 19% and 14% which might fill up the possible deficiency of our semantic representation (Park et al. 2007).

Language: Advances in Theory and Applications. Special issue of *Research on Computing Science*.

George Doddington. 2002. Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics. In *Proceedings of the Second Conference on Human Language Technology* (*HLT-2002*). San Diego, CA. pp. 128-132.

Karin Kipper, Martha Palmer. 2000. Representation of Actions as an Interlingua. In *Proceedings of the Third Workshop on Applied Interlinguas*, held in conjunction with ANLP-NAACL 2000.

Estelle Maillebuau. 2002. *Sémantique lexicale et dictionnaires électroniques*. Rapport de stage de DESS. Université Nancy 1&2.

Igor Mel'čnk. 1988. Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria. In *International Journal of Lexicography* 1988, volume 1(3): pp.165-188.

Teruko Mitamura, Keith Miller, Bonnie Dorr, David Farwell, Nizar Habash, Stephen Helmreich, Eduard Hovy, Lori Levin, Owen Rambow, Florence Reeder, Advaith Siddharthan. 2004. Semantic Annotation for Interlingual Representation of Multilingual Texts. In *Proceedings of the Workshop on Beyond Named Entity Recognition Semantic labelling for NLP*, LREC'04, Lisbon, Portugal. Pages 19-24.

Eric Nyberg, Teruko Mitamura, Kathryn Baker, David Svoboda, Brian Peterson, Jennifer Williams. 2002. Deriving Semantic Knowledge from Descriptive Texts Using an MT System. In *Proceedings of AMTA 2002*.

Bill Ogden, Jim Cowie, Eugene Ludovik, Hugo Molina-Salgado, Sergei Nirenburg, Nigel Sharples, Svetlana Sheremtyeva. 1999. CRL's TREC-8 Systems: Cross-Lingual IR, and Q&A. In *Proceedings of the 8th Text Retrieval Conference* (TREC-8).

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (*ACL2002*), Philadelphia, July 2002, pp. 311-318.

Jungyeul Park, Emilie De Guimier, Jérôme Vinesse. 2007. *Should we learn semantic transfer rules from a corpus for an interlingual MT system?* Internal technical report. France Télécom R&D, Lannion.

Projet Transat. 2006. *France Télécom's third-generation rule-based interlingual MT system*. Internal technical report. France Télécom R&D, Lannion.