Using Constraints on a general Knowledge lexical networK for domain-specific semantic relation extraction and modeling

Nadia Bebeshina-Clairet, Lionel Ramadier, Mathieu Lafourcade

Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2016"

Moscow, June 1–4, 2016

USING CONSTRAINTS ON A GENERAL KNOWLEDGE LEXICAL NETWORK FOR DOMAIN-SPECIFIC SEMANTIC RELATION EXTRACTION AND MODELING

Clairet N. (clairet@lirmm.fr)[1,2,3],
Ramadier L. (ramadier@lirmm.fr)[1,4],
Lafourcade M. (mathieu@lafourcade@lirmm.fr)[1]

[1]Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Montpellier, France

[2]Laboratoire d'Informatique Médicale et d'Ingénieurie des Connaissances en e-Santé (LIMICS), Paris France

[3]Lingua et Machina, Boulogne-Billancourt, France

[4]IMAIOS, 672, Montpellier, France

We introduce a pattern-based approach applied to the semantic relation retrieval and semantic modeling. Our method relies upon the use of a general knowledge lexical semantic network built, shaped, and handled by crowdsourcing and GWAPs (games with a purpose). Implementing constraints on semantic relations available in the network increases the efficiency of the relation extraction process but also opens a semantic modeling perspective. In terms of (mostly horizontal) relation extraction, we tested our method on radiology reports in French. Our results show the interest of using a general knowledge lexical semantic network for the domain specific textual analysis as well as the interest of implementing series of constraints on semantic relations for the relation retrieval. We recently turned to the analysis of cooking recipes that stand for examples of domain specific instructional texts. Thus, in addition to the semantic relation discovery, we are building a method for the semantic modeling and conceptualization of cooking instructions. Its first results are presented below. Today, our results are available for French but we target extending the lexical network coverage to other languages in the next few years.

Keywords: semantic relation retrieval, semantic modeling, domain specific raw text analysis, lexical-semantic network

ИСПОЛЬЗОВАНИЕ ОГРАНИЧЕНИЙ В ЛЕКСИКО-СЕМАНТИЧЕСКОЙ СЕТИ ДЛЯ ИЗВЛЕЧЕНИЯ СЕМАНТИЧЕСКИХ ОТНОШЕНИЙ ИЗ СПЕЦИАЛИЗИРОВАННЫХ ТЕКСТОВ И СЕМАНТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Клере Н. (clairet@lirmm.fr)[1,2,3],
Рамадье Л. (ramadier@lirmm.fr)[1,4],
Лафуркад М. (mathieu.lafourcade@lirmm.fr)[1]

[1]Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Монпелье, Франция

[2]Laboratoire d'Informatique Médicale et d'Ingénieurie des Connaissances en e-Santé (LIMICS), Париж, Франция

[3]Lingua et Machina, Boulogne-Billancourt, Франция

[4]IMAIOS, Монпелье, Франция

Мы предлагаем подход к извлечению семантических отношений из неструктурированных текстов и семантическому моделированию основанный на использовании лексико-семантической сети общих знаний и семантических шаблонов. Используемая лексико-семантическая сеть развивается и поддерживается пользователями при помощи целевых онлайн игр и прямого участия (краудсорсинга). Специфика её структуры и применение ограничений (правил) к дугам сети повышают эффективность анализа семантики текста, а также открывают новые перспективы для (полу)автоматического семантического моделирования и концептуализации. Мы протестировали наш метод на материале корпуса рентгенологических заключений на французском языке с целью извлечения неиерархических отношений. Параллельно с исследованием этой проблематики, мы используем нашу методику для разработки модели кулинарного рецепта с целью концептуализации так как данный материал является примером процедурального текста. В настоящее время наши результаты касаются французского языка но мы планируем расширить нашу лексико-семантическую сеть включив в неё другие языки и, вместе с ней, возможности семантического анализатора.

Ключевые слова: семантический анализ, извлечение неиерархических семантических отношений, семантическое моделирование, анализ неструктурированного текста

Introduction

In recent years, the semantic analysis of the domain specific texts has been conducted on the basis of specific resources often without using any general knowledge repository. Instead of building a specific resource for our analysis, we immersed the domain specific knowledge into a general knowledge lexical semantic network. Then, we added constraints on the relations present in the lexical semantic network which improved our relation extraction results. Finally, we targeted an instructional text perspective and extend our approach to the modeling the sequences of actions corresponding to the cooking instructions. The paper is structured as follows. First, we give an overview of the state of the art and detail the resource we use for our experiments. Second, we introduce the IMAIOS system for semantic relation extraction and its results. Finally, we propose the MAKI system for the instructional text analysis and semantic modeling.

1. State of the art

In the literature, the "semantic relation" term corresponds to a variety of definitions. In the lexicographic perspective ex. Wordnet (Fellbaum, 1998), semantic relations are above all taxonomy relations (hyperonymy, hyponymy, meronymy). The semantic role approach understands semantic relations as roles handled by the terms in a context as proposed by (Dong et al., 2006) and described by (Morris and Hirst, 2004). Lexico-semantic networks such as BabelNet (Navigli and Ponzetto, 2012), ConceptNet (Liu and Singh., 2004), and JeuxDeMots (Lafourcade, 2007)[1] implement this kind of approach.

Most research work concerning the extraction of semantic relations focus on domain-independent relations (Snow et al., 2006; Chklovski and Pantel., 2004). In the paradigm of information retrieval, open information extraction systems such as (Banko and al., 2009) are also able to retrieve unknown relations. In the biomedical domain, there are four main techniques for relation extraction: finding co-occurrences (Jelier et al., 2005), using patterns or rules (Auger et al., 2008; Song et al., 2015; Rindflesch et al., 2000), implementing supervised learning-based approaches (Song et al., 2015; Rink et al., 2011), and using hybrid approaches (Suchanek et al., 2006; Chowdhury et al., 2012). The relation extraction between verbs (based on their arguments) yielded a number of methods. In particular, those of (Brody, 2007) and (Chambers and Jurafsky, 2008) focus respectively on arguments' role and discourse relations and the real order of actions instead of following the textual order.

---

[1] As a GWAP, JeuxDeMots includes a number of games for lexical acquisition and validation. The main word game involves two players who are asked questions to populate some knowledge type (ex. *What is the typical characteristic of cake*?) their answers are recorded and compared, the terms that appear in both answers are validated and either they are added into the network or their weight is augmented. More details are available in (Lafourcade, 2007).

The approaches to the semantic analysis of the cooking recipes fall into two major trends. The first one is centered on the concept of aliment, its possible features and adaptation perspectives. The IBM Chef Watson system implements this kind of approach. The second trend is related to the case-based reasoning paradigm. It uses cooking recipes as examples of instructional texts with the scope of analyzing and modeling work-flows. The Taaable (Badra et al., 2008) project is being developed in this perspective. Cooking instructions can be considered as short texts such as defined by (Pedersen, 2008[2]). Their analysis can also be conducted in the *mapping instructions to actions* perspective. In this paradigm the instructional text can be represented in formal language (Chen and Mooney, 2011), approahced from the machine learning linear policy estimation (Vogel and Jurafsky, 2010) in particular semantic role labeling (Malmaud et al., 2014) or alignment-based compositional semantics (Andreas and Klein, 2014) perspectives. In the knowledge engineering domain, a number of dedicated resources and models have been developed for food and nutrition: BBC Food Ontology[3], PIPS Food ontology[4], SOUR CREAM (Tasse and Smith, 2008) etc.

2. Crowd-sourced lexical semantic network

The lexical semantic network JeuxDeMots (Lafourcade, 2007) is an oriented, typed, weighted graph that contains 40M arcs (relations) linking 800K nodes. Unlike some of the similar resources (such as Wiktionnary), the JeuxdeMots graph features more than 100 relation types and includes various kinds of lexical, morphological, and semantic information. Therefore, it is relevant for the mining of horizontal relations such as *location*, *part-of*, *synonym*, causal and temporal relations, *characteristic*, *manner*, and more. The JeuxDeMots graph also implements inference and annotation schemes. The inference scheme (Zarrouk, 2015) for the graph semantic relations spreading introduces deductive, inductive, abductive and refinement approaches. The deduction and the induction mechanisms test the assumption of the transitivity of the is-a relation and use the logical blocking in case of polysemy. The blocking scheme is based upon the refinement concept and the annotation process applied to the premisses (typed and weighted outgoing relations inferred on the basis of hyperonyms/hyponyms of a term). Its final stage is the validation processed by a human expert. The refinement corresponds to the "real life" use of a term and may stand quite far from its lexicographic representation. This feature of our resource helps untangling some knotty problems such as the multi-word term detection and the polysemy management while analyzing raw text. The abduction scheme uses

[2] "A short written context consists of one to approximately 200 words of text that is presented to a human reader as a coherent source of information from which a conclusion can be drawn or an action taken"(Pedersen, 2008).

[3] http://www.bbc.co.uk/ontologies/fo/1.1

[4] http://cordis.europa.eu/project/rcn/71245en.html

synonym and similarity relations in order to infer new relations in the graph. To benefit from the inference scheme, a term is supposed to have at least one hyperonym, hyponym, synonym or refinement. The annotation scheme (Ramadier et al., 2014) allows defining and spreading the annotations over the relations already existing in the graph. It amplifies the inference scheme described above without adding new relation types into the graph.

The features of our resource grant the use of robust and somehow crude mining algorithms. The identification of compound terms can be made upstream by comparing the entities found in the text to the JeuxDeMots network. We use underscore to aggregate the two parts of a compound word. Thus, it is considered as an entity by the extractor (tibia_fracture). The polysemy resolution is based on the refinement. In the graph, all the refinements are linked to the core term. The disambiguation process is based on a triangular scheme including the term to disambiguate, its context and its refinements. The context is compared to the term refinements found in the graph. If a refinement do have some relation with the context (has at least one ingoing or outgoing relation to/from the context terms and these relation(s) has a weight w > 50) it is retained as it expresses the sense close to the context. The weight of a relation corresponds to the force of association between two words: how many crowdsourcers figured out the second term while considering the first one.

3. Extracting semantic relations from domain specific unstructured text: the IMAIOS system

The relation extraction approach implements series of semantic patterns. We understand semantic patterns as linguistic patterns similar to (Embarek and al, 2008) coupled with series of constraints on the relations of the JeuxDeMots graph. In the scope of the radiology report analysis and indexation and after being advised by radiologists, we have chosen 15 semantic relations relevant within this domain. These are in particular *r_isa* (generic terms), *r_synonym* (synonyms or quasi synonyms), *r_carac* (typical characteristics), *r_location* (typical location), *r_target* (disease target such as social group, organ), *r_part_of* (typical parts), *r_cause* (typical causes) etc. These relations can be of any general purpose. Some authors have already noticed that the use of patterns is an effective method for automatic information extraction from corpora if they are efficiently designed (Embarek et al., 2008; Cimino et al., 1993). For each relation type, we build patterns and match them with the sentences to identify the correct relation. These patterns are (for now) manually built through partial analysis of our corpus. In our experiment, we restricted ourselves to 42 semantic patterns, 12 of which are specific to medicine.

Table 1. Semantic pattern examples

Relations	Examples of patterns in English	Exemples of patterns in French
location	E1 on the level of E2	E1 au niveau de E2
location	E1 in E2	E1 dans E2
location	E1 is in the E2	E1 se trouve dans E2
location	E1 passing by E2	E1 passant par E2
causes	E1 may trigger E2	E1 déclenchant E2
characteristic	E1 is characterized by E2	E1 est caractérisé par E2
characteristic	Noun Adj	Nom + Adjectif
synonym	E1 also called E2	E1 encore appelé
causes	E1 can produce E2	E1 peut produire E2
consequence	E1 causes E2	E1 provoque E2
hyperonym	E1 is a E2	E1 est un E2
consequence	E1 leading to a E2	E1 menant à E2
target	E1 touching E2	E1 touchant E2
treatement	E1 treated by E2	E1 traité par E2
clinical sign	E1 accompanied by E2	E1 accompagné par E2

For some of the relations listed above, we encountered difficulties related to the ambiguity issue. For the location relation, we can distinguish two kinds of possible semantic relations depending on the pattern. The first pattern refers to the *r_location* relation (hepatocellular carcinoma is at the level of the liver). The second relation is holonomy (femur *r_holo* lower limb). For some connectors (of in caudate lobe of liver) both relations are correct (caudate lobe *r_location* liver and caudate lobe *r_holo* liver). We also make use of immediate co-occurences of entities for characteristic relation. For instance mutifocal hepatocellular carcinoma (HCC) appears five times together, so we consider multifocal as a probable characteristic of HCC (HCC *r_characteristic* multifocal).

Some linguistic patterns are inexpressive and it is hardly possible to determine the kind of the associated relation (ex. the french connector *de*, "of, from, because of"). Thus, we have added some semantic constraints on linguistic patterns. A semantic constraint is a condition that should verify the reification of one of variable of the pattern. There may be any number of constraints on $x and $y. Basically, a semantic constraint is a rule defined as follows: "if x is related to B than x is related to C" or x, Rb(x, B) =>Rc(x,C).

Table 2. Semantic pattern structure

Semantic pattern structure	Example
Lexical pattern (linguistic scheme)	"$x de $y"
Premisses (conditions or semantic rules)	*If $x r_isa illness & $y r_isa anatomical_location*
=> Conclusion (action)	$x r_location $y

To apply the semantic constraints' principle over our corpus and extract semantic relations, we use the following algorithm:

Let *S* the result set, being the empty set at initialization
Finding pattern occurrence in the text by moving a word window of size *n*
For each pattern occurrence applying constraints to the instantiated variables
If constrains are verified then the associated semantic relation is associated to $x and $y, that is to say added to *S*
Return S

From a corpus of more 30,000 medical reports, we extracted a random subset of around 120,000 relation instances for the different relation types[5]. About 800 of these relations were manually checked for evaluating precision. For assessing recall, we manually identified the relations in about 300 medical reports. Then we applied our algorithm.

Table 3. The IMAIOS system: relation extraction results

Relations	***Precision w/o constraints***	**Precision with constraints**	**Recall**	***F-measure w/o constraints***	**F-measure with constraints**	**Contribution of the IMAIOS method (F-score)**
cause	*74%*	90%	60%	*66%*	72%	+6.0
consequence	*70%*	89%	62%	*63.4%*	73%	+9.6
location	*48%*	83%	40%	*43.6%*	54%	+10.4
treatment	*70%*	88%	60%	*64.6%*	71.3%	+6.7
part-of	*32%*	75%	30%	*31%*	42.9%	+11.9
target	*45%*	80%	40%	*42.4%*	53.3%	+10.9
characteristic	*60%*	88%	58%	*60%*	70%	+10.0
lieu	*45%*	86%	40%	*41.7%*	54.6%	+12.0

The IMAIOS system has also been applied to other corpora. For a corpus of 45,000 cooking recipes, 245,000 semantic relations have been extracted with a precision of 95% (manually evaluated on a sample of 755 relations). Furthermore, we extracted 789,000 relations for randomly Wikipedia pages with a precision of 92% (manually evaluated on a sample of 1,250 relations). Hyperonym extraction on Wikipedia articles has a precision of about 94%.

[5] Even though we target some of the relation types according to our objectives, our system can extract any of the relation types present in the JeuxDeMots network on the basis of appropriate semantic patterns: taxonomic (*isa*, *hypo*, *has-part*), predicative (*agent*, *patient*), horizontal (*location*, *place*, *action place*, *instrument*, *manner*, *cause*, *consequence*, **qualia structure** (Pustejovsky, 1995) inspired relations (telic role, agentive role), and more.

4. Instructional text analysis, semantic relation extraction, and modeling

The MAKI System focuses on the analysis of cooking recipes taken as examples of instructional text, the extraction of temporal relations, and the modeling of the sequences of actions. Its ultimate goal is to build a conceptualizing work-flow to discover the canonical recipe on the basis of its variants found in the texts of recipes. It extends the use of semantic patterns by introducing the rules with not only two but more variables.

> $\forall X^n\ \forall x$, *patient*$(X^n, x) \wedge$ *is-a* $(x, aliment)$ $\exists y$, *pos* $(y, Nom) \wedge$ *carac*$(y, z) \wedge$ *pos*$(z, Ver{:}PP) \Rightarrow$ *consequence*$(X^n, y) \wedge$ *successeur_temps*(X^n, y) *cf. "for each transforming action of the instructional text there is a state such as the consequence and the temporal successor of this action".*

The graph corresponding to the representation of the cooking instructions as sequences of actions is a bipartite graph with two types of nodes: states and actions. Our analysis strategy prompts that of (Bonfante et al., 2010) and also that of (Poria et al., 2014). We move from a syntactic dependency surface representation that can be obtained by using a parser such as Bonsaï PCFG-LA parser and MElt (Denis et Sagot, 2009), or by using the JeuxDeMots graph which contains such information. For each segment of the text we build an oriented, typed acyclic graph such as

$$\lfloor G \rfloor \models \forall x \exists (y, z.\ (edg\,(x, y, r') \wedge edg\,(x, z, r''))).$$

From the linguistic point of view, the MAKI system focuses on the following phenomena (among other): predominance of predicative structures; monotony of the argumental field (from one textual segment to another, we rediscover the same arguments such as *patient*, *instrument*, *place*, *quantifier* etc.); adjectival value of the past participle forms visible when observing the characteristic semantic feature (ex. *légumes blanchis,* "blanched vegetables"). These observations corroborate the assumption that any recipe action is spatially situated (utensil, table, kitchen), transforming (each action is followed by some new state of ingredients (patients)), and temporary finite.

In terms of computation, we introduce the dynamic creation of entities (nodes and directed arcs). The general analysis process starts from a surface representation built using the syntactic relation typed as *succ* (successor). Then, a context free grammar introduces constraints on the JeuxDeMots relations. As the rules (semantic constraints) keep on being applied, the JeuxDeMots graph is browsed and the semantic relations between these nodes keep on being discovered. The reification mechanism defined by (Zarrouk, 2015) is implemented. In this scheme, the graphs of the variables are dynamically built, the nodes typed as "rules" are created and linked to the reifications of a corresponding relation. We extend this scheme as we allow the creation of nodes to type the graphs of variables which facilitates their comparison. The variable graphs are linked to generic alignment entities using the relation type *r_head*. The lexical shape of these generic entities is that of the domain key terms (and not conceptual entities specifically created for the analysis) present in the JeuxDeMots lexical semantic network. Thus, each generic alignment entity has a syntactic and semantic behavior and we can benefit from the power of our resource and its range of relation types. Main entities of this kind are: *action*, *état* (state), *préparation*, *ingrédient*. Moreover, we use such elements as grammatical features (parts of speech, gender, number etc.), word order, antecedence and reference markers in the text.

For the modeling of sequences of actions, we consider the instructional graphs built during our recipe analysis. We assume that actions pertaining to the same sequence share a certain number of resources (arguments) such as *patient* (ingredient), *instrument*, *place* (utensils, recipients). Therefore, graphs that do not have any of such resources in common and do not have any similar nodes (synonyms or quasi-synonyms otherwise linked by the relations typed as *part-of, substance, place of action, location*) are not part of a same sequence hyper-graph. We also noticed that the shortest path to the resources or groups of resources in the graph could be an indicator of the order of actions in a sequence.

To test our modeling system, we applied our scheme to 1,500 cooking instructions selected according to their predicative structure (at least one predicate and one patient), grammatical correctness, and length (4 to 12 tokens). For the evaluation purposes, we manually extracted semantic relations from our corpus (we obtained 3,878 relations) and built the generic alignment entities (1,720 entities). Then, we compared our semantic parser's results to this reference. The actions behind the textual instructions have been modeled exactly in 57% of cases (855 graphs have been built with all the expected nodes and all the appropriate semantic relations). 28% of instructions have been partly modeled (which corresponds to 435 instructions modeled). Finally, in 14% of cases our system failed in building alignment hyper-graphs which is due to the need of improving the coverage of the Jeux DeMots graph (new relation types). The semantic relations we extracted and the entities we built according to the rules show the following results:

Table 4. The MAKI system: relation extraction and semantic modeling first results

Relation type	**Reference**	**Extraction**	**Recall**	**Precision**	**F-score**	**Modeling**	
patient	1,500	1,700	1	88%	94%	action	966
characteristic	505	570	94%	86%	90%	state	70
manner	378	288	50%	98%	66%	event	15
successor_time	420	67	78%	1	87%	set_of_ingredients	210
has-part	168	145	73%	1	84%	mixture ::	63
quantifier	84	77	83%	1	91%		
place	462	444	69%	88%	77%		
place of action	336	329	61%	82%	70%		
instrument	25	21	68%	1	81%		
agent (not expected)	-	205	-	-	-		
agent-1 (not expected)	-	200	-	-	-		
Totals	3,878	4,152	-	-	-	Total	1,324
average	-	-	**75%**	**93%**	**82%**	**Modeling accuracy**	**77%**

Conclusion

While the machine learning techniques are dominant in the research field of computational linguistics within the text analysis, the graph based approach with crafted knowledge remains a very promising area for fine semantic analysis of raw text as well as for the terminological and relation retrieval. Indeed, a common knowledge graph gives access to the extra-linguistic information which is not part of the text as a percept and which rarely appears in texts. Thus, this type of resource is relevant for the domain specific texts' analysis.

References

1. *Abacha, A. B., & Zweigenbaum, P.* (2011). Automatic extraction of semantic relations between medical entities: a rule based approach. J. Biomedical Semantics, 2(S-5), S4.
2. *Auger, A., & Barrière, C.* (2008) Pattern-based approaches to semantic relation extraction: A state-of-the-art. Terminology, 14(1), pp. 1–19.
3. *Badra, F., Bendaud, R., Bentebibel, R., Champin, P. Cohan, J., Cordier, A.* (2008) TAAABLE : Text Mining, ontology Engineering, and Hierarchical Claasification for Textual Case-Based Cooking. Actes de 9th European Conference on Case-Based Reasoning, 5239, 219–228.
4. *Banko, M., Michael J., Cafarella, M. J., Stephen Soderland, S., Broadhead, M., and Etzioni, O.* (2007) Open Information Extraction from the Web, International Joint Conference on Artificial Intelligence (IJCAI).
5. *Bodenreider, O.* (2004). The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research, 32(suppl 1), D267–D270.
6. *Chambers, N., & Jurafsky, D.* (2008). Unsupervised learning of narrative event chains. Proceedings of the Association of Computational Linguistics, 31(14), 789–797.
7. *Chklovski, T. and Pantel, P.* (2004). Verbocean : Mining the web for fine-grained semantic verb relations. In Dekang Lin and Dekai Wu, editors, Proceedings of EMNLP 2004, 33–40, Barcelona, Spain, July. Association for Computational Linguistics.
8. *Chowdhury, M. F. M., & Lavelli, A.* (2012, April). Combining tree structures, flat features and patterns for biomedical relation extraction. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 420–429). Association for Computational Linguistics.
9. *Cimino, J. J., & Barnett, G. O.* (1993). Automatic knowledge acquisition from MEDLINE. Methods of information in medicine, 32, 120–120.
10. *Das, D., Chen, D., Martins, A., Schneider, N., and Smith, N.* (2014). Frame-semantic parsing. Computational Linguistics.
11. *Denis, P. and Sagot, B.* (2012). Coupling an annotated corpus and a lexicon for state-of-the-art pos tagging. Language Resources and Evaluation, (46): 721–736.
12. *Dong, Z., Dong Q.* (2006). HowNet and the Computation of Meaning. World Scientific Publishing.

13. *Dufour-Lussier, V., Le Ber, F., Lieber, J., & Nauer, E.* (2014). Automatic case acquisition from texts for process-oriented case-based reasoning, Information systems, 40, 153–167.
14. *Embarek, M., & Ferret, O.* (2008). Learning Patterns for Building Resources about Semantic Relations in the Medical Domain. In LREC, may 2008.
15. *Esuli, A., Marcheggiani, D., & Sebastiani, F.* (2013). An enhanced CRFs-based system for information extraction from radiology reports. Journal of biomedical informatics, 46(3), pp. 425–435.
16. *Fabre, C., Bourigault, D.* (2006). Extraction de relations sémantiques entre noms et verbes au-delà des liens morphologiques. TALN 2006, Leuven, 10–13 avril 2006, pp. 121–129.
17. *Fellbaum, C.* (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.
18. *Lafourcade, M.* (2007, December). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In SNLP'07: 7th international symposium on natural language processing (p. 7).
19. *Lafourcade M.*(2011). Lexique et analyse sémantique de textes—structures, acquisitions, calculs, et jeux de mots, HDR, Université Montpellier II.
20. *Liu, H., Singh, P.* (2004). ConceptNet—a practical commonsense reasoning toolkit. BT technology journal, 22(4):211–226.
21. *Malmaud, J., Wagner, E. J. Chang, N., Murphy, K.* (2014). Cooking with Semantics, Actes de ACL 2014, 33–38.
22. *Morris, J., & Hirst, G.* (2004). Non-classical lexical semantic relations. Actes de Workshop on Computational Lexical Semantics, pp.46–51.
23. *Navigli, R., Ponzetto, S. P.* (2012). BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. Artificial Intelligence, 193, Elsevier, pp. 217–250.
24. *Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, K., Marinov, S., and Marsi, E.* (2007). Maltparser : A language-independent system for data-driven dependency parsing. Natural Language Engineering 13(2) :95–135.
25. *Pedersen, T.*(2008). Computational Approaches to Measuring the Similarity of Short Contexts : A Review of Applications and Methods. CoRR abs/0806.3787
26. *Poria, S., Agarwal, B., Gelbukh, A., Hussain, A., Howard, N.* (2014). Dependency-Based Semantic Parsing for Concept-Level Text Analysis, in Computational Linguistics and Intelligent Text Processing: 15th International Conference, CICLing 2014, Nepal, April 6–12, Springer Berlin Heidelberg :113–127.
27. *Pustejovsky, J.*(1995) The Generative Lexicon, MIT Press, Cambridge.
28. *Ramadier, L., & Zarrouk, M.* (2014). Annotations et inférences de relations dans un réseau lexico-sémantique :application à la radiologie, Actes de TALN, 103–112.
29. *Rindflesch, T. C., Bean, C. A., & Sneiderman, C. A.* (2000). Argument identification for arterial branching predications asserted in cardiac catheterization reports. In Proceedings of the AMIA Symposium (p. 704). American Medical Informatics Association.
30. *Rink, B., Harabagiu, S., & Roberts, K.* (2011). Automatic extraction of relations between medical concepts in clinical texts. Journal of the American Medical Informatics Association, 18(5), 594–600.

31. *Song, M., Kim, W. C., Lee, D., Heo, G. E., & Kang, K. Y.* (2015). PKDE4J: Entity and relation extraction for public knowledge discovery. Journal of biomedical informatics, 57, 320–332.
32. *Song, M., Yu, H., & Han, W. S.* (2011). Combining active learning and semi-supervised learning techniques to extract protein interaction sentences. BMC bioinformatics, 12(Suppl 12), S4.
33. *Snow, R., Jurafsky, D., & Ng, A. Y.* (2006, July). Semantic taxonomy induction from heterogenous evidence. Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (pp. 801–808). Association for Computational Linguistics.
34. *Suchanek, F. M., Ifrim, G., & Weikum, G.* (2006, August). Combining linguistic and statistical analysis to extract relations from web documents. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 712–717). ACM.
35. *Tasse, D. and Smith, N.*(2008). SOUR CREAM: Toward Semantic Processing of Recipes. TechnicalReport CMU-LTI-08–005, Carnegie Mellon University, Pittsburgh, PA.
36. *Zarrouk, M.* (2015) Consolidation endogène de réseaux lexico-sémantiques : inférence et annotation des relations, règles d'inférence et langage dédié. Thèse de doctorat, Université Montpellier II.