

Relation Inference in Lexical Networks ... with Refinements

Manel Zarrouk, Mathieu Lafourcade

UM2-LIRMM,
161 rue Ada
34095 Montpellier, FRANCE
manel.zarrouk@lirmm.fr, mathieu.lafourcade@lirmm.fr

Abstract

Improving lexical network's quality is an important issue in the creation process of these language resources. This can be done by automatically inferring new relations from already existing ones with the purpose of (1) densifying the relations to cover the eventual lack of information and (2) detecting errors. In this paper, we devise such an approach applied to the JeuxDeMots lexical network, which is a freely available lexical and semantic resource for French. We first present the principles behind the lexical network construction with crowdsourcing and *games with a purpose* and illustrated them with JeuxDeMots (JDM). Then, we present the outline of an *elicitation engine* based on an *inference engine* using schemes like deduction, induction and abduction which will be referenced and briefly presented and we will especially highlight the new scheme (Relation Inference Scheme with Refinements) added to our system. An experiment showing the relevance of this scheme is then presented.

Keywords: relation inference, lexical network improvement, reconciliation

1. Introduction

In Computational Linguistics (CL), constructing lexical resources on which NLP applications can rely is one of the crucial matters on the field. Most of these existing resources have been constructed by hand such as for the popular WordNet, a lexical network based on synsets which can be roughly considered as concepts (Fellbaum, 1988). Others like EuroWordnet (Vossen., 1998) a multilingual version of WordNet and WOLF (Sagot., 2008) a French version of WordNet, were built by automated crossing of WordNet and other lexical resources along with some manual checking. Navigli. (2012) constructed automatically BabelNet a large multilingual lexical network from term co-occurrences in the Wikipedia encyclopedia. Crowdsourcing approaches are flowering in CL especially with the advent of Amazon Mechanical Turk (M.Winter, 2010) or in a broader scope Wikipedia and Wiktionary, to cite the most well known examples.

In this context, a lexical resource built collaboratively, should be considered as being constantly evolving and as such the state of an entry can not be definitively verified. The construction of this kind of resources, especially lexical networks, can be realised according to two strategies. Firstly, it can be designed as a contributive system (like Wikipedia or Wiktionary) where people willingly add and complete entries. Secondly, contributions can be made indirectly thanks to games (better known as GWAP as proposed by (vonAhn, 2008)) and in this case, players do not need to be aware that while playing they are helping building a lexical resource. In both cases, the built lexical network is not free of errors which are corrected upon their discovery.

Thus, a large number of obvious relations are missing in the lexical network but are indeed necessary for a high-quality resources usable in various NLP applications and notably semantic analysis. For example, contributors seldom indicate that a particular bird type can fly, as

it is considered as an obvious generality. Only notable facts which are not easily deductible are naturally contributed. Well known exceptions are also generally contributed and take the form of a negative weight (for example, $fly \xrightarrow{agent:-100} ostrich$).

For the lexical network consolidation, we adopt a strategy based on a simple inference system to propose new relations from those already existing. The approach is strictly endogenous (i.e. self-contained) as it does not rely on any other external resources.

Inferred relations are submitted either to contributors for voting or to experts for direct validation/invalidation. A large percentage of the inferred relations has been found to be correct however, a non negligible part of them are found to be wrong and understanding why is both interesting and useful. The explanation process can be viewed as a *reconciliation* between the inference engine and contributors who are guided through a dialog to explain why they found the considered relation incorrect. The causes for a wrong inferred relation may come from three possible origins: false premises that were used by the inference engine, exception or confusion due to some polysemy.

In (Sajous et al., 2013) an endogenous enrichment of Wiktionary is done thanks to a crowdsourcing tool. A quite similar approach of using crowdsourcing has been considered by (Zeichner, 2012) for evaluating inference rules that are discovered from texts. In (Krachina, 2006), some specific inference methods are conducted on text with the help of an ontology. Similarly, (Besnard, 2008) capture explanation with ontology-based inference. OntoLearn (Velardi, 2006) is a system that automatically builds ontologies of specific domains from texts and also makes use of inferences. There have been also researchs on taxonomy induction based on WordNet (see (Snow, 2006)). Although extensive work on inference from texts or handcrafted resources has been done, almost none endogenously on lexical network built by the crowds.

Most probably the main reason of that situation is the lack of such specific resources.

In this paper, we first present the principles behind the lexical network construction with crowdsourcing and *games with a purpose* (also know as human-based computation games) and illustrated them with the JeuxDeMots (JDM) a french project launched in 2007. Then, we present the outline of an *elicitation engine* based on an *inference engine* using schemes like deduction, induction and abduction which will be briefly presented and we will especially highlight a new scheme added to our system. An experiment showing the relevance of this scheme is then presented.

2. Crowdsourcing lexical networks

When building a lexical network, some crucial factors as the quality of data, cost and time are taken into consideration. Beside manual or automated strategies, contributive approaches are flowering and becoming more and more popular as they are both cheap to set up and efficient in quality. There is an increasing trend of using online GWAPs (game with a purpose (Thaler et al., 2011)) methods for feeding such resources.

The JDM lexical network is constructed through a set of associative games. The game's principle is to make players contribute on lexical and semantic relations between terms (nouns, verbs, expressions, named entities) which are contained in the network. The information in the JDM network are gathered by an unnegotiated crowd agreement (classical contributive systems rely on a negotiated crowd agreement (the contributor must settle on the information so that it's integrated on the resource)). JeuxDeMots is a projet aiming at building collaboratively a lexical network (JDM Network) through two approaches:

JeuxDeMots¹ is a two player GWAP, launched in September 2007 (Lafourcade, 2007), whose goal is to build a large lexical-semantic network ((Lafourcade., 2008) and (Lafourcade, 2012)). This resource has been made freely available by its authors with a monthly update (as a dump) and as such, it is invaluable to make various NLP experiments (for example: program codes analysis with *Orange* (P. Warintarawej, 2014), report analysis in radiology with *Imaios* (L. Ramadier, 2014), public debate management with *SucceedTogether* (C. da Costa Pereira, 2014)) . The network is composed of terms (as vertices) and typed relations (as links between vertices) with weight. There are more than 50 types for relations, that range from ontological (hypernym, hyponym), to lexico-semantic (synonym, antonym) and to semantic roles (agent, patient, instrument). The weight of a relation is interpreted as a strength, but not directly as a probability of being valid. The JDM network is therefore constructed by connecting terms with typed/weighted relations, through iterated validation by pairs of players. The weight of a given relation between two terms is related to the number of times a pair of player proposed this relation.

Diko² is a web based tool for displaying the information contained in the JDM lexical network which can also be used as a *negotiated* contributive tool. People contributing can discuss and negotiate each proposal, hence Diko being to this respect complementary to JDM.

3. CIR approach - Consolidation by Inference and Reconciliation

Increasing the number of relations in the JDM lexical network relies on two cooperating components: (a) *an inference engine* and (b) *a reconciliator*.

The inference engine proposes relations, as if being a normal contributor, to be validated by other human contributors or validators. In case of invalidation of an inferred relation, the reconciliator is invoked to try to ascertain the cause of the problem. Consolidation here should be seen as the process to transform some implicit knowledge of the user into explicit relations in the lexical network.

3.1. Inference Engine

The inference engine is founded on *relations inference schemes* (RIS) which serve as patterns for the engine behavior.

3.1.1. Deduction, Induction and Abduction

Deduction is a top-down scheme based on the transitivity of the ontological relation *is-a* (hypernym). The scheme can be formally written as follows:

$$(\exists A \xrightarrow{is-a} B \wedge \exists B \xrightarrow{R} C \Rightarrow A \xrightarrow{R} C)$$

Since the scheme itself is simplistic, we devised a logical and statistical blocking strategy to increase the pertinence of inferred relations (Zarrouk, 2013).

Induction is the inverse scheme of deduction as being a bottom-up scheme exploiting as well the transitivity of *is-a*. More formally, we can write:

$$(\exists A \xrightarrow{is-a} B \wedge \exists A \xrightarrow{R} C \Rightarrow B \xrightarrow{R} C)$$

The principle is similar to the one applied to the deduction scheme and similarly some logical and statistical filtering may be undertaken (Zarrouk, 2013).

The *abduction* scheme can be viewed as an example based strategy. Hence abduction relies on similarity between terms, which may be formalized in our context as sharing some outgoing relations between terms (M. Zarrouk, 2014). The abductive inferring layout supposes that relations held by a term can be proposed to similar terms. Here, abduction first selects a set of similar terms to the target term which are considered as proper examples. The outgoing relations from the examples which are not common with those of the target term are proposed as potential relations for it and then presented for validation/invalidation to users. Unlike induction and deduction, abduction can be applied on terms with missing or irrelevant ontological relations, and can generate ontological relations to be used afterward by the inference

¹<http://jeuxdemots.org>

²<http://www.jeuxdemots.org/diko.php>

loop. A filtering strategy should be applied too to avoid the proposition of dubious relations.

3.1.2. Relation Inference Scheme with Refinements

A given polysemous word, as identified by locutors, has several usages that might differ substantially from word senses as classically defined. A given usage can also in turn have several deeper *refinements* and the whole set of usages can take the form of a decision tree. For example, *frigate* can be a bird or a ship. A *frigate*>*boat* can be distinguished as a modern ship with missiles and radar *frigate*>*boat*>*modern* or an ancient vessel with sails *frigate*>*boat*>*ancient*.

This scheme as its name indicates requires the term A to have at least a refinement A' . The Relation Inference Scheme with Refinement (RIS_R) scheme, for each synonym, hypernym or hyponym (ontological relations only) B of the start term A , tries to share the outgoing relations between A' and B . The relations exchanged are the inferred relations to be validated or rejected later.

To increase the relevance of the proposed relations, we make sure that some relation exists between the refinement term A' and the term B . For example, suppose we have A : *rose* which has two refinements at least A' : *rose*>*flower* and *rose*>*color* and a hypernym B : *plant*. In this example, the terms A' : *rose*>*flower* and B : *plant* are related (some relation exists between them) unlike the terms A' : *rose*>*color* and B : *plant*. This strategy avoids proposing for example *rose*>*color* $\xrightarrow{\text{has-part}}$ *leaf* (an outgoing relation coming from B).

Another strategy is not to propose outgoing relations from an hypernym to its hyponyms (as shown in algorithm1). The direction of the transfer is always from the hyponym to the hypernym and thus because in general cases, outgoing relations of an hypernym are not valid for its hyponyms, like for example, for the term A : animal having a refinement A' : animal>animalia which can have as parts: fin, scale, fang... Those relations $x \xrightarrow{\text{has-part}}$ *fin*, *scale*, *fang* are not valid for the hyponym *cow* for example.

This scheme has a behavior subtly different according to the nature of the term B (synonym, hypernym or hyponym) and this is clearly illustrated by the pseudocode below (algorithm 1). In the algorithm 1 we use the following notations:

- $\text{out}(X)$ refers to the outgoing relations of the term X ;
- $\text{out}(X) \overset{\mathbb{P}}{\leftrightarrow} \text{out}(Y)$: propose all the outgoing relations of X as outgoing relations for the term Y (other notation as \mathbb{C} to copy relations and \mathbb{D} to displace them are available but not used here);
- $A \overset{X}{\rightleftarrows} B$: a relation between A and B in any direction.

3.2. Performing reconciliation

Inferred relations are presented to the validator to decide of their status. In case of invalidation, a reconciliation procedure is committed in the purpose to try to identify

the reasons through a dialog initiated with the user. Then, the reconciliation engine generates then corrected relations (wrong existing premises are marked, exceptional relation are annotated, refined terms are added...) to be integrated in the network and used later by the inference engine (Zarrouk, 2013).

4. Experimentation

We applied the scheme with three different seed relations:

- RIS_R (synonym): the scheme applied with R=syn (specifications: in case of existence of B' the terms A' and B' exchange relations between them)
- RIS_R (hyponym): the scheme applied with R=hypo (relations are transferred from B or B' to A')
- RIS_R (hypernym): the scheme applied with R=hyper (relations are transferred from A' to B or B' (see algorithm (1))

4.1. Productivity

Since the schema has a condition to be applied as explained above, the propositions (inferred relations) are made for only **6 349** terms fulfilling the constraints. The whole process produced **308 532** inferences presenting totally new relations not existing before in the network which make about **49** new relations per entry. The RIS_R(syn) produced **2.7** times the existing relations which make it the most productive version, followed by the RIS_R(hypo) producing **2.6** times and the RIS_R(hyper) with a productivity of **0.73** (table 1).

	# existed	# proposed	productivity
RIS_R (syn)	38 792	105 288	271.41%
RIS_R (hyper)	139 490	101 908	73.05%
RIS_R (hypo)	38 756	101 336	261.47%

Table 1: The number of relations existing before application of the scheme and those proposed by the scheme. The statistics were made on the terms proposed by the inference scheme.

The inferred relations are detailed by relation type in the table 2. The different relation types are variously productive, and this is mainly due to the number of existing relations and the distribution of their type. The *associated idea* type is the most proposed from both three schemes and this is explained by the large semantic spectre of this relation type since it refers to every term associated to the target term.

The figures are inverted for some other relations that are not so well populated in the lexical network but still are potentially valid.

4.2. Accuracy

The validation processus was applied manually on a sample of around 1 000 propositions randomly chosen for each scheme. The synonym version has the highest accuracy with **90.76 %** valid relations, hypernym version with **72.69 %** and **66.24 %** for the hyponym version (table 3).

foreach (A *such as* $\exists A \xrightarrow{raff} A' \wedge \exists A \xrightarrow{R \in \{syn/hyper/hypo\}} B \wedge \exists B \xrightarrow{X} A'$) **do**

if ($\exists B \xrightarrow{raff} B' \wedge \exists A' \xrightarrow{X} B'$) **then**

if $A \xrightarrow{R=syn} B$ **then**
 $out(B) \stackrel{P}{\leftrightarrow} out(A)$;
 $out(A) \stackrel{P}{\leftrightarrow} out(B)$;
end

if $A \xrightarrow{R=hyper} B$ **then**
 $out(A) \stackrel{P}{\leftrightarrow} out(B)$;
end

if $A \xrightarrow{R=hypo} B$ **then**
 $out(B) \stackrel{P}{\leftrightarrow} out(A)$;
end

end

if ($\nexists B \xrightarrow{raff} B' \vee \nexists A' \xrightarrow{X} B'$) **then**

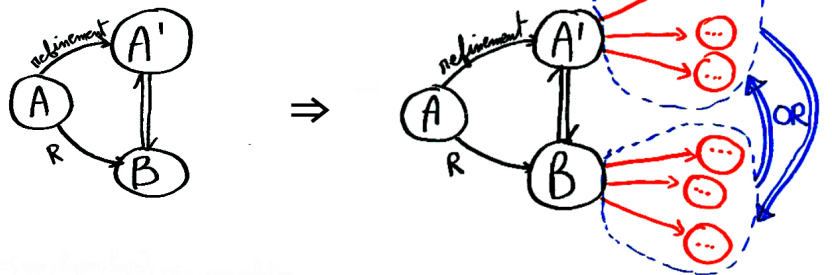
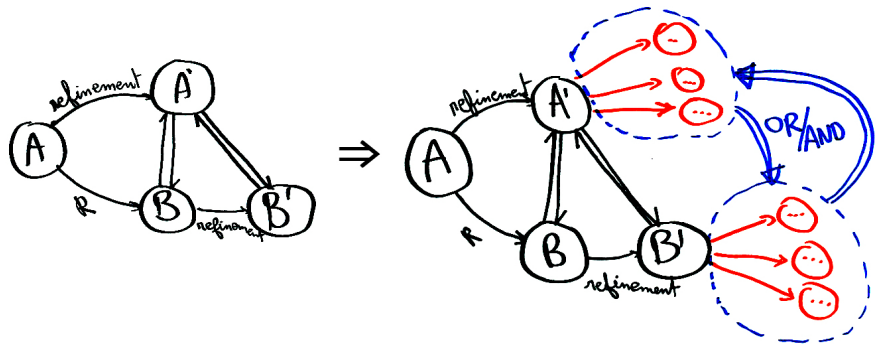
if $A \xrightarrow{R=syn} B$ **then**
 $out(A) \stackrel{P}{\leftrightarrow} out(B)$;
end

if $A \xrightarrow{R=hyper} B$ **then**
 $out(A) \stackrel{P}{\leftrightarrow} out(B)$;
end

if $A \xrightarrow{R=hypo} B$ **then**
 $out(B) \stackrel{P}{\leftrightarrow} out(A)$;
end

end

end



$R \in \{syn, hyper, hypo\}$;
 \Rightarrow : any relation;
raff: refinement.

Algorithm 1: The progress of the RIS_R scheme according to the existence of a term B' and the type of the R relation.

The synonym version of the scheme has systematically the best accuracy for all the relation types. Some accuracy percentages are lower than others for these reasons. In certain cases, some outgoing relations of an hyponym do not suit for the hypernym. For example:

• A : animal • A' : animal>animalia • $B(hypo)$: cat
 \Rightarrow The inference scheme will propose the outgoing relation of *cat* ($cat \xrightarrow{is-a} pet$) to *animal>animalia* ($animal>animalia \xrightarrow{is-a} pet$) which is wrong and this explain the weak percentage of accuracy for example of the relation *is - a* (56.4% by the RIS_R(hypo) and 46% by the RIS_R(hyper)) and *has - part* (46.9% by the RIS_R(hypo)).

Another reason is that in the network, some terms are not refined (or not completely refined) which can lead to some wrong relations, as for example:

• A : milk • A' : milk>dairy product • $B(hypo)$: cow³

\Rightarrow The inference scheme will propose the relation ($milk > dairy\ product \xrightarrow{has-part} teats$) which is wrong and thus because the term *cow* is not yet refined into *cow > dairy product* and *cow > animal*.

From the figures, we can make the following observations. First, global results show that produced inferences are strongly valid with synonyms. The resultat are much poorer with hypernyms and hyponyms (table 3) which is obvious regarding that with synonym, the terms exchanging relations are roughly at the same level of the taxonomy hierarchy which is not the case when they are related with an hyponym or hypernym relation.

³In french, some dairy products are called sometimes by the name of the producer animal, like *chèvre*(goat) for the cheese made from the goat's milk

Relation type	RIS_R(syn)	RIS_R(hypo)	RIS_R(hyper)
associated	50 946	39 325	51 960
has-part	13 362	13 120	8 049
is-a	3 711	5 114	5 707
hyponym	6 463	10 186	6 326
holonym	1 927	1 407	3 757
charac	10 378	10 063	7 614
location	5 921	9 251	5 529
agent-1	6 887	9 366	3024
other	5 693	4 076	9 370

Table 2: Relations proposed by type of scheme and relation type. The *associated idea* type is the most proposed from both three schemes and the figures are inverted for some other relations that are not so well populated in the lexical network.

Relation type	RIS_R(syn)	RIS_R(hypo)	RIS_R(hyper)
associated	92.4%	65%	60.8%
has-part	93.2%	46.9%	80.8%
is-a	86.2%	56.4%	46%
hyponym	69.7%	60%	65%
holonym	74.4%	60.7%	64.2%
charac	91.5%	73.9%	90.5%
location	91.1%	81%	79.5%
agent-1	92.1%	78.9%	90.9%

Table 3: Percentage of valid relations by type of scheme and relation type. RIS_R(syn) version of the scheme has systematically the best accuracy for all the relation types.

5. Conclusion

We presented some issues in building a lexical semantic network with games and user contributions and about inferring new relations from existing ones. To be able to enhance the network quality and coverage, we proposed a consolidation approach based on an inference engine and a reconciliator. If an inferred relation is proven wrong, a reconciliation process is conducted in order to identify the underlying cause in order to solve the problem.

This inference engine is considered as a contributor like a human one. It infers new relations from those already existing ones in the network by using some behavior patterns: deduction, induction, abduction and a refinement based scheme which has been deeply detailed in this paper.

The inference relation scheme with refinements relies on a pre condition to be applied which limitates the number of terms on which it can be applied (6 349) and so constrains the quantity of relations inferred (308 532). However, it increases the accuracy of these ones (which varies from 66.24 % to 90.76 % depending on the ontological relation used in the scheme).

Our scheme, by using only the synonym relations, infers 2.7 times the number of the relations existing on the choosen terms with an accuracy of 90.76% which is quite

interesting and promising. This scheme, added to our elicitation engine as an inference scheme, is as proved a very efficient way to enrich endogenously the network leading to an increase in quality and in lexical coverage.

6. References

- von Ahn, L. and Dabbish, L. 2008. *Designing games with a purpose*. in bouahhahahahaha Communications of the ACM, number 8, volume 51.p58-67.
- Besnard, P. Cordier, M.O., and Moinard, Y. 2008. *Ontology-based inference for causal explanation*. Integrated Computer-Aided Engineering , IOS Press, Amsterdam , Vol. 15 , No. 4 , 351-367 , 2008.
- da Costa Pereira, C., Lafourcade, M., Lloret, P., Lopez, C. and Roche, M. 2014. *Vectorisation paramétrée des données textuelles*. Demonstration in the 14ème conférence Extraction et Gestion des Connaissances (EGC 2014), Rennes, France, 28 au 31 Janvier 2014, 2014, 4 p.
- Fellbaum, C. and Miller, G. 1988. (*eds*) *WordNet*. The MIT Press.
- Krachina, O. and Raskin, V. 2006. *Ontology-Based Inference Methods* CERIAS TR 2006-76, 6p.
- Lieberman, H, Smith, D. A and Teeters, A. 2007. *Common consensus: a web-based game for collecting common-sense goals*. In Proc. of IUI, Hawaii,2007.12p .
- Lafourcade, M. 2007. *Making people play for Lexical Acquisition*. In Proc. SNLP 2007, 7th Symposium on Natural Language Processing. Pattaya, Thaïlande, 13-15 December. 8 p.
- Lafourcade, M. and Joubert, A. 2008. *JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes*. In proc of JADT'2008, Ecole normale supérieure Lettres et sciences humaines , Lyon, France, 12-14 mars 2008 .
- Lafourcade,M. and Joubert, A. 2012. *Long Tail in Weighted Lexical Networks*. In proc of Cognitive Aspects of the Lexicon (CogAlex-III), COLING, Mumbai, India, December 2012.
- Lafourcade, M. and Joubert, A. 2012. *Long Tail in Weighted Lexical Networks*. In proc of Cognitive Aspects of the Lexicon (CogAlex-III), COLING, Mumbai, India, December 2012.
- Marchetti, A., Tesconi, M., Ronzano, F., Mosella, M. and Minutoli, S. 2007. *SemKey: A Semantic Collaborative Tagging System*. in Procs of WWW2007, Banff, Canada. 9 p.
- Mihalcea, R and Chklovski, T. 2003. *Open MindWord Expert: Creating large annotated data collections with web users help*. In Proceedings of the EACL 2003, Workshop on Linguistically Annotated Corpora (LINC). 10 p.
- Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K.J. 1990. *Introduction to WordNet: an on-line lexical database*. International Journal of Lexicography. Volume 3, p 235-244.
- Navigli, R and Ponzetto, S. 2012. *BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network*. Artificial Intelligence. 193: pp. 217-150.
- Ramadier, L., Zarrouk M., Lafourcade, M. and Micheau, A. 2014. *Spreading Relation Annotations in a Lexical*

- Semantic Network Applied to Radiology* 15th International Conference on Intelligent Text Processing and Computational Linguistics. April 6–12, 2014 - Kathmandu, Nepal.
- Sagot, B. and Fier, D. 2010. *Construction d'un wordnet libre du français à partir de ressources multilingues*. in Proceedings of TALN 2008, Avignon, France, 2008.12 p.
- Sajous, F., Navarro, E., Gaume, B., Prévot, L. and Chudy, Y. 2013. *Semi-Automatic Enrichment of Crowdsourced Synonymy Networks: The WISIGOTH system applied to Wiktionary*. Language Resources & Evaluation, 47(1), pp. 63-96.
- Siorpaes, K. and Hepp, M. 2008. *Games with a Purpose for the Semantic Web*. in IEEE Intelligent Systems, number 3, volume 23, p 50-60.
- Snow, R., Jurafsky, D. and Ng, A.Y. 2006. *Semantic taxonomy induction from heterogenous evidence*. in Proceedings of COLING/ACL 2006, 8 p.
- Thaler, S., Siorpaes, K., Simperl, E. and Hofer, C. 2011. *A Survey on Games for Knowledge Acquisition*. STI Technical Report, May 2011.19 p.
- Velardi, P., Navigli, R., Cucchiarelli, A. and Neri, F. 2006. *Evaluation of OntoLearn, a methodology for Automatic Learning of Ontologies*. in Ontology Learning and Population, Paul Buitelaar Philipp Cimmianno and Bernardo Magnini Editors, IOS press 2006).
- Vossen, P. 2011. *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic Publishers. Norwell, MA, USA. 200 p.
- Warintarawej, P., Laurent, A., Huchard, M., Lafourcade, M. and Pompidor, P. 2014. *software understanding: Automatic classification of software identifiers*. Intelligent Data Analysis (IDA Journal) 18(6) (2014) in press.
- Winter, M. and Siddharth, S. 2010. *Conducting Behavioral Research on Amazon's Mechanical Turk*. Behavior Research Methods, Volume 44, Issue 1, pp 1-23.
- Zarrouk, M., Lafourcade, M. and Joubert, A. 2013. *Inference and reconciliation in a lexical-semantic network*. 14th International Conference on Intelligent Text Processing and Computational Linguistic (CICLING-2013), 13 p.
- Zarrouk, M., Lafourcade, M. and Joubert, A. 2013. *Inférences déductives et réconciliation dans un réseau lexico-sémantique*. 20ème conférence du Traitement Automatique du Langage Naturel 2013 (TALN 2013), 14 p.
- Zarrouk, M., Lafourcade, M. and Joubert, A. 2013. *Inductive and deductive inferences in a Crowdsourced Lexical-Semantic Network*. 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013), 6 p.
- Zarrouk, M., Lafourcade, M. and Joubert, A. 2014. *About Inferences in a Crowdsourced Lexical-Semantic Network*. In proc of 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014), 8 p.
- Zeichner, N., Berant J., and Dagan I. 2012. *Crowdsourcing Inference-Rule Evaluation*. in proc of ACL 2012 (short papers).