HAL
open science

Computing trees of named word usages from a crowdsourced lexical network

Mathieu Lafourcade, Alain Joubert

To cite this version:

Mathieu Lafourcade, Alain Joubert. Computing trees of named word usages from a crowdsourced lexical network. CLA'10: Computational Linguistics - Applications in IMCSIT'10: International Multi-Conference on Computer Science and Information Technology, Oct 2010, Wisla, Poland. pp.439-446. lirmm-00536793

HAL Id: lirmm-00536793

https://hal-lirmm.ccsd.cnrs.fr/lirmm-00536793v1

Submitted on 16 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proceedings of the International Multiconference on Computer Science and Information Technology pp. 439–446
ISBN 978-83-60810-22-4
ISSN 1896-7094

Computing trees of named word usages from a crowdsourced lexical network

Mathieu Lafourcade, Alain Joubert
LIRMM – Université Montpellier 2 – CNRS
Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
161, rue Ada – 34392 Montpellier Cédex 5 – France
Email: {lafourcade, joubert}@lirmm.fr

***Abstract*—Thanks to the participation of a large number of persons via web-based games, a large-sized evolutionary lexical network is available for French. With this resource, we approached the question of the determination of the word usages of a term, and then we introduced the notion of similarity between these various word usages. So, we were able to build for a term its word usage tree: the root groups together all possible usages of this term and a search in the tree corresponds to a refinement of these word usages. The labelling of the various nodes of the word usage tree of a term is made during a width-first search: the root is labelled by the term itself and each node of the tree is labelled by a term stemming from the clique or quasi-clique this node represents. We show on a precise example that it is possible that some nodes of the tree, often leaves, cannot be labelled without ambiguity. This paper ends with an evaluation about word usages detected in our lexical network.**

I. Introduction

In this paper, we describe an approach for acquiring lexical structures useful for word sense disambiguation (WSD). Having a word sense inventory is often mandatory prior to many text analysis application but those resources are quite rare especially for French. By text analysis, we mean here a task where word senses and relations amongst word senses are identified. Furthermore, being constructed either by experts (generally lexicographers) or from corpora, they barely reflect what people would think of. An adequate resource would then mostly refer to word usages than word senses (in the classical meaning of dictionaries) and would also try to evaluate the proper weight of an usage[1]. A very immediate usage of a term would have a higher weight than a rare one. We can then expect a quite large discrepancy between a proper inventory of usages and inventory of meanings (in dictionaries). Some usages would not be in dictionary either being elisions (like in French *sapin* for *sapin de Noël* – eng. *fir* for *Christmas tree*) or popular (like in French *caisse* as *car*). Some meaning would be absent of a *popular meaning inventory* when too technical, too rare or simply unknown by the vast majority of people. Can we try to define precisely what is the difference between a word sense and a word usage? Word usages are broader and would certainly include senses, but not the other way around. In French, for example, *sapin* (*fir*) beside being the tree, has a strong usage related to Christmas (*fir* as *Christmas tree*) and tends to become autonomous. A putative definition of a word usage could be a *specific meaning in a given context, popular enough to be spontaneously given by someone.*

The method and evaluation presented in this paper are based on a resource under construction but already freely available: the JeuxDeMots lexical network[2]. We first briefly remind the reader of the principles of two games which aim at building a base of relations between terms. The first of these two games (JeuxDeMots[3]) allows the construction of a lexical network, while the second game (PtiClic[2]) allows the user to strengthen associations acquired thanks to JeuxDeMots. With the network thus obtained, we tackle the problem of the word usage determination, by analysing the relations between every term and its immediate neighbours. The similarity between the various usages of the same term can be computed allowing us to build the classification tree of the usages of a term, the nodes of which being labelled.

[1]Definition for word usage is given several lines below, in this paragraph, and the weight of an usage is explained in section III.

[2] This resource is available at http://www.lirmm.fr/jeuxdemots/rezo.php for the lexical network and at http://www.lirmm.fr/jeuxdemots/diko.php for the obtained dictionary.

[3] JeuxDeMots and PtiClic are available at http://jeuxdemots.org and http://pticlic.org. An English version has recently been added, as well as a Thai, Japanese and Spanish versions (http://www.lirmm.fr/jeuxdemots/world-of-jeuxdemots.php)

Such a word usage tree structure as of primary interest for WSD should be evaluated against users before considering using it in applications. Furthermore, one of the objectives is to be able to connect the relations, not on the very terms (with ambiguities for polysemous terms), but on their usages (thus by clearing up lexical ambiguities).

II. Lexical network construction

The basic principles of JeuxDeMots (JDM) software, the game design, as well as the incremental construction of the lexical network, have already been described in [1]. A game takes place between two players, in an asynchronous way, based on the concordance of their propositions. When a first player begins a game, an instruction concerning a type of competence (synonyms, opposite, domains …) is displayed, as well as a term[4] T randomly picked in a base of terms. This player has then a limited time to answer by giving propositions which, to his mind, correspond to the instruction applied to the term T. The same term, along the same instruction, is afterwards proposed to another player; the process is then identical. For the same target term T and a same instruction (synonyms, domains, free associations…), we record the answers common to both players. Validations are thus made by concordance of the propositions between pairs of players. This validating process is similar to the one used by [2] to index images and, as far as we know, this has never been done in the field of the lexical networks. In Natural Language Processing, some other Web-based systems exist, such as *Open Mind Word Expert* [3] that aims to create large sense tagged corpora with the help of Web users, or *SemKey* [4] that exploits WordNet and Wikipedia in order to disambiguate lexical forms to refer to a concept, thus identifying a semantic keyword.

The structure of the lexical network we are building relies on the nodes and relations between nodes, as it was initially introduced by [5]. More precisely, JDM game leads to the construction of a lexical network connecting terms by typed and weighted relations[5]. These relations are labelled by the instruction given to the players and they are weighted according to the number of pairs of players who proposed them. The morphosyntactic category was not initially indicated in JDM. A recent evolution, published in [6], has allowed us to introduce the notion of refinement of a term. This refinement can depend on the meaning (case of polysemous terms) or on the morphosyntactic category of the term.

In a similar way to JDM, a PtiClic game takes place in an asynchronous way between two players. A target term T, origin of relations, as well as a cluster of words resulting from terms connected with T in the lexical network produced by JDM are proposed to a first player. Several instructions corresponding to types of relations (synonym, hypernym, hyponym, predicate relations like possible/typical agent, patient or instrument, part-of and substance, ...) are also displayed. The player associates words of the cluster with instructions he thinks correspond by a drag and drop. The same term T, as well as the same cluster of words and the same instructions, are also proposed to a second player. According to a principle similar to that set up for JDM, only the propositions common to both players are taken into account, thus strengthening the relations of the lexical network. Contrary to JDM, the players of PtiClic cannot suggest new terms, but are forced to choose among those proposed. This choice of conception has to allow to reduce the noise due to misspelt terms or to the confusions of meanings.

The collaborative building of resources by non-experts may induce some errors. In fact, as one may expect, we detected some of them, such as classical orthographic mistakes (eg: *théatre* for *théâtre*) or traditional confusions (eg: French singer *Dalida* with the biblical character *Dalila*) … These well-known mistakes are relatively rare and they can be manually detected.

According to the JeuxDeMots Web site, at the time of the writing of this paper, the lexical network contains around one million relations linking 221000 terms. Around 800000 games have been played corresponding to more than 13000 hours of cumulated play.

III. Similarity between usages of a term

A. Word Usage Determination

If a term T is polysemous, the terms which are directly connected with it (semantically connected in the lexical network) form several different groups, each of these groups constituting a word usage of T. The notion of word usage (often referred to as "usage") is much more accurate and relevant than the notion of meaning which, as shown by [7], is relatively poor when we refer to traditional dictionaries or to resources as WordNet. Our hypothesis is that the usages of a term correspond in the network to the various cliques this term belongs to. A clique is a set of terms constituting a fully connected subgraph in the lexical network. Two terms T_i and T_j belong to the same clique if there is at least one relation between T_i and T_j and at least one relation between T_j and T_i. Our approach is similar to the one developed by [8] from dictionaries of synonyms.

Why using the JeuxDeMots lexical network for our experiment instead of WordNet [9], EuroWordNet or WOLF [10] ? These resources are handcrafted contrary to the JeuxDeMots lexical network which is crowdsourced through some games, and by itself it is interesting to assess if common word usage can be identified. Similar approaches like [11] have been conducted on contexonyms but on

[4] A term can be a compound word (for example: *Christmas tree*) and each of its words may be a term (*Christmas* and *tree*).

[5] A relation can be thus considered as a quadruplet: origin term, destination term, type and weight of the relation. Between two same terms, several relations of different types can exist.

resources trained on very large corpora (and again not extracted from people).

For the clique identification, we take into account all relation types available in the lexical network. Of course we might certainly consider that they do not contribute equally to the induced word usage, but the principle of JeuxDeMots induce that the most important relations have the highest weights and that the most important relation types (for a given term) are the most populated. So, there is, *a priori*, no need to stress on specific relation type, as this information is already implicitly present in the network. Estimating the relevance of a usage consists in obtaining a measure of its importance both in terms of frequency and of lexical coverage. Considering the principle of the weighting of the relations, the weight of a usage is correlated with the weights of the relations between the terms of the clique which characterizes this usage. We have the following notations :

- C is a given clique for the term T (this is a set of terms and instances of relation).
- C_{all} is the union of all C (that can be seen as the full pseudo clique for T).
- W(C) is the sum of the weights of the relations between the terms of C.
- Card(C) is the number of terms in C.

So, for a clique C related to the term T, we define formally the *relevance* as :

$$Rel(C) = W(C) * \log(Card(C_{all})/Card(C))$$

The Rel measure for a clique may be seen as an adaptation of the tf/idf measure where, for the sake of simplicity, we have not divided by $W(C_{all})$. If there is only one clique, the relevance is equal to 0 and of course in that case we consider that there is only one usage. Figure 1 presents the obtained usages for the term *sapin* (*fir*), and the relevance of each of these usages.

0: 'sapin' 'fiacre' **REL = 52**
fir, hansom
1: 'sapin' 'cercueil' **REL = 55**
fir, coffin
2: 'sapin' 'montagne' **REL = 38**
fir, montain
3: 'sapin' 'épicéa' 'ginkgo' 'conifère' 'cèdre' 'mélèze' 'résineux' **REL = 66**
fir, spruce, ginko, conifer, cedar, larch,conifer
4: 'sapin' 'vert' 'arbre' **REL = 126**
fir, green, tree
5: 'sapin' 'épicéa' 'épinette' 'conifère' **REL = 59**
fir, spruce, spruce, conifer
6: 'sapin' 'aiguille' **REL = 43**
fir, needle
7: 'sapin' 'conifère' 'arbre' **REL = 139**
fir, conifer, tree
8: 'sapin' 'guirlande' 'Noël' **REL = 111**
fir, garland, Christmas
9: 'sapin' 'boule' 'boules' **REL = 51**
fir, ball, balls
10: 'sapin' 'boule' 'Noël' **REL = 108**
fir, ball, Christmas
11: 'sapin' 'Noël' 'sapin de Noël' 'sapin de noël' **REL = 84**
fir, Christmas, Christmas Tree, Christmas tree
12: 'sapin' 'Noël' 'fête' **REL = 152**
fir, Christmas, celebration
13: 'sapin' 'arbre' 'bois' 'forêt' **REL = 219**
fir, tree, wood, forest
14: 'sapin' 'arbre' 'bois' **REL = 148**
fir, tree, wood
15: 'sapin' 'conifères' **REL = 71**
fir, conifers

Fig 1 : 16 usages for the term *sapin* (*fir*) as found in the lexical network at the writing time.

The most relevant clique is {**sapin, arbre, bois, forêt**} *(fir, tree, wood, forest)* with a score of 219, followed by {**sapin, Noël, fête**} *(fir, Christmas, celebration)* with a score of 152.

B. Clique Similarity

The similarity between two objects can be defined according to [12] as being a function of their common characteristics with regard to all their characteristics. In NLP, we find several definitions of the similarity, for example [13] or [14]. More recently, [15] evaluated several different measures of lexical semantic relatedness, while [16] presents a general survey on this question. In our case, it corresponds to the ratio between the weight of the relations connecting two cliques and the total weight of the relations on all the terms of these two cliques. We note W(E) the weight sum of the relations between the terms of the set E (as in 3.1). The similarity between two cliques C1 and C2 will be equal to the *Jaccard* indice :

$$Sim(C1,C2) = W(C1 \cap C2) / W(C1 \cup C2)$$

We should note here, that the Jaccard indice is in our case applied on the set of relations and that the 'cardinality' of this set is the sum of the weights. Usually the Jaccard indice is applied on the true cardinality of the sets, considering equally all elements of the set. Figure 2 shows the similarities between the cliques of the term *sapin* (*fir*).

IV. Classification Tree of Word Usages

A. Construction

Our aim is to obtain a representation of the various usages of a term T in the form of a tree, with the root grouping together all the meanings of T and the branches corresponding to its various usages. Generally, most of terms possess several not separate cliques. In that case, the further away we go from the root of the tree, the more we

meet fine distinctions of usages. In fact, we build the tree of the usages of a term T according to a "bottom - up" method: from all of its cliques, that is, from its leaves and going back up to its root which groups together all the meanings of T. For that purpose, we apply an agglomerative hierarchical clustering algorithm: we merge the cliques, two by two, beginning with those whose coefficient of similarity is the highest: thus, we build quasi-cliques representing groups of usages, close during the first fusions, less and less close during the successive fusions. The merging algorithm ends when all coefficients of similarity are equal to zero.

The usage tree of a term is a structure expressing the refinements of its various meanings as deduced from the state of the lexical network. It thus constitutes a decision tree, a data structure which can be exploited for disambiguation. Furthermore, nodes of this tree are weighted allowing to identify usages that are the most common, which is both useful for guessing default cases and ordering usages from the most activated for people to the least activated.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0.04	0	0.08	0.04	0	0	0	0	0	0.01	0.03	0
3	0	0	0	1	0.31	0.66	0.06	0.54	0.02	0	0.02	0.02	0.02	0.2	0.18	0.36
4	0	0	0.04	0.31	1	0.29	0	0.75	0.03	0	0.03	0.03	0.03	0.68	0.72	0.15
5	0	0	0	0.66	0.29	1	0.09	0.56	0.03	0	0.03	0.03	0.03	0.16	0.13	0.48
6	0	0	0.08	0.06	0	0.09	1	0.05	0	0	0	0	0	0	0	0.33
7	0	0	0.04	0.54	0.75	0.56	0.05	1	0.03	0	0.03	0.03	0.03	0.64	0.66	0.46
8	0	0	0	0.02	0.03	0.03	0	0.03	1	0.14	0.69	0.6	0.77	0.01	0	0
9	0	0	0	0	0	0	0	0	0.14	1	1	1	1	0	0	0
10	0	0	0	0.02	0.03	0.03	0	0.03	0.69	1	1	1	1	0.01	0	0
11	0	0	0	0.02	0.03	0.03	0	0.03	0.6	1	1	1	1	0.01	0	0
12	0	0	0	0.02	0.03	0.03	0	0.03	0.77	1	1	1	1	0.01	0	0
13	0	0	0.01	0.2	0.68	0.16	0	0.64	0.01	0	0.01	0.01	0.01	1	1	0.06
14	0	0	0.03	0.18	0.72	0.13	0	0.66	0	0	0	0	0	1	1	0.08
15	0	0	0	0.36	0.15	0.48	0.33	0.46	0	0	0	0	0	0.06	0.08	1

Fig 2 : Similarity matrix between the cliques of the term *sapin* (*fir*).
The clique numbers are those from figure 1.

B. Labelling

The labelling of the various nodes of the usage tree of a term is made during a width-first search, that is, according to a "top - down" method. The tree root is labelled by the term itself. Every node of the tree is labelled by a term stemming from the clique or the quasi-clique this node represents; the selected term is the one whose sum of the weights of its relation with the root term is the highest, after eliminating all the terms labelling the nodes of the tree situated in a depth lower than that of the concerned node. Thus, it is possible that a node cannot be labelled if all the terms which define it have already been used in the labelling of nodes previously done. In this case, this node, but also its brother and all its successors, are not labelled. Figure 3 shows the tree obtained for the term *sapin* (*fir*).

C. Tree pruning

The relevance of an inner node of the tree is computed as defined above considering the quasi clique obtained during the merging. We ignored the usages of the tree which relevance is below a given threshold (empirically set to 50 in our experiment). This threshold correspond to the configuration where both terms have been associated to each other to only two pairs of player (one pair for each direction) and maybe it could have been accidental. In figure 3, in the tree for the term *sapin*, we discard the *sapin (montagne)* node at level 1 which relevance is only equal to 38 (clique number 2 in Figure 1). The nodes for *sapin* as *fiacre*, *cercueil*, *arbre* and *Noël* have respectively a relevance of 52, 55 and more than 300 for the last two.

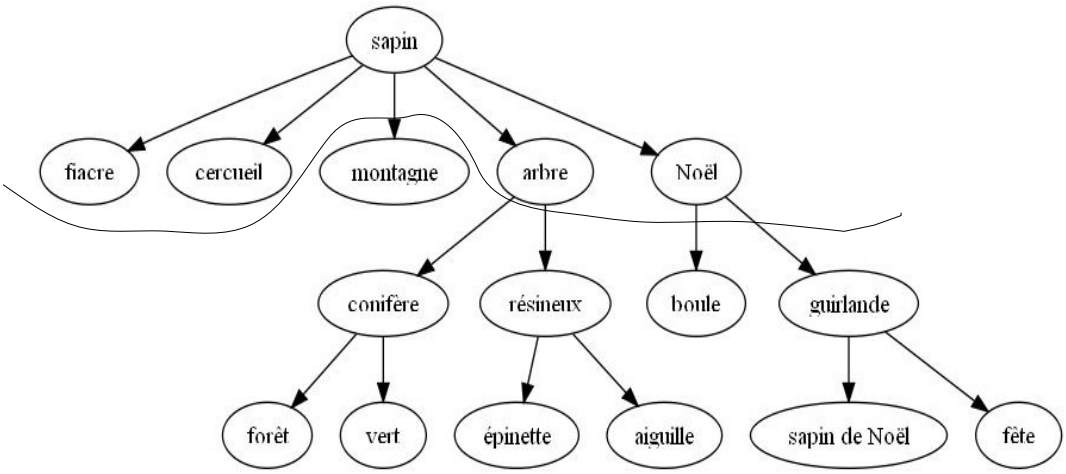

Fig 3 : Labelled word usage tree for the term *sapin* (*fir*). We pruned this tree by deleting nodes which cannot be labelled or which the relevance is too low (below 50 in our experiment)

V. Evaluation

A. Obtained Word Usages

Regularly, we examine the highly polysemous terms, and we ask an expert lexicographer to validate the meanings detected by JDM, after pruning the usage tree. The role of the expert does not extend beyond this validation: he doesn't add any meaning, even if he thinks some are missing, and he removes only those that obviously correspond to mistakes. As the time of writing this article, with the expert's help, we thus obtain 4412 validated word usages for 1263 terms, which correspond to a mean of around 3.5 usages per term. Out of these 1263 terms more than 80% are labeled as very common terms (at least one meaning should be known at the age of 12). The terms are mostly common nouns also some usages are tagged with other part-of-speech like verb or adjective. For example *dîner* in French can be at the same time a noun or a verb. The expert's role must go on to validate much more meanings detected in our lexical network.

B. User Evaluation

Evaluating the quality of such word usages is difficult, dramatically in the absence of adequate gold standard. As far as we aware of there is no such resources for French. If we had got any, a (semi-)automatic evaluation might have been feasible on a very much large scale. So, we decided to make a user based evaluation, trying to access qualitatively and quantitatively the word usages we obtained so far. We undertook the evaluation only at the first level of the usage tree computed for a given term.

We based our evaluation on naive users (i.e. not lexicographers) for two reasons. First, finding lexicographers for this task is not easy to say the least and certainly not more than few ones. Secondly, we wanted to have an evaluation confronting common people to our data. The idea is to identify word usages the same way (as a result at least) an average person would do. We remind here that the JeuxDeMots lexical network does not aim at being more than an average representation of associations between terms.

We asked 30 non-expert persons to undertake four slightly different tasks. Given a word and the set of associated named usages, they had to evaluate the number of missing usages and the number of supernumerary usages (either too specific or plain wrong). The task has to be done on two sets of different 50 words. On the first set, persons are not allowed to consult any dictionary (task Dict-), on the second they can check on dictionaries if they want to do so (task Dict+). The dictionary we proposed for reference is Wiktionary for French, but user where allowed to use any resources they want. Furthermore, the set of words is either taken from common words or non-common words (usually known or unknown at the age of 12).

An example of what is asked to the people is the following (translated for the purpose of this article):

> For the word *sapin*, we propose the following usages:
>
> - sapin(arbre) *(fir - tree)*
> - sapin(Noël) *(fir - Christmas)*
> - sapin(cercueil) *(fir - coffin)*
> - sapin (fiacre) *(fir - hansom)*
>
> Do you think some usages are missing, if yes how many ? Do you think some proposed usages are inappropriate, if yes how many ?

Word usages are ordered by decreasing relevance (as defined in section 3.1).

The four sets of words proposed to each evaluator are completely random (although verifying the constraints described before, that is to say either common or uncommon ones) and they are distinct (a given word may be present only in one set). Two different evaluators may have no distinct sets. The evaluators have all an age above 20 and had a similar proportion of 17 females and 13 males. The level of education was basically 2 years of university or more. The following tables present the collated result of this evaluation.

Common words	**Dict -**	**Dict +**
Missing usages	0.45	1.52
Added usages	0.66	0.37

Uncommon words	**Dict -**	**Dict +**
Missing usages	0.25	1.67
Added usages	0.76	0.28

Tables 1 & 2 : Average number of missing usages and added usages without and with the help of a dictionary for common (table 1) and uncommon (table 2) words used in our evaluation.

For users, added usages are those considered wrong (or at least far fetched). Missing usages are those which should have been present.

C. Result Analysis

How can we interpret those results? Without dictionary there is systematically less than one usage felt as missing or added. For the Dict- task, by debriefing the evaluators it appears that the added usage is quite often a proper usage that was unknown to the user (technical, old or rare). Conversely in the Dict+ task, the missing usages value rises as more usages, unknown to the user are found in the dictionary. It seems that globally we are missing much more usages than adding wrong ones. This is quite inline which the way the lexical network is constructed (by players indirect contribution). Missing usages are those quite specific, rare and basically unknown to users.

The task on uncommon words tends to strengthen this analysis. Indeed, without any dictionary people feel that they are more added usages than with common words and less missing usages. The result of the Dict+ is contravariant with the Dict-. Indeed, missing usage value rises and added usage value diminishes.

We can take several precise examples for illustrations. For the word *sapin* (a common word) we got :

sapin	**Dict -**	**Dict +**
Missing usages	0.5	0.9
Added usages	0.2	0.4

Table 3 : Average number of missing usages and added usages without and with the help of a dictionary for the term *sapin (fir)*.

The wiktionnary definitions are: **sapin** *masculin*

1. (Botanique) Arbre conifère résineux de la famille des abiétinées à aiguilles persistantes, au tronc droit, dont le fruit est un cône.
2. Bois de cet arbre utilisé en menuiserie.
3. (Par métonymie) Cercueil.
4. (Familier) (Vieilli) Fiacre.

Some people where doubting about the *sapin (fiacre)* usage although this is a correct one. The usage of *sapin* as wood (mater) is missing but roughly only one person out of two have been thinking of it without checking in a dictionary. The added usage compared to the dictionary is the *sapin (Noël)* although it is present as a locution.

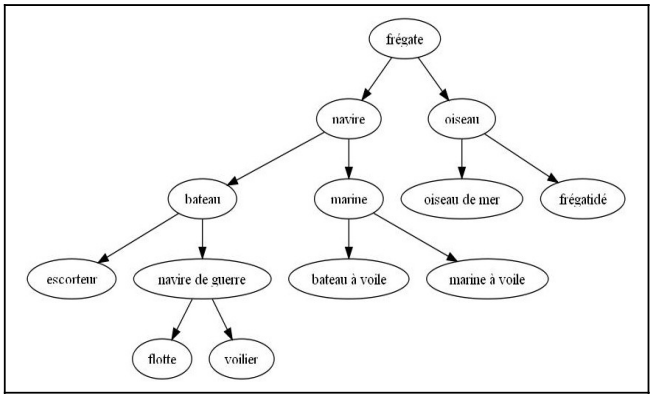

Fig 4 : Usage tree for the term *frégate (frigate)*.

For the word *frégate*, we found out the following usages:

- frégate (navire) *(frigate boat)*
- frégate (oiseau) *(frigate bird)*

frégate	**Dict -**	**Dict +**
Missing usages	0.15	0
Added usages	0.1	0

Table 4 : Average number of missing usages and added usages without and with the help of a dictionary for the term *frégate (frigate)*.

The wiktionnary definitions are: **frégate** *féminin*

1. (Histoire) (Marine) (Militaire) Bâtiment de guerre qui n'avait qu'une seule batterie couverte et qui portait de vingt à soixante bouches à feu.

2. (Zoologie) Oiseau de mer palmipède, d'une très grande envergure, et qui saisit à la surface de l'eau les poissons dont il se nourrit.

For *frégate,* some people made the distinction between the ancient boat and the modern boat. On a rare occasion the evaluator was doubtful on the bird meaning. Comparing with the dictionary we got an exact match.

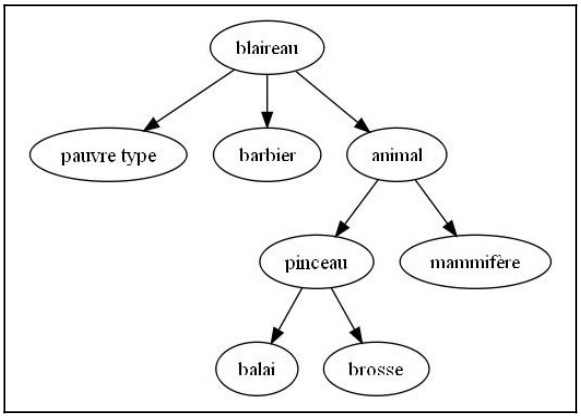

Fig 5 : Usage tree for the term *blaireau (badger, dork, shaving brush)*.

For the word *blaireau*, we found out and proposed the following usages:

- blaireau (animal) *(badger)*
- blaireau (pauvre type) *(dork)*
- blaireau (barbier) *(shaving brush)*

blaireau	**Dict -**	**Dict +**
Missing usages	0.3	1.1
Added usages	0	0

Table 5 : Average number of missing usages and added usages without and with the help of a dictionary for the term *blaireau (badger, dork, shaving brush)*.

The wiktionnary definitions are: **blaireau** *masculin*

1. Mammifère omnivore, bas sur pattes, au pelage noir, gris et blanchâtre, qui se creuse de profonds terriers.
2. (Arts) Brosse en poils de cet animal dont se servent les peintres et les doreurs.
3. Pinceau garni de ces poils dont on se sert, en se rasant, pour étaler et faire mousser le savon.
4. (Argot) (Vieilli) Nez.
5. (Argot) Individu grossier et antipathique ; imbécile, idiot.

For *blaireau* some people found that the *(painting) brush* (meaning 2 of wikipedia) is indeed missing. Most people missed the *nez* (nose – meaning 4) meaning, which is quite old.

All in all, although being preliminary, those results are very encouraging both on the soundness of the method for determining and naming word usage and the quality of the resource collected so far (although evaluating this resource was not the primary goal of this paper). They seem to correspond to what people know and not specifically to some resources made by lexicographers or experts.

What is the effect of the label on the evaluation? If for a given word usage a different label would have been chosen to which extend the results might be modified? In fact there is no much choice for a reasonable label, and generally choosing a substitute like an hyperonym (for example animal instead of bird, in case of frigate) does not alter the results. Of course, this is less and less true as we go deeper in the tree, but also there is less and less choices (if we stick to our labeling approach described in 4.2). Perhaps a deeper evaluation on this particular point should be conducted.

What can we do with the unlabeled usages? So far the answer is simple: nothing. But we should keep in mind that the network is in constant evolution and that some cliques existing now may be fusionned in the near future due to the players activity, or on the contrary being reinforced with new terms allowing them to be labeled.

VI. Conclusion

Viewing the results first obtained, these trees of named usages seem to correspond in their main structures to those a human non-expert would build. In particular, the main branches, directly stemming from the root, correspond in the majority of cases to the meanings of the root term as we could find them in a dictionary. These main branches are subdivided into sub-branches which are so many refinements in the usages.

In their detailed structures however, we notice elements that are different from what a human would have written as we showed on examples above. Are these differences (can we really speak about abnormalities?) due to our method of construction of the trees of the labelled usages with help of players who are not experts, or are they due to the fact that the lexical network is not "complete" enough yet? Moreover, do not the word usage trees sometimes distinguish too subtle refinements? Using the similarity between nodes, how would it be necessary to prune them?

A strong perspective of our work is to propose to the authors of the JeuxDeMots game to insert the identified word usages and proposed them to the players. Hence, the word usages are going to be associated to other terms of lexical database. It would be interesting to assess whether or not reapplying our algorithm leads to some convergence as expected. If it is not the case for a given term, either the usage identified or the label (or both) are not appropriate and should be revised.

VII. References

[1] M. Lafourcade and A. Joubert (2008) Détermination des sens d'usage dans un réseau lexical construit grâce à un jeu en ligne, *Conférence sur le Traitement Automatique des Langues Naturelles (TALN'08)*, Avignon, pp. 189-199

[2] L. von Ahn and L. Dabbish (2004) Labelling Images with a Computer Game, *ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 319-326.

[3] R. Mihalcea and T. Chklovski (2003) Open Mind Word Expert: Creating Large Annotated Data Collections with Web Users' Help, *Proceedings of the EACL 2003 Workshop on Linguistically Annotated Corpora (LINC 2003)*, Budapest.

[4] A. Marchetti, M. Tesconi, F. Ronzano, M. Rosella and S. Minutoli (2007) SemKey: A Semantic Collaborative Tagging System, *Proceedings of WWW2007*, Banff, Canada

[5] A. Collins and M. R. Quillian (1969) Retrieval time from semantic memory, *Journal of verbal learning and verbal behaviour*, 8 (2), pp. 240-248.

[6] M. Lafourcade and A. Joubert (2010) Détermination et pondération des raffinements d'un terme à partir de son arbre des usages nommés, *Conférence sur le Traitement Automatique des Langues Naturelles (TALN'10)*, Montréal, (to be published)

[7] J. Véronis (2001) Sense tagging: does it make sense?, *Corpus linguistics' 2001 Conference*, Lancaster, U.K.

[8] S. Ploux and B. Victorri (1998) Construction d'espaces sémantiques à l'aide de dictionnaires informatisés de synonymes. *TAL*, 39(1), pp. 161-182.

[9] C. D. Fellbaum. (1998) WordNet: An Electronic *Lexical Database*. MIT Press, New York.

[10] D. Fišer and B. Sagot (2008). Combining multiple resources to build reliable wordnets. In *TSD 2008*, Brno, Czech Republic

[11] H. Ji, S. Ploux and E. Wehrli. (2003) Lexical knowledge representation with contexonyms. In *Proceedings of the 9th MT summit*, pp. 194-201

[12] A. Tversky (1977) *Features of similarity*, Psychological Review, 84, pp. 327-352

[13] C. D. Manning and H. Schütze (1999) *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge.

[14] C. Fairon and N. D. Ho (2004) Quantité d'information échangée : une nouvelle mesure de la similarité des mots, *Journées internationales d'Analyse statistiques des Données Textuelles (JADT'04)*, Louvain-la-Neuve (Belgique).

[15] A. Budanitsky and G. Hirst (2006) Evaluating WordNet-based Measures of lexical semantic Relatedness, *Computational Linguistics*, 32(1), pp. 13-47.

[16] P. D. Turney and P. Pantel (2010) From Frequency to Meaning: Vector Space Models of Semantics, *Journal of Artificial Intelligence Research*, 37, pp. 141-188.