Inferring Knowledge with Word Refinements in a Crowdsourced Lexical-Semantic Network

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

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

Abstract

Automatically inferring new relations from already existing ones is a way to improve the quality and coverage of a lexical network and to perform error detection. In this paper, we devise such an approach for the crowdsourced JeuxDeMots lexical network and we focus especially on word refinements. We first present deduction (generic to specific) and induction (specific to generic) which are two inference schemes ontologically founded and then propose a transfer schema devoted to infer relations with and for word refinements.

1 Introduction

Efficiently building useful resources for Computational Linguistics (CL) is of a crucial interest. Most of existing lexical-semantic networks have been built by hand (like for instance WordNet (Miller et al., 1990)) and, despite that assisting tools are generally designed for consistency checking, the task remains time consuming and costly. Fully automated approaches are generally limited to term co-occurrences as extracting precise semantic relations between terms from corpora remains at best difficult. Crowdsourcing approaches are flowering in CL especially with the advent of Amazon Mechanical Turk or in a broader scope Wikipedia, to cite the most well known examples. WordNet is such a lexical network, constructed at great cost, based on synsets which can be roughly considered as concepts (Fellbaum, 1988). EuroWordnet (Vossen., 1998) a multilingual version of WordNet and WOLF (Sagot., 2008) a French version of WordNet, were built by automated crossing of the original Princeton WordNet and other lexical resources along with some more or less manual checking. Navigli (2010) constructed automatically BabelNet a large multilingual lexical network from term co-occurrences in Wikipedia. Although being very large and multilingually connected (which is tremendously usefull for machine translation, for instance) it contains few various lexical-semantic relations.

An *ideal* lexical-semantic network contains interconnected lemmas, word forms and multi-word expressions as entry points (nodes) along with word meanings and concepts. The idea itself of *word senses* as forwarded in the lexicographic tradition may be debatable in the context of resources for semantic analysis, and we generally prefer to consider the psycholinguistic idea of *word usages*. 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 a very classical example, *bank* can be related to money or river : *bank* ⋗ *'bank>money'* and *bank* ⋗ *'bank>river'*. A *'bank>money'* can be distinguished as the financial institution or the actual building.

In the context of a collaborative construction, such a lexical resource should be considered as being constantly evolving and a general pragmatic rule of thumb is to have no definite certitude about the state of an entry. For a polysemous term, some refinements might be just missing at a given time

Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 365–375, Dublin, Ireland, August 23-29 2014.

notwithstanding the evolution of language which might be very fast, especially in technical domains. There is no way (unless by inspection) to know if a given entry refinements are fully completed, and even if this question is really relevant.

Creating collaboratively a lexical-semantic network (or, in all generality, any similar resource) can be devised according to two broad strategies. Firstly, it can be designed as a contributive system like Wikipedia where people willingly add and complete entries (like for Wiktionary). Secondly, contribution can be undertaken indirectly thanks to games (also known as GWAP (vonAhn, 2008)). In this case, players do not need to be aware that while playing they are helping building a lexical and semantic resource. In any case, the built network is not free of errors which are (or should be) corrected along their discovery. Thus, a large number of obvious relations may be missing in the lexical network but are indeed necessary for a high quality resources usable in various NLP applications, or even crucial notably for textual semantic analysis.

For example, contributors seldomly 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. Conversely, well known exceptions are also generally contributed and take the form of a negative weight and anotated as such (for example, *fly* $\xrightarrow{agent:-100}$ *ostrich* [exception: bird]). In order to consolidate the lexical network, we adopt a strategy based on a simple inference mechanism to propose new relations from those already existing. The approach is strictly endogenous (i.e. self-contained) as it doesn't 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 possible 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 build 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. In this article, we first present the principles behind the lexical network construction with crowdsourcing and *games with a purpose* (also known as human-based computation games) and illustrated them with the JeuxDeMots (JDM) project. Then, we present the outline of an *elicitation engine* based on an *inference engine* using deduction, induction and especially relation transfer schemes. The reconciliation engine which presents the second part of the elicitation engine is detailed on previous papers (Zarrouk, LREC2014) (Zarrouk, TALN2013). An experimentation with a discussion is then detailed.

2 Crowdsourced lexical networks

For validating our approach, we used the JDM lexical network, which has been made freely available by its authors, and constructed thanks to a set of associatory games (Lafourcade, 2007). There is an increasing trend of using online GWAPs (game with a purpose (Thaler et al., 2011)) method for feeding such resources. 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.

The network is composed of terms (as vertices) and typed relations (as links between vertices) with weights. It contains terms and possible refinements. There are more than 50 types for relations, that range from ontological (hypernym, hyponym), to lexical-semantic (synonym, antonym) and to semantic role (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 not an ontology with some pristine, factorized and well-thought hierarchy of concepts or terms. A given term can have a substantial set of hypernyms that covers a large part of the ontological chain to upper concepts. For example, hypernym(cat) = {feline, mammal, living being, pet, vertebrate, ...}. Heavier weights associated to terms are those felt by users as being the most relevant. On the 1st of January 2014, there are more than 6 800 000 relations and roughly 310 000 lexical items in the JDM lexical network (according to the figures given by the game site: http://jeuxdemots.org). To our knowledge, there is no other, in French at least, existing freely available crowdsourced lexical-network, especially with weighted relations, thus enabling strongly heuristics or psycho-linguistically motivated methods.

3 Inferring Semantic Relations...

Adding new relations to the JDM lexical network may rely on two components: (a) an inference engine and (b) a reconciliator. The inference engine proposes relations as if it was a contributor, to be validated by other human contributors or experts. In case of invalidation of an inferred relation, the reconciliator is invoked to try to assess why the inferred relation was found wrong. Elicitation here should be understood as the process to transform some implicit knowledge of the user into explicit relations in the lexical network. The core ideas about inferences in our engine are the following:

- inferring is to derive new premises (taking the form of relations between terms) from previously known premises, which are existing relations;
- candidate inferences may be logically blocked on the basis of the presence or the absence of some other relations;
- candidate inferences can be filtered out on the basis of a strength evaluation. The strong assumption here is to consider strengh as a confidence level, which is in fact only partially exact. More precisely, high strengh values clearly correlate to confidence, but we cannot say much about low strength values.

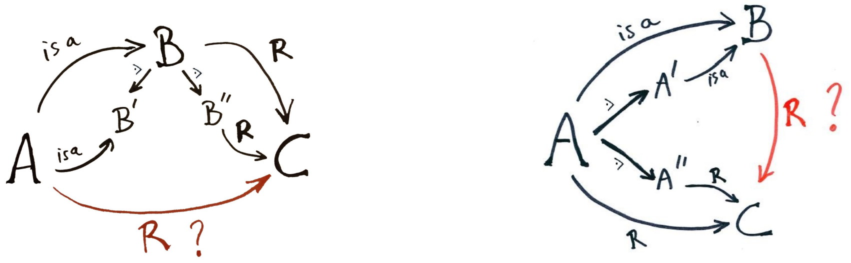

Figure 1: On the left, triangular deductive inference scheme where logical blocking based on the polysemy of the central term B which has two distinct meanings B' and B'' is applied. Arrows labelled $>$ are word meaning/refinements. The relation R? is the conclusion that may be blocked. On the right, (A *is-a* B) and (A R C) are the premises, and (B R C) is the induction proposed for validation. Term A may be polysemous with refinements holding premises, thus inducing a probably wrong relation.

3.1 ... by Deduction and by Induction...

Inferring by deduction (Zarrouk, RANLP2013) is a top-down scheme based on the transitivity of the ontological relation *is-a* (hypernym). If a term A is a kind of B and B holds some relation R with C, then we can expect that A holds the same relation type with C. The scheme can be formally written as follows:

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

For example, *shark* $\xrightarrow{is-a}$ *fish* and *fish* $\xrightarrow{has-part}$ *fin*, thus we can expect that *shark* $\xrightarrow{has-part}$ *fin*. The inference engine is applied on terms having at least one hypernym (the scheme could not be applied otherwise). Of course, this scheme is far too naive, especially considering the resource we are dealing with and may produce wrong relations. Indeed, the central term B is possibly polysemous and ways to avoid probably wrong inferences can be done through a *logical blocking*: if there are two distinct meanings for B that hold respectively the first and the second relation, then most probably the inferred relation is wrong (see figure 1) and hence should be blocked. Moreover, if one of the premises is tagged by contributors as *true but irrelevant*, then the inference is blocked.

It is possible to evaluate a confidence level (on an open scale) for each produced inference, in a way that dubious inferences can be eliminated out through *statistical filtering*. The weight w of an inferred relation is the geometric mean of the weights of the premises (relations (A *is-a* B) and (B *R* C) in figure 1). If the second premise has a negative value, the weight is not a number and the proposal is discarded. As the geometric mean is less tolerant to small values than the arithmetic mean, inferences which are not based on two rather strong relations (premises) are unlikely to pass.

$$w(A \xrightarrow{R} C) = (\ w(A \xrightarrow{is-a} B) \times w(B \xrightarrow{R} C)\ )^{1/2} \quad \Rightarrow \quad w3 = (w1 \times w2)^{1/2}$$

Although making a transitive closure over a knowledge base is not new, doing so considering word usages (refinements) over a crowdsourced lexical network is an original approach. As for the deductive inference, induction (Zarrouk, RANLP2013) exploits the transitivity of the relation *is-a*. If a term A is a kind of B and A holds a relation R with C, then we might expect that B could hold the same type of relation with C. More formally we can write: $\exists\, A \xrightarrow{is-a} B \quad \wedge \quad \exists\, A \xrightarrow{R} C \quad \Rightarrow \quad B \xrightarrow{R} C$
For example, *shark* $\xrightarrow{is-a}$ *fish* and *shark* $\xrightarrow{has-part}$ *jaw*, thus we might expect that *fish* $\xrightarrow{has-part}$ *jaw*. This scheme is a generalization inference. The principle is similar to the one applied to the deduction scheme and similarly some logical and statistical filtering may be undertaken. The central term here A, is possibly polysemous (as shown in figure 1). In that case, we have the same polysemy issues with the deduction, and the inference may be blocked. The estimated weight for the induced relation is:

$$w(B \xrightarrow{R} C) = (w(A \xrightarrow{R} C))^2 \ / \ w(A \xrightarrow{is-a} B) \quad \Rightarrow \quad w2 = (w_3)^2 / w_1$$

3.2 ... and Performing Reconciliation

Inferred relations are presented to the validator to decide of their status. In case of invalidation, a reconciliation procedure is launched in order to diagnose the reasons: error in one of the premises (previously existing relations are false), exception or confusion due to polysemy (the inference has been made on a polysemous central term). A dialog is initiated with the user. To know in which order to proceed, the reconciliator checks if the weights of the premises are rather strong or weak.

Errors in the premises. We suppose that the relation (A *is-a* B) (in figures 1) has a relatively low weight. The reconciliation process asks the validator if that relation is true. It sets a negative weight to this relation if it is false so that the inference engine blocks further inferences. Else, if the relation (A *is-a* B) is true, we ask about the second relation (B *R* C or A *R* C) and proceed as above if the answer is negative. Otherwise, we check the other cases (exception, polysemy).

Errors due to exceptions. For the *deduction*, in case we have two trusted relations, the reconciliation process asks the validators if the inferred relation is a kind of exception relatively to the term B. If it is the case, the relation is stored in the lexical network with a negative weight and annotated as *exception*. Relations that are exceptions do not participate further as premises for deducing. For the *induction*, in case we have two trusted relations, the reconciliator asks the validators if the relation ($A \xrightarrow{R} C$) (which served as premise) is an exception relatively to the term B. If it is the case, in addition to storing the false inferred relation ($B \xrightarrow{R} C$) in the lexical network with a negative weight, the relation ($A \xrightarrow{R} C$) is annotated as *exception*. In the induction case, the exception is a true premise which leads to a false induced relation. In both cases of induction and deduction, the *exception* tag concerns always the relation ($A \xrightarrow{R} C$). Once this relation is annotated as an exception, it will not participate as a premise in inferring generalized relations (bottom-up model) but can still be used in inducing specified relations (top-down model).

Errors due to Polysemy. If the central term (B for deduction and A for induction) presenting a polysemy is mentioned as polysemous in the network, the refinement terms $term_1$, $term_2$, $\ldots term_n$ are presented to the validator so he can choose the appropriate one. The validator can propose new terms as refinements if he is not satisfied with the listed ones (inducing the creation of new appropriate refinements). If there is no meta information indicating that the term is polysemous, we ask

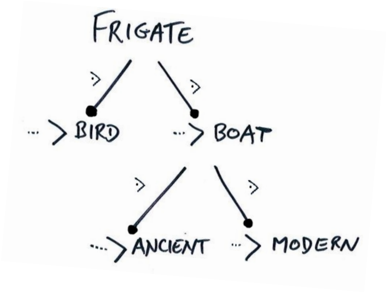

Figure 2: Refinement (noted >) tree of the term *frigate*. The first level discriminates between *frigate>bird* and *frigate>boat* which itself is refined between *(frigate>boat)>ancient* and *(frigate>boat)>modern*. This tree is a part of the lexical network which makes use of a specific refinement relation. Each refinement is connected to other terms of the network.

first the validator if it is indeed the case. After this procedure, new relations will be included in the network with positive values and the inference engine will use them later on as premises.

3.3 Transferring Relations 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*). Having proper relations between refinements and other terms or refinements is crucial for word sense disambiguation.

The purpose of this scheme is to *enrich refinements and terms that are ontologically connected*. As its name indicates, this scheme requires the term A to have at least a refinement A' and at least one support relation that is ontological. The Relation Inference Scheme with Refinements (RIS_R) scheme, for each synonym, hypernym or hyponym (the support) 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 latterly. 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 avoid proposing for example *rose>color* $\xrightarrow{has-part}$ *leaf* (an outgoing relation coming from B).

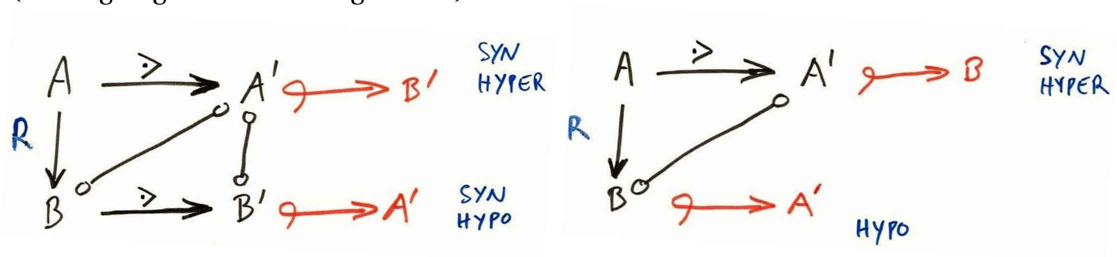

Figure 3: Relation Inference Scheme with Refinements (RIS_R). Above A (resp. B) has a refinement A' (resp. B'). Outgoing relations of A' are copied as outgoing relations of B' and vice-versa, according to the support relation (syn, hyper, hypo). On the right, we are in a *minimal* situation where B has no refinement.

Another strategy is not to propose outgoing relations from an hypernym to its hyponyms. The direction of the transfer is always from the hyponym to the hypernym because generally, outgoing relations of an hypernym are not all valid for its hyponyms. For example, for the term A: *animal* having a refinement A': *animal>zoology* which can have as parts *fin, scale, fang...* Those relations x $\xrightarrow{has-part}$ *(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) relatively to A. In figure 3, we use the following notations:

- $A \looparrowright B$: propose all the outgoing relations of A as outgoing relations for the term B (other notation as $\mathbb{C}$ to copy relations and $\mathbb{D}$ to displace them are available but not used here);
- $A \circ\!\!-\!\!\!-\!\!\circ B$: a relation between A and B in any direction exists.

4 Experimentations and Discussion

Our experiments consisted in applying and assessing the schemes presented above on the entire lexical network. This has been once during one run. At the time of writing of this article, the JeuxDeMots consists in more than 6 800 000 relations betweeen 310 000 terms. Specifically, it contains over 150 000 hypernym *is-a* relations, 170 000 syn relations and 27 000 hyponym relations.

Relation type	**Proposed %**
is-a (x is a type of y)	6.2
has-parts (x is composed of y)	25
holonyms (y specific of x)	7.2
typical place (of x)	7.2
charac (x as characteristic y)	13.7
agent-1 (x can do y)	13.3
instr-1 (x instrument of y)	1.7
patient-1 (x can be y)	1
place-1 (x located in the place y)	9.8
place > action (y can be done in place x)	3.4
object > mater (x is made of y)	0.3

Table 1: Percentages of relation proposed per relation type globally for deduction and induction.

4.1 Assessing Deduction and Induction

We applied the inference engine on around **32 000** randomly selected terms having at least one hypernym or one hyponym and thus produced by deduction more than **2 700 000** inferences and produced by induction over **430 000** relation candidates. The threshold for filtering was set to a weight of 25. This value is relevant as when a human contributor proposed relation is validated by experts, it is introduced with a default weight of 25 (the choice of this particular value is arbitrary and could have been different). The transitive *is-a* (Table1) is not very productive which might seem surprising at first glance. In fact, the *is-a* relation is already quite populated in the network, and as such, fewer new relations can be inferred. The figures are inverted for some other relations that are not so well populated in the lexical network but still are potentially valid. The *has-parts* relation and the agent semantic role (the *agent-1* relation) are by far the most productive types.

Deduction	**% valid**		**% error**		
Rel type	**rlvt**	**¬ rlvnt**	**prem**	**excep**	**pol**
isa	76%	13%	2%	0%	9%
has-parts	65%	8%	4%	13%	10%
holonyme	57%	16%	2%	20%	5%
typ place	78%	12%	1%	4%	5%
charac	82%	4%	2%	8%	4%
agent-1	81%	11%	1%	4%	3%
instr-1	62%	21%	1%	10%	6%
patient-1	47%	32%	3%	7%	11%
place-1	72%	12%	2%	10%	6%
place>act	67%	25%	1%	4%	3%
obj>mater	60%	3%	7%	18%	12%

Induction	**% valid**		**% error**		
Rel types	**rlvt**	**¬rlvnt**	**prem**	**excep**	**pol**
isa	-	-	-	-	-
has-parts	78%	10%	3%	2%	7%
holonyme	68%	17%	2%	8%	5%
typ place	81%	13%	1%	2%	3%
charac	87%	6%	2%	2%	3%
agent-1	84%	12%	1%	2%	1%
instr-1	68%	24%	1%	4%	3%
patient-1	57%	36%	3%	2%	2%
place-1	75%	16%	2%	5%	2%
place>act	67%	28%	1%	3%	1%
obj>mater	75%	10%	7%	5%	3%

Table 2: On the left, number of propositions produced by *deduction* and ratio of relations found as true or false. On the right, Number of propositions produced by *induction* and ratio of relations found as true or false.

Table 2 presents some evaluations of the status of the inferences proposed by the inference engine through deduction and induction respectively. Inferences are valid for an overall of 80-90% with around 10% valid but not relevant (like for instance $dog \xrightarrow{has-parts} proton$). We observe that error number in premises is quite low, and errors can be easily corrected. Of course, not all possible errors are detected through this process. More interestingly, the reconciliation allows in 5% of the cases to

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

Table 3: The number of relations existing before application of the scheme and those proposed by the scheme. The statistics were made on the terms on which the scheme has proposed inferences

identify polysemous terms and refinements. Globally false negatives (inferences voted false while being true) and false positives (inferences voted true while being false) are evaluated to less than $0,5\%$. For the induction process, the relation *is-a* is not obvious (a lexical network is not reductible to an ontology and multiple inheritance is possible). Result seems about 5% better than for the deduction process: inferences are valid for an overall of 80-95%. The error number is quite low. The main difference with the deduction process is on errors due to polysemy which is lower with the induction process. To try to assess a baseline for those results, we compute the full closure of the lexical network, i.e. we produce iteratively all possible candidate relations until no more could be found, each candidate being considered as correct and participating to the process. We got more than 6 million relations out of which 45% were wrong (evaluated on around 1 000 candidates randomly chosen).

4.2 Assessing Relation Transfer

We applied the scheme of refinements relation transfer with three different support relations:
- RIS_R (synonym): the scheme applied with syn as support (in case of existence of B' the terms A' and B' share relations.)
- RIS_R (hyponym): the scheme applied with hypo (relations are shared from B or B' to A')
- RIS_R (hypernym): the scheme applied with R=hyper (relations are shared from A' to B or B').

RIS_R stands for *Relation Inference Schema with Refinements.*

Relation type	**syn**	**hypo**	**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

Relation type	**syn**	**hypo**	**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 4: On the left, relations proposed by type of the support relation and relation type of the conclusion. On the right, percentage of valid relations by type of the support relation and relation type of the conclusion.

Relation Transfer Productivity - Since the schema has a condition to be applied, the propositions (inferred relations) are made for only **6 349** terms fullfilling 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 3). The inferred relations are detailed by relation type in the left table 4. 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" 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. In the network, the most possessed relations of a term are typed with the associated relations. The amount of the relations proposed is related to the one existing in the network. If a relation type is quite populated in the network, fewer new relations can be inferred. The figures are inverted for some other relations that are not so well populated in the lexical network but still are potentially valid.

Relation Transfer Accuracy - The validation process was applied manually on a sample of around 1 000 propositions randomly choosen 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 4). The synonym version of the scheme has systematically the best accuracy for all the relation types. Some accuracy percentages are lower than others for some reasons. In certain cases, some outgoing relations of an hyponym do not suit for the hypernym. For example:
$\bullet A$: animal $\bullet A'$: animal>animalia $\bullet B(hypo)$: 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: $\bullet A$: cheese $\bullet A'$: cheese>dairy product $\bullet B(hypo)$: goat [1]
$\Rightarrow$ The inference scheme will propose the relation (*cheese>dairy product* $\xrightarrow{has-part}$ *teats*) which is wrong and thus because the term $goat$ is not yet refined into *goat>dairy product* and *goat>animal*.

From the figures, we can make the following observations. First, global results show that produced inferences are strongly valid with synonyms. The results are poorer with hypernyms and hyponyms (table 4) which is obvious regarding that with synonym, the terms exchanging relations are roughly at the same level of the taxonomic hierarchy which is not the case when they are related with an hyponym or hypernym relation.

5 Conclusion

We have presented some issues in inferring new relations from existing ones to consolidate a lexical-semantic network built with games and user contributions. To be able to enhance the network quality and coverage, we proposed an elicitation engine based on inferences (induction, deduction and relation transfer with refinements) and reconciliation. If an inferred relation is proven wrong, a reconciliation process is conducted in order to identify the underlying cause and solve the problem.

We focused our work on the transfer of relations related to word usage (refinements) with help of a support relation being either synonym, hypernym or hyponym. Unlike deduction and induction, the transfer scheme does not rely directly on the relation (*is-a*), but merely on terms that may be ontologicaly connected to the target. Experiments showed that relation transfer for refinements is quite productive (compared to deduction and induction), and is satisfying in correctness especially with synonym as support relation. The most obvisous reason is that in general a (quasi-)synonym is almost at the same level with the target term, and at least much more often than a hypernym or hyponym. User evaluation showed that wrong inferred relations (between around 20-15% of all inferred relations) are still logically sound and could not have been dismissed *a priori*. Relation transfer with refinements can conclusively be considered as a usefull and efficient tool for relation inference, and it may be really crucial as support for building information to be used in word sense disambiguation. In particular, it can help proposing hypernyms for the target term when they are missing, making possible further deductions or inductions. Hence, a virtuous circle may be initiated.

Still, the main difficulty of such approach relies in setting the various parameters in order to achieve an appropriate and fragil tradeoff between an over-restrictive filter (many false negatives, resulting in information losses) and a too lenient engine (many false postive, resulting in more human effort). The elicitation engine we presented through schemes based on deduction, induction and more precisely on relation transfer is an efficient error detector and a polysemy identifier. The actions taken during the reconciliation forbid an inference proven wrong or exceptional to be inferred again. Each inference scheme may be supported by the two others in particular for refinements, and if a given inference has been produced by more than one of these three schemes, it is almost surely correct.

[1] In french, some dairy products are called sometimes by the name of the producer animal, like $chevre(goat)$ for the cheese made from the goat's milk

An additional inference scheme, *abduction*, reinforced our inference engine and guided it through producing accurate new relations with an interesting accuracy. This 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. The abductive inferring layout supposes that relations held by a term can be proposed to similar terms. Abduction first selects a set of similar terms to the target term A which are considered as proper examples. The outgoing relations from the examples which are not common with those of A are proposed as potential relations for A 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 loop. This scheme was detailed in our paper (M. Zarrouk, EACL2014).

Researches are undertaken on (semi)automating the inference schemes or inference rules (scheme with just one or two unknown terms) discovery by our elicitation system. Enhancements are also considered on our previous schemes as for exemple defining the inference's scope especially in deduction and induction (example: what to do to avoid transferring invalid inferences from the term *animal* as *has-part wings* to its hyponyms like *cat* or *fish*).

We are also modelling a declarative query language that allows users to manipulate the lexical-semantic network and to apply our elicitation engine according to their needs while remaining focused on their request and without drifting in database access or linguistic domain.

References

von Ahn, L. and Dabbish, L. 2008. *Designing games with a purpose.* in 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.

Fellbaum, C. and Miller, G. 1988. *(eds) WordNet..* The MIT Press.

Krachina, O., Raskin, V. 2006. *Ontology-Based Inference Methods.* CERIAS TR 2006-76, 6p.

Lafourcade, M. 2007. *Making people play for Lexical Acquisition..* In Proc. SNLP 2007, 7th Symposium on Natural Language Processing. Pattaya, Thailande, 13-15 December. 8 p.

Lafourcade, M., Joubert, A. 2008. *JeuxDeMots : un prototype ludique pour l'ÃĪmergence de relations entre termes..* In proc of JADT'2008, Ecole normale supÃĪrieure Lettres et sciences humaines , Lyon, France, 12-14 mars 2008 .

Lafourcade, M., Joubert, A. 2012. *Long Tail in Weighted Lexical Networks..* In proc of Cognitive Aspects of the Lexicon (CogAlex-III), COLING, Mumbai, India, December 2012.

Lieberman, H, Smith, D. A and Teeters, A 2007. *Common consensus: a web-based game for collecting commonsense goals..* In Proc. of IUI, Hawaii,2007.12p .

Marchetti, A and Tesconi, M and Ronzano, F and 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. and Beckwith, R. and Fellbaum, C. and 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. 2010. *BabelNet: Building a very large multilingual semantic network..* in Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 11-16 July 2010.p 216-225.

Sagot, B. and Fier, D. 2010. *Construction d'un wordnet libre du franï£¡ais ï£¡ partir de ressources multilingues..* in Proceedings of TALN 2008, Avignon, France, 2008.12 p.

Thaler, S and Siorpaes, K and Simperl, E. and Hofer, C. 2011. *A Survey on Games for Knowledge Acquisition..* STI Technical Report, May 2011.19 p.

Sajous, F., Navarro, E., Gaume, B,. Prï£¡vot, 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., Y. Ng., A. 2006. *Semantic taxonomy induction from heterogenous evidence.* in Proceedings of COLING/ACL 2006, 8 p.

Velardi, P. Navigli, R. Cucchiarelli, A. Neri, F. 2006. *Evaluation of OntoLearn, a methodology for Automatic Learning of Ontologies.* in Ontology Learning and Population, Paul Buitelaar Philipp Cimmiano 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.

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. *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.

Zarrouk, M., Lafourcade, M. and Joubert, A. 2013. *InfÃľrences dÃľductives et rÃľconciliation dans un rÃľseau lexico-sÃľmantique.* 20Ãľme confÃľrence du Traitement Automatique du Langage Naturel 2013 (TALN 2013), 14 p.

Zarrouk, M.and Lafourcade, M. 2014. *Relation Inference in Lexical Networks ... with Refinements.* The 9th edition of the Language Resources and Evaluation Conference, 26-31 May, Reykjavik, Iceland, 6 p.

Zeichner, N., Berant J., and Dagan I. 2012. *Crowdsourcing Inference-Rule Evaluation.* in proc of ACL 2012 (short papers).