A Lexical Ontology to Represent Lexical Functions

Alexsandro Fonseca[1], Fatiha Sadat[1], François Lareau[2]

[1]Université du Québec à Montréal, [2]Université de Montréal

[1]201 Président Kennedy, Montreal, Canada, [2]C.P. 6128 succ. Centre-Ville, Montreal, Canada

affonseca@gmail.com, sadat.fatiha@uqam.ca, francois.lareau@umontreal.ca

Abstract

Lexical functions are a formalism that describes the combinatorial, syntactic and semantic relations among individual lexical units in different languages. Those relations include both paradigmatic relations, i.e. vertical or "in absence", such as synonymy, antonymy and meronymy, and syntagmatic relations, i.e. horizontal or "in presence", such as intensification (*deeply committed*), confirmative (*valid argument*) and support verbs (*give an order, subject to an interrogation*). We present in this paper a new lexical ontology, called Lexical Function Ontology (LFO), as a model to represent lexical functions. The aim is for our ontology to be combined with other lexical ontologies, such as the Lexical Model for Ontologies (lemon) and the Lexical Markup Framework (LMF), and to be used for the transformation of lexical networks into the semantic web formats, enriched with the semantic information given by the lexical functions, such as the representation of syntagmatic relations (e.g. collocations) usually absent from lexical networks.

Keywords: Lexical Functions, Ontology, Lexical Relations, Lexical Network, Lexical Semantics, Collocations, Multiword Expressions

1. Introduction

We present in this paper an ongoing project that aims to represent the lexical functions (Mel'čuk, 1996) of the Meaning-Text Theory (MTT) (Mel'čuk, 1997) as a lexical ontology, called Lexical Functions Ontology (LFO).

A lexical ontology is a representation of the different aspects of the lexicon, such as meaning, morphology, part of speech, as well as the relation among lexical units, such as syntactic, semantic and pragmatic relations, using the semantic web formalisms (RDF/OWL languages).

Our objective in this project is to use this ontology in order to represent the relations among lexical units in lexical networks, especially in those networks based on lexical functions (LFs), such as the *Réseau lexical du français* (RLF) (Lux-Pogodalla and Polguère, 2011). However, it can also be used to represent different lexical relations in other lexical networks, such as WordNet in RDF/OWL format. This is an important aspect, since most of the existing lexical networks do not implement the syntagmatic information (Schwab et al., 2007) provided by some of the LFs. Moreover, we show how this model can be used to represent collocations in a lexical network, since the relation among lexical units in a collocation is a syntagmatic relation (Mel'čuk 1998).

We do not intend to recreate lexical representations already realized by previous works, such as lemon (McCrae et al., 2012), LexInfo (Buitelaar, 2009) or LMF (Francopoulo, 2007). Our proposal is to use, whenever possible, the lexical information already implemented by those models, such as the classes "LexicalEntry" and "LexicalSense" in the lemon model, and create the necessary classes for the implementation of lexical functions information.

2. Foundations and related work

We present in this section the theoretical information about lexical functions and related work.

2.1. Lexical functions

Bolshakov and Gelbukh (1998) defined a lexical function (LF) as a formalism for the description and use of combinatorial properties of individual lexemes. A more technical definition, given by Mel'čuk (1998), says that a "*Lexical Function f* is a function that associates with a given lexical unit *L*, which is the argument, or keyword, of *f*, a set $\{L_i\}$ of (more or less) synonymous lexical expressions – the *value* of *f* – that are selected contingent on *L* to manifest the meaning corresponding to *f*:

$$f(L) = \{L_i\}$$

The LFs considered in this paper are the standard ones, differentiated from the non-standard by the fact that the former can be coupled with a higher number of possible keywords and value elements (Mel'čuk 1998). For example, the LF *Magn*, which represents the sense 'intensification', can be coupled with many keywords (e.g. $shave_N$, *easy*, *to condemn*, *naked*, *thin*, *to rely*, and many others) to give different values : Magn(*shave*) = {*close*, *clear*}; Magn(*easy*) = {*as pie*, *as 1-2-3*}; *Magn*(*to condemn*) = *strongly*; Magn(*naked*) = *stark*; Magn(*thin*) = *as a rake*; *Magn*(*to rely*) = *heavily*; (Mel'čuk 1998). On the other hand, the sense "*additionné de…"* (with the addition of…) is a non-standard LF in French, because it can only be coupled with a few number of keywords (*café*; *fraises*; *thé*), to create the expressions: *café crème*, *fraises à la crème* (and not **café à la crème*, **fraises crème*); *café au lait*; *café arrosé*; *café noir*; *thé nature*; etc (Mel'čuk 1992).

About 70 simple standard LFs have been identified (Kolesnikova, 2011). Complex LFs are formed by the combination of simple standard ones.

LFs can be classified as paradigmatic or syntagmatic, according to the kind of lexical relation they model. Figure 1 illustrates the difference between the two kinds of relations. The paradigmatic LFs model the vertical, "in absence" or "in substitution" relation among lexical units (Saussure, 1983). For example, antonymy, Anti(*big*) = *small*; synonymy, Syn(*car*) = *automobile*; hyponymy, Hypo(*feline*) = {*cat*, *tiger*, *lion*, etc.}. Syntagmatic LFs model the horizontal, "in presence" or "in composition" relations among lexical units (Saussure, 1983). For example: magnification, Magn(*committed*) = *deeply*; confirmation, Ver(*argument*) = *valid*; laudatory, Bon(*advice*) = {*helpful*, *valuable*}. And simple standard LFs can be combined to form complex ones. For instance, AntiBon(*criticism*) = *harsh*; AntiMagn(*similarity*) = *vague*.

		the cake	
	buys	the cheese	
The man	sells	the bread	↑ paradigmatic relations
The woman	eats	her/his food	

syntagmatic relations ➔

Figure 1: The difference between syntagmatic and paradigmatic relations. There are syntagmatic relations between "*The man*" and "*buys*", "*buys*" and "*the cheese*", "*The man*" and "*sells*", "*The woman*" and "*buys*", etc. There are paradigmatic relations between "*the cake*" and "*the cheese*", "*the cheese*" and "*the bread*", etc., which belong to the same syntactic/semantic fields and can substitute each other in a phrase.

Another important concept is that of semantic actant (Sem-actant) (Mel'čuk, 2004). In logic, a predicate is a falsifiable assertion. Each predicate has one or more arguments. For example, in the assertion "Rome is the capital of Italy", we can define the predicate 'capital' having two arguments, 'Rome' and 'Italy': *capital*(*Italy*, *Rome*).

In linguistics, the predicate is called "predicative sense" and the arguments are its "semantic actants". Each LF represents a different predicative sense and the semantic actants are represented by subscripts. For example, the LF *S* gives the equivalent noun of the value to which it is applied. S_1 gives the first actant (the one who executes the action), S_2 gives the second actant (the object of the action) and S_3 gives the third actant (the recipient of the action): S_1(*to teach*) = *teacher*; S_2(*to teach*) = {*subject*; *matter*}; S_3(*to teach*) = {*pupil*; *student*}. Other subscripts give circumstantial information. For example: S_{loc} – local of the action/event; S_{instr} – instrument used; etc.

LFs can be classified according to their semantic or syntactic behaviour. For example, in (Mel'čuk, 1998) we find the following classification:

- Semantic derivatives: S_1(*to teach*) = *teacher*; S_3(*to teach*) = *pupil*; S_{loc}(*to fight*) = *battlefield*; S_{instr}(*murder*$_{V,N}$) = *weapon;* A_1(*anger*$_N$) = *angry;* Adv_1(*anger*) = *angrily*;
- Semi-auxiliary verbs: $Oper_1$(*support*) = [*to*] *lend* [~ *to N*]; $Oper_1$(*promise*$_N$) = [*to*] *make* [ART ~]; $Func_2$(*proposal*) = *concerns* [*N*];
- Realization verbs: $Real_1$(*bus*) = [*to*] *drive* [ART ~]; $Real_2$(*bus*) = [*to*] *ride* [*on* ART ~]; $Real_1$(*promise*$_N$) = [*to*] *keep* [ART ~];
- Modifiers: Magn(*injury*) = *serious*; Ver(*citizen*) = *loyal*; Ver(*argue*) = *convincingly*; Bon(*analysis*) = *fruitful*.

In this work, we combine different classifications of the LFs, especially those presented by Mel'čuk (1998) and Mel'čuk et al., (1995) to create the classes of our ontology.

2.2. lemon model

lemon (McCrae et al., 2012) is a model for sharing lexical information on the semantic web. It is based on earlier models, such as LexInfo (Buitelaar, 2009) and LMF (Francopoulo, 2007). As its main advantages over these previous models, we cite:

- separation between the linguistic and the ontological information;
- linguistic information, such as "partOfSpeech" and "writenForm" are represented as RDF properties, differently of LMF, which represent them as attributes of a property, which makes easier the use of other resources, like the SPARQL query language;
- lemon uses ISOCat, data categories homologated by ISO (for example, "partOfSpeech", "gender" and "tense");
- lemon is an easily extensible model;
- there are already many linguistic resources in lemon format, like WordNet and DBPedia Wiktionary.

Lexical units are represented in the lemon model using the classes "LexicalEntry" and "LexicalForm". The "LexicalEntry" class is connected to the lexical unit sense, which is represented by the "LexicalSense" class. The connection between the lemon model and external ontologies are made through this last class.

In our model, the keyword and the value of a LF will be represented as a lemon "LexicalSense" class. In MTT, the different senses of a word are represented by subscripts, using Roman and Arabic numbers and Latin letters (Mel'čuk 1995), which we illustrate here with an example. Consider the word "*ocean*". It has concrete senses, like "*a body of water that covers the planet*" and abstract senses, like in "*ocean of people*". In MTT, the concrete senses of "*ocean*" would be represented as "$Ocean_I$" and the abstract senses as "$Ocean_{II}$". Inside "$Ocean_I$" we could have subdivisions:

- $Ocean_{I.1a}$: "*extension of water that covers the planet*" (always in singular, referring to the entire body of water);
- $Ocean_{I.1b}$: the set of oceans in general (always in plural) – "*the oceans are becoming more polluted.*";
- $Ocean_{I.2}$: a part of $Ocean_{I.1a}$ in a specific region – *Atlantic Ocean, Pacific Ocean, Arctic Ocean*, etc.

In our model, the word "*ocean*" is represented by a lemon object "LexicalEntry" and $Ocean_{\mathrm{I.}}$, $Ocean_{\mathrm{I.1a}}$, $Ocean_{\mathrm{I.1b}}$, $Ocean_{\mathrm{I.2}}$ and $Ocean_{\mathrm{II}}$ are each represented by a "LexicalSense" lemon object. The reason for this is explained as follows: the semantic connection represented by an individual LF is between senses, and not between lexical forms or lexical entries. Doing so, we can have an already disambiguated lexical network when connecting lexical units with a LF.

2.3. ILexicOn lexical ontology

Lefrançois and Gandon (2011) present a lexical ontology based on MTT for the construction of a dictionary.
Their approach is based on a three layers architecture:
- the meta-ontology layer;
- the ontology layer;
- the data layer.

The meta-ontology layer is formed by what the authors call meta-classes, which are super classes for the classes in the ontology layer. For example, the meta-class "ILexicalUnit" is a super class of all types of lexical units and the meta-class "ISemanticRelation" is a super-class for all the semantic relations appearing in the ontology layer. The ontology layer is formed by classes that represent concepts, such as "Entity", "Person" and "State". They are connected to the meta-class "ILexicalUnit" by a "is-a" relation and to each other by semantic relations that are instances of the ISemanticRelation meta-class. The data layer contains instances of the classes in the ontology layer. For example, "Mary01" can be an instance of the class "Person" and "Alive01" an instance of the class "Alive".
The authors justify those layers saying that this ensures three of the four redaction principles of an explanatory and combinatory dictionary (Mel'cuk et al, 1995), in MTT: the principles of formality, internal coherence and uniform treatment. The principle of exhaustivity is not ensured. In their model, the collocations and locutions are represented as dictionary entries of the keyword's collocation or locution.
The difference between their work and our model is that ILexicOn is intended to represent an entire dictionary following the Meaning-Text Theory precepts, while our model is intended to represent lexical relations in a lexical network. Moreover, using ILexicOn, collocations and locutions are represented as dictionary entries, while with our model, they will be represented as a graph, representing connections between lexical units.

3. The LFO Model

Figure 2 illustrates the LFO core model. The central class in our model is the "LexicalRelation". It connects to the LexicalFunction class, to the lexical relation type (which can be paradigmatic or syntagmatic), and to the value and to the key of the lexical relation (LR).
We decided to connect the LF keyword and the LF value using an intermediate class (LexicalRelation), instead of connecting them directly with the LexicalFunction class because in this way we can connect to the LexicalRelation information that is specific to the relation between two lexical units, independently of the LF connection them, and we can connect to a LF information that is independent of the lexical units that it connects. Also, the paradigmatic/syntagmatic information (LRType) is connected to the LexicalRelation class instead of being connected to the LexicalFunction class. Although the LFs usually have a definite type (paradigmatic or syntagmatic), some of them do not have, which will depend on the lexical units they model.
Figure 3 illustrates how the collocation "*close friend*" would be represented. It is modelled by the LF Magn (predicative sense = intensification): Magn($friend_{\mathrm{I.1}}$) = $close_{\mathrm{III.1a}}$; Since also Magn($friend_{\mathrm{I.1}}$) = $good_{\mathrm{II}}$, we could have another LexicalRelation (*Magn_02*) connecting the LexicalSense $good_{\mathrm{II}}$ and the LexicalSense $friend_{\mathrm{I.1}}$.

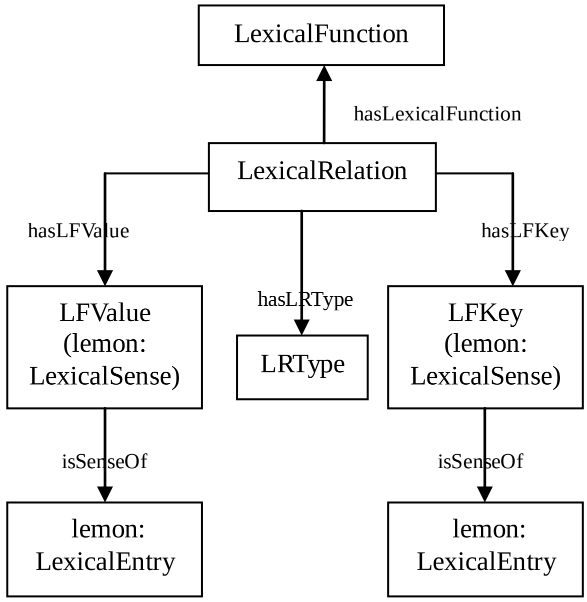

Figure 2: The LFO core model combined with the LexicalSense and LexicalEntry lemon objects

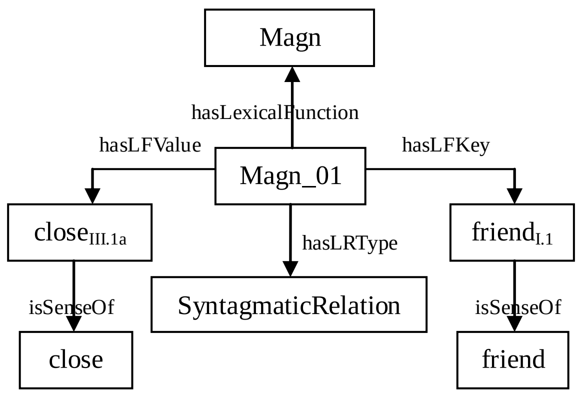

Figure 3: The representation of the collocation "*close friend*"

The lexical relation is connected to the value of the collocation using the property "hasLFValue" and to the keyword using the property "hasLFKey". The property "hasLRType" informs that the relation between

"$close_{\mathrm{III.1a}}$" and "$friend_{\mathrm{I.1}}$", modelled by the LF "Magn", is a syntagmatic relation.

Example 1 illustrates a possible RDF/OWL code of our previous example. As explained in Section 2.2, it is important to note that the lexical units that appear in our example, "$friend_{\mathrm{I.1}}$", and "$close_{\mathrm{III.1a}}$" will be modeled as "LexicalSense" and not as a "LexicalEntry" lemon object. This means that our model will connect to the lemon model via the sense of the lexical units. This allows the construction of already disambiguated lexical networks. Finally, the lexical variations (e.g. plural) can be treated at the level of the LexicalEntry lemon object, already implemented by the lemon model.

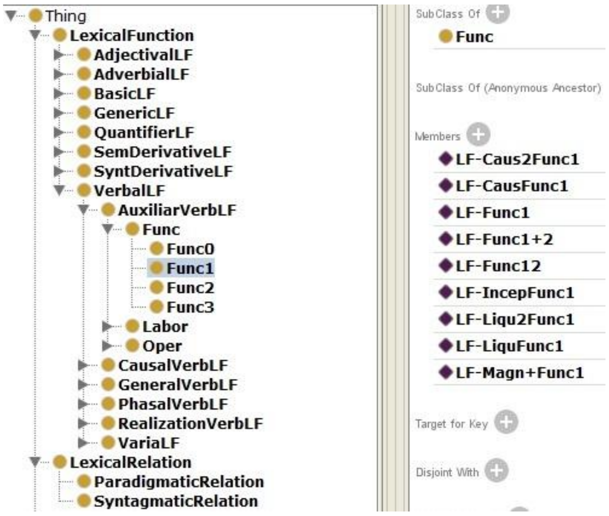

Figure 4: An overview of the LFs and some complex LFs implemented as members of the class *Func*$_1$

```
:close a lemon:LexicalEntry.
:closeIII.1a
    a lemon:LexicalSense;
    lemon:isSenseOf :close.
:friend a lemon:LexicalEntry.
:friendI.1
    a lemon:LexicalSense;
    lemon:isSenseOf :friend.
:LexicalFunction rdf:type owl:Class.
:AdjectivalLF rdf:type owl:Class;
    rdfs:subClassOf :LexicalFunction.
:Magn rdf:type owl:Class;
    rdfs:subClassOf :AdjectivalLF.
:LF-Magn rdf:type :Magn, owl:NamedIndividual.
:LexicalRelation rdf:type owl:Class.
:SyntagmaticRelation rdf:type owl:Class;
    :rdfs:subClassOf :LexicalRelation.
:Magn_01 a lfo:LexicalRelation;
    lfo:hasLRType  lfo:SyntagmaticRelation;
    lfo:hasLFValue :closeIII.1a;
    lfo:hasLFKey :friendI.1;
    lfo:hasLexicalFunction lfo:LF-Magn;
```

Example 1: The RDF/OWL representation, using the Turtle notation, of the collocation "*close friend*"

Figure 4 is an overview of how the simple LFs are organized by OWL classes and how the complex LFs are modelled as members of a class. For example, the LF *Func*$_1$ has the following members: CausFunc$_1$, Func$_1$, Func$_{1+2}$, Func$_{12}$, etc. In total, about 600 simple and complex LFs are already represented in our model, which were extracted from a relational database representation of the RLF (Lux-Pogodalla and Polguère, 2011).

4. Conclusion and Future Work

We presented in this paper an ongoing project, called Lexical Function Ontology (LFO), aimed at the representation of the lexical functions of Meaning-Text Theory as a lexical ontology.

Most of the existing lexical networks lack important semantic information, especially the syntagmatic relations among lexical units. Lexical functions are a powerful tool for the representation of linguistic relations. In particular, syntagmatic lexical functions can fill the present gap in the representation of syntagmatic relations in lexical networks.

Moreover, the combination of the descriptive logic embedded in the OWL language with the semantic and syntactic information provided by lexical functions creates a strong tool for studying human reasoning and for interesting psycholinguistic studies.

Finally, this work can be seen as a new form for the representation of multiword expressions.

As future work, we intend to complete the representation of lexical functions in our model with the combinatorial, semantic, syntactic, and communicational perspectives presented by Jousse (2010). The implementation of such classification perspectives will allow invaluable semantic, syntactic and pragmatic information to be coded directly in a lexical network. For example, the LF *CausFunc1* has as perspectives:

- combinatorial: ThreeActants;
- part of speech of the value of the function: verb;
- semantics: represents a cause;
- target: the target of the function is its first actant;

This information can be connected to each LexicalFunction using owl:objectProperties. For example:

```
:CausFunc1 hasCombinatorialPersp :ThreeActants.
:CausFunc1 hasSemanticPersp :CauseSemPersp.
:CausFunc1 hasValuePOS isoCat:Verb.
:CausFunc1 hasTarget :FirstActant.
```

Also as a future work, we intend to use our model to transform the *Réseau Lexical du Français* (Lux-Pogodalla and Polguère, 2011), from its present relational database format to an ontology format (already in progress).

5. Acknowledgments

We want to thank Professor Alain Polguère for providing us with the datasets from RLF and the reviewers for helpful comments and suggestions.

6. Bibliographical References

Bolshakov, I. and Gelbukh, A. (1998). Lexical functions in Spanish. *Proceedings CIC-98, Simposium Internacional de Computación*, pp. 383-395. November 11-13, 1998, Mexico D.F.

Buitelaar, P., Cimiano, P., Haase, P. et Sintek, M. (2009). Towards linguistically grounded ontologies. In *L. Aroyo et al. (Eds.): ESWC 2009, LNCS 5554*, pp. 111-125, Spring-Verlag Berlin. Heidelberg 2009.

Francopoulo, G., Bel, N., Georg, M., Calzolari, N., Monachini, M., Pet, M. and Soria, C. (2007). Lexical markup framework: ISO standard for semantic information in NLP lexicons. In: *Proceedings of the Workshop of the GLDV Working Group on Lexicography at the Biennial Spring Conference of the GLDV*.

Jousse, A. (2010). Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales. Thèse de doctorat. Université de Montréal et Université Paris Diderot (Paris 7), 340 p.

Kolesnikova, O. Automatic extraction of lexical functions, PhD Thesis (2011). Instituto Politecnico Nacional – Centro de Investigacion en Computacion, Mexico, D.F., Mexico, 116 p.

Lefrançois, M. et Gandon, F. (2011). ILexicOn: Toward an ECD-compliant interlingual lexical ontology dscribed with semantic web formalisms. In *5th International Conference on Meaning-Text Theory (MTT '11)*, September 2011, pp. 155-164, Barcelona, Spain.

Lux-Pogodalla, V. and Polguère, A. (2011). Construction of a French lexical network: Methodological issues. In *Proceedings of the First International Workshop on Lexical Resources, WoLeR'11.* An ESSLLI 2011 Workshop, pp. 54-61, Ljubljana, Slovenia.

McCrae, J., Aguado-de-Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D. and Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Lang Resources & Evaluation (2012)* 46:701–719.

Mel'čuk, I. (2004). Actants in semantics and syntax, I: Actants in semantics. *Linguistics* 42(1):1-66.

Mel'čuk, I. (1998). Collocations and lexical functions - *A.P. Cowie (Ed.), Phraseology. Theory, Analysis and Applications*, Oxford: Clarendon Press, pp. 23-53.

Mel'čuk, I. (1997). Vers une linguistique Sens-Texte. Leçon inaugurale. Paris: Collège de France, 78 p.

Mel'čuk, I. (1996). Lexical functions: A tool for the description of lexical relations in the lexicon. In *L. Wanner (ed.): Lexical Functions in Lexicography and Natural Language Processing*, pp. 37-102, Amsterdam/Philadelphia: Benjamins.

Mel'čuk, I., Clas, A. et Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*, Coll. Champs linguistiques/Universités francophones, Louvain-la-Neuve/Paris, Éditions Duculot/AUPELF-UREF, 256 p.

Mel'čuk, I. (1992). Paraphrase et lexique: La théorie Sens-Texte et le *Dictionnaire explicatif et combinatoire* in Mel'cuk *et al*. 1992: 9-58.

Saussure, F. de (1983). Course in general linguistics. Eds. Charles Bally and Albert Sechehaye. Trans. Roy Harris. La Salle, Illinois: Open Court. 1983, 236p.

Schwab, D., Tze, L. L. et Lafourcade, M. (2007). Les vecteurs conceptuels, un outil complémentaire aux réseaux lexicaux. TALN'07: Traitement Automatique des Langues Naturelles, pp. 293-302, Jun 2007, Toulouse, France, ATALA.