

Long Tail in Weighted Lexical Networks

Mathieu Lafourcade Alain Joubert

LIRMM, Montpellier, France

mathieu.lafourcade@lirmm.fr, alain.joubert@lirmm.fr

ABSTRACT

Lexical networks can be used with benefit for semantic analysis of texts, word sense disambiguation (WSD) and in general for graph-based Natural Language Processing. Usually strong relations between terms (e.g.: cat --> animal) are sufficient to help for the task, but quite often, weak relations (e.g.: cat --> ball of wool) are necessary. Our purpose here is to acquire such relations by means of online serious games as other classical approaches seems impractical. Indeed, it is difficult to ask the users (non experts) to define a proper weighting for the relations they propose, and then we decided to relate weights with the frequency of their propositions. It allows us to acquire first the strongest relations, but also to populate the long tail of an already existing network. Furthermore, trying to get an estimation of our network by the very users thanks to a tip of the tongue (TOT) software, we realized that they rather tend to favor the relations of the long tail and thus promote their emergence. Developing the long tail of a lexical network with standard and non-standard relations of low weight can be of advantage for tasks such that words retrieval from clues or WSD in texts.

KEYWORDS : LEXICAL NETWORK, LONG TAIL, GAME WITH A PURPOSE, TIP OF THE TONGUE SOFTWARE, TYPED RELATIONS, WEIGHTED RELATIONS, WSD

Introduction

Lexical/semantic networks are very precious resources for NLP applications in general and for Word Sense Disambiguation (WSD) in particular. Their construction is delicate as automated approaches from corpora may have various shortcomings (mainly high noise level and/or low recall) and a manual approach may be long, tedious, costly and of unsatisfactory quality or coverage. A way of handling the building of such resources can be direct crowdsourcing (as contributive approaches) or indirect crowdsourcing through for instance serious games.

What is a long tail in a lexical network?

A lexical/semantic network (thereafter dubbed JDM) for French is under construction with methods based on popular consensus by means of games with a purpose named JeuxDeMots (Lafourcade 2007). Thus, in 5 years, a high number of players lead to the construction a large scale lexical network for the French language (currently more than 240 000 terms with around 1.4 million semantic relations) representing a common general knowledge but also including word senses referred as *word usages* (Lafourcade and Joubert, 2010). The relations of the lexical networks created this way are directed and typed, with classical ontological relations (like hypernym, hyponyms, part-of, whole, material/substance, ...), lexical relations (synonyms, antonyms, lexical family, ...);

semantic roles (agent, patient, instrument, ...) and less standard relations (typical location and time, cause, consequence, ...). Furthermore, relation occurrences are weighted which constitutes a quite original aspect in the lexical network domain exemplified by (for example) WordNet (Miller, 1990). The interpretation of a weight might be difficult but can be related to the *strength* of the relation as collectively perceived by speakers/players. The weight computation is done by emergence along with the gaming activity. Obviously by intuition, the relation *cat* --> *animal* is stronger than *cat* --> *ball of wool*, none withstanding their types.

The lexical network has been made available (at <http://jeuxdemots.org>) and free to use by their authors, giving the research community a resource to play with. The question of the evaluation of its quality, usability in WSD and word recollection (Tip of the Tongue problem), and distributional properties are the main subjects of this article. One specific question is whether low weight but still important relations can be captured by some similar approaches and to which extent they are useful.

We observed that many (if not most) relations in JDM are “frontal/direct/obvious” relations (e.g.: *chat* --> *feline*), but some others are more farfetched/indirect. We wish to evaluate but also find practical ways to *densify* the network increasing the number of “indirect” relations (e.g.: *chat* --> *allergy*) belonging to the long tail. To do so, we use a TOT tool in a *taboo mode*, that is, refraining from using the strongest relations.

In a first section, we will briefly remind to the reader the principles of *long tail* and the link with the network construction. Then, we introduce our TOT (tip of the tongue) tool, named AKI and we will explain the *taboo mode*, and show how it leads to *densifying* the JDM network. An evaluation of the long tailed network obtained is done for AKI and for a simplified WSD task.

1 Long tailed lexical networks

Lexical networks, either general or specialized, are quite well known, especially with the advent of WordNet (Fellbaum, 1998). But relations in those lexical networks are not weighted, that is to say relations between terms are just enumerated and being viewed as equivalent in their influence (not considering their type). Introducing weights to relations to discriminate between *strong* and *loose* relations seems interesting but leads to also critical issues like: *how it could be done in practice and how to evaluate the obtained lexical network relatively to weights?* Propagation algorithms in WSD can take advantage of weighted relations, and especially in case of loose but numerous connexions between words of the text.

In (Sigman and Cecchi, 2002), a study of the organisation of the WordNet lexicon showed that the statistical distribution of the relation shows long tail behaviour, although they are not weighted. In fact, the study focused on the relations distribution amongst terms of WordNet, not of the distribution of the relations weights. Gaume (2008) studied various lexical networks and particularly graph of synonyms, and showed that they are “Small worlds graphs”, and as such amongst other properties, having a long tail in the relation distribution relatively to terms. But again, such long tail doesn't relate to the strength of the relations by themselves, even they are highly applicable between synonyms.

Some works aim at introducing weights in lexical network and especially WordNet. generally weights are added to synsets (and not relations between synsets) for handling default cases in WSD tasks. Such approaches relate generally to term frequency or various evaluation of terms pairs computed in the basis of the network itself. For instance (Boyd-Graber , 2006) and (Budanitsky & Hirst, 2001) amongst others, added numerically evaluated WordNet relations, weights being computed from various similarity measures. Weights are generally added either by asking people to evaluate the strength of term pairs, or 3-uples (when a relation type is added, like hypernym, synonymy, cause, consequence, etc) by giving a value on a closed scale (between 0 and 100, for example), or automatically by counting occurrences of such pairs from corpora.

1.1 Which Tail to Look at in Weighted and Typed Lexical Network?

The concept of *long tail* has been first popularized by (Zipf, 1965) for word occurrences in texts. Also in different domains, (Anderson, 2004) actually coined the phrase *long tail* about selling strategies of providing a large number of unique items in small quantities of each, usually combined to selling less popular items in large quantities. More precisely in our context, a long tail is a statistical property that a large share of a population belong to the tail of a probability distribution (larger than a normal "Gaussian" distribution) usually related to a power-law distribution.

The tail in a weighted lexical network is

the lower part of the distribution of relation weights for a given term

It is not the distribution of relations amongst terms, nor the distribution of term weights (if there is any). The tail can be considered with advantage separately for incoming or outgoing relations, as relations or even free associations are seldom symmetric. A question arises as when does the tail start in the distribution? The answer to this question is highly debatable and falls outside the scope of this paper. A simple (if not simplistic) approach is to consider that the tail starts at the point where

the cumulated weight of the relations of the tail **equals**
the cumulated weights of the relations which do not belong to the tail.

For example, in figure 1 is shown the distribution of outgoing relation weights for the term *chat* (eng. *cat*) in the JDM network. The pike (at around 45 on the x-axis, around 9,5% of the relation number) is an indication of the limit where the surface below the curve at the left of the pike is equal to the surface below the curve at the right. In this case, the first 45 relations have together the same importance than the rest of 405 relations.

However, in WSD we generally consider than the strongest relations (those on the left of the pike, in what is called the *belly zone*) are able to disambiguate around 70-75% on the ambiguity. The 25-30% could be solved with relations of the long tail, of course only if they are available in the knowledge base. None withstanding these figures (some literature would rather refer to the 20/80 rule), capturing the long tail is not only a challenge but a requirement to increase resolution percentage of WSD.

Terme 'chat' et relation all sortantes (442 données)

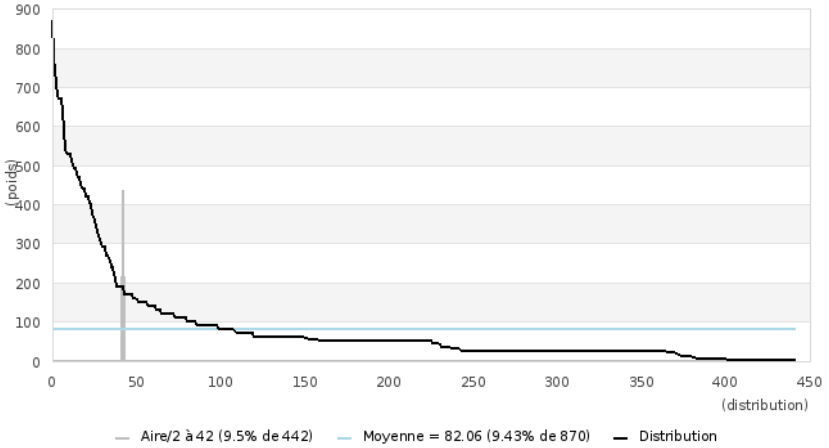


Figure 1a: distribution of outgoing relations for the term *chat* (eng. *cat*). The x-axis is the rank of the relations, the y-axis is the strength (weight). The frontier between the *belly* and the *tail* of the curve is indicated by the pike. On the left, the belly part of the curve stops after the first 9.5% of strongest relations. The tail in this case covers the 90.5% weakest relations.

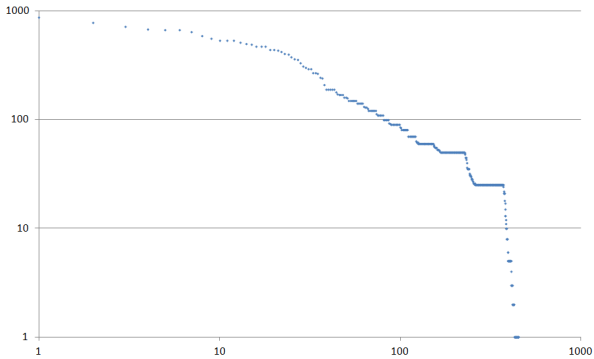


Figure 1b : the log-log version of the Figure 1a.

Put another way, Figure 1 can be interpreted, that the descriptive impact (in terms of weights) of the remaining 90.5% of the relations is equivalent that that of the first 9.5%. The curve reminds a zipfian power law (and is usually presented under a doubly-

logarithmic scale, but the issue here is to pinpoint the frontier pike between the belly and the tail) or perhaps more precisely a Mandelbrot law of the form : $K/(a+bn)^c$. However, we should stress this is not because such a curve is zipfian in shape that the data are actually related to a power law. Moreover, knowing the actual distribution law of the relations is by no mean any help in either the construction of the lexical network nor its use in lexical assistance or WSD. The question of lexical assistance have been largely presented in (Zock et al., 2010) as a difficult problem by itself. Indeed, it is not that a word is contained in such a resource that it is *de facto* easy to retrieve either by a speaker native of not, nor by any automated process.

1.2 Long Tailed Lexical Network Construction

The basic principles of JeuxDeMots (thereafter JDM) software, the game design, as well as the incremental construction of the lexical network, have already been described in (Lafourcade and Joubert, 2010). A game takes place between two players, in an asynchronous way. For the same target term¹ T and a same instruction (synonyms, domains, free associations ...), the answers common to both players are recorded. Validations are thus made by concordance of the propositions between pairs of players.

This validating process is similar to the one used by (von Ahn and Dabbish, 2004) to index images and by (Lieberman et al., 2007) to collect *common sense knowledge*. As far as we know, this is the first time it is done for lexical/semantic networks. However, using games for collecting resources of use in NLP is nowhere new, as (Chamberlain et al.) used it for anaphora annotations and (Mihalcea and Chklovski, 2003) for annotating corpora, to name a few.

The structure of the lexical network built in JDM relies on the nodes and relations between nodes, as it was initially introduced by (Collins and Quillian, 1969) and more recently explicated by (Polguère, 2006). More precisely, JDM game leads to the construction of a lexical network connecting terms by typed and weighted relations², some of them being quite non-classical. 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. Also similar at first sight, this a strong departure from collecting concurrences (typed or not) form corpora. Indeed, there is less guarantee, if any, that term associations extracted from corpora faithfully reflect what people have in their mind than asking them directly.

In a similar way to JDM, a PtiClic game (Zampa and Lafourcade, 2009) 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 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

¹ A term can be a compound word (for example: *Christmas tree*)

² 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.

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 design choice should allow to reduce the noise due to misspelt terms or to meaning confusion. There are at least two aims to this game: 1) to make the weights of the relation more reliable, and 2) to cast freely associated terms to more specific relations when possible. The first one is crucial as it counterbalances a strong bias in JDM: people tend to *over propose* terms to be associated.

It is generally assumed that when a relation holds between two terms, it is of only one type. However this should be mitigated as polysemy comes into play. For example, *café* can be located in a *café* (the *beverage* and the *place*, respectively), *café* can be made of *café* (*beverage* and the *plant/grain*). Some relation might not always be clearly distinct : is a *seat* part of a *car* or located in a *car*, or both? For semantic roles, it is quite common that an agent can also be the patient of a predicate (an *animal* can kill or be killed).

With the help of more than 3000 players, relations between pairs of terms have been collected, most of them being spontaneous, and thus "frontal" ones. Other "indirect" relations, are more uncommon, which seems quite logical considering the network creating mode (consensus filtered by player pairs). More formally, a clue can be said to as frontal for a target term if it belong to the belly of the distribution curve of that target.

Finally, looking at actual weight values isolated is of little significance. Instead comparing at least two values, for the same term and the same relation is of interest and may have meaning. Some terms are more played than others for various reasons (popularity, funniness, etc.), and tend to have higher strength values. The more played a term, the more reliable are the distribution of its relations and their *relative* values.

2 A Tip of the Tongue System: AKI

The questions we answer are the following: for a given term are its relations with other terms able to characterize it in a unique way? When it is the case, is it useful for a Tip of the Tongue Software? Such a tool aimed at helping someone retrieving a word that is "on the tip of the tongue" by the help of clue words. As the user is supposedly unable to retrieve the target word, he can only provide words that are related. Those words are the clues given to the system.

If the answer to the first question is positive, any term may be found via one or several reduced sets of typed clues. A tool helping the resolution of "word on the tip of the tongue" is a way to undertake the evaluation of the lexical network. Through such a tool made available on the web, the evaluation can thus be made permanent in time and rely on a large number of evaluators (not necessarily knowing that they are part of a global evaluation process).

The system we developed (named AKI) is a tool for helping retrieving some word on the tip of the tongue. Alternatively, it can be viewed as a game, whose goal is to make the system find a given word through clue, or to trick it.



Figure 2 (a and b): examples of AKI plays. In the first play (on the left), the clues given are *cinéma (movie)*, *ville (town)* and *Bollywood*, leading the system to propose in turn *film*, *place* and finally *Bombay*. In the second play (on the right), the clues are *film*, *salle (room)*, and *pop-corn* leading in the end to *cinéma (as movie theater)*.

Figures 2 are typical AKI games. At the stages displayed, the player, can either click on the button "C'est la bonne réponse" (Eng. *This is the proper answer*) if the proposition made by the system is the target term, or introduce another clue to get another proposition. The second plays, lead to a specific meaning of the word *cinéma* which may relate in French to *movie* or *theater*.

Players can introduce typed clues. A type relates to the kind of relation holding between the clue and the target word. For example, a clue of the form *:isa town*, indicates that the target word is a town. When the clues are not typed (as in the above plays), they are assumed to be related to the target no considering any specific relation type. The available relations types that can be chosen by players are as follows:

:isa	Hypernym, <i>:isa dog</i> means the target word is a <i>dog</i>
:hypo	Hyponym, <i>:hypo eagle</i> , means that the target word is an hypernym of <i>eagle</i>
:syn	Synonym, the target word and <i>clue</i> are synonyms.
:anto	Antonym, the target word is antonym of <i>clue</i> . For example, <i>:anto cold</i>
:subst	The target word has <i>clue</i> as substance. For example, <i>:subst silver</i>
:loc	The target word can be found in <i>clue</i> . For example, <i>:loc garden</i> , <i>:loc desert</i>
:locfor	The target word is a location for <i>clue</i> . For example, <i>:locfor money</i>
:carac	The target word has <i>clue</i> as a property. For example, <i>:carac cold</i>
:part	The target word has <i>clue</i> as part. For example, <i>:part wheel</i>
:partof	The target word is a part of <i>clue</i> . For example, <i>:partof car</i>
:do	The target word can do <i>clue</i> . For example, <i>:do roar</i>
:patientof	The target word can be an patient of <i>clue</i> . For example, <i>: patientof paint</i>
:cause	The target word can cause <i>clue</i> . For example, <i>:cause disease</i> .
:hascause	The target word is a consequence of clue. For example, <i>: hascause virus</i>

The reader can refer for example to (Morris and Hirst, 2004) for a discussion of non-classical semantic relations and their relevance for NLP.

2.1 Principle and General Algorithm

When viewing AKI as a game, the user tries consciously to make the computer guess a term, supplying, one by one, a succession of typed clues. After each clue, AKI makes the most probable proposition. If it corresponds to the searched term, the user confirms the proposition as the proper one; otherwise he introduces a new clue. This dialogue goes on, until either AKI finds the target term, or gives up asking the user to supply the solution. The algorithm relies both on the intersection of sets of terms activated by the clues and the fuzzy set of concepts linked to the clues.

The algorithm is based on manipulating sets of weighted words (named thereafter *lexical signatures*). We call a *clue* a term proposed by the user for the system to guess what could be the term to be found (called thereafter *target term*). Finally, we call a *proposition*, a term returned by the system from a set of clues.

From the first clue i_1 , a lexical signature is computed on the basis of what can be found in the lexical network: $S(i_1) = S_1 = t_1, t_2, \dots$ where the t_i are the terms related to the clue and sorted by descending activation (weight). By default, we consider all terms in the lexical network to be eligible as propositions and potential target terms. Put another way, t_1 is the term for which the sum of all relations related to the clue i_1 is the strongest. The first proposition made by AKI, p_1 is this term. The player is supposed to acknowledge it, if it is the target term, otherwise he/she is invited to propose another clue. In this case, the clue and the proposition is removed from the signature : $S'_1 = S_1 - \{p_1, i_1\}$.

With the second clue i_2 , the next lexical signature is computed : $S_2 = (S'_1 \cap S(i_2)) - i_2$. The generalized formula at stage n is :

$$S_n = (S'_{n-1} \cap S(i_n)) - i_n \quad \text{and} \quad S'_n = S_n - p_n$$

where i_n is the n -th clue given by the user and p_n the n -th proposition returned by AKI. With such a process, the size of signatures steadily diminishes as clues are added. The weight of each term of the signature is then the geometric mean of the weight of this term in the previous signatures.

If the signature becomes empty, the system has not found the target term. We could stop the process at this stage, but it is more valuable to set a recovering procedure which will try a simple heuristic. In this case, a boolean union of signatures are made instead of intersections:

$$S_n = (S'_{n-1} + S(i_n)) - i_n \quad \text{and} \quad S'_n = S_n - p_n$$

The weight if a term in the signature is then the sum of its occurrences in the previous signatures. This is a form of majority vote, where the proposal with the most votes is returned by AKI. This recovery induce a form of learning for the system as if the target term is found this way, as unlinked clues are added in the lexical network. We have found that using the recovering procedure two times before making AKI giving up, leads to satisfactory results. Be more lenient then the system tends to propose very general and too loosely related terms, be more strict the system tend to learn less or not at all.

About $\frac{1}{4}$ of the games concern common words and are played with "indirect" clues. The other games concern non common words, often connected to the current events, and are

played with “frontal” clues. Thus, as with JDM, most of the created relations are “frontal” ones.

2.2 AKI in Taboo Mode

As we find out, the JDM network contains mainly “frontal” relations, and we wish to extend it by creating or reinforcing “indirect” ones. In other words, we would like to increase the population of the long tail.

The aim of this work is to make the system guess a target term, without using clue terms which are the most strongly connected with the target term in the lexical network. We generally limit this list of *forbidden* (or *taboo*) terms to the first 20. It means clues given by the user cannot be any of these terms, and thus the user has to give other clues, less strong connected with the target term and belonging to the long tail. Using this network extension, it increases the recall of the system.

How to play in the taboo mode? In AKI, players has access to a list of recently played words, guessed or not. They can then choose one of these terms to make AKI guess it, avoiding as clues the terms indicated as taboos (forbidden by the system). These are in fact the terms in the *belly* in the lexical network. Alternatively, the player can send by email a term of its choice to be played to another person. *Tabooing* allows either to create new relations, or to strengthen already existing but relatively weak relations.

Figure 3 (a and b) shows a typical game under taboo mode. The target term along with the forbidden clues is first presented to the user. The player succeeded in making the system find the term *Bollywood* not using the forbidden clues.

How the taboo approach affect the relation frequency (or strength)? We can wonder that explicitly excluding the most common terms we might as well influence the *natural strength* in an artificial way. In the experiments we conducted, it has been observed that people *do not only* play in taboo mode, and that strongest and most immediate relations have their weight increased as well. The distribution curve (as exemplified in Figure 1) is globally pushed upward, revealing new more distant and low weight relations.



Figure 3 (a and b): AKI play with taboo words. On the left, the target term is Bollywood and the forbidden clues are *Bombay*, *Inde (India)*, *cinéma (movie)*, *danse*, *indien (indian)*, *cinéma>art (cinema as art)*, *film*, *bollywood* (no upercase) *cinéma indien (indian movies)* and *Hollywood*. On the right, the user made the system find the target term without using those forbidden clues, but with *acteur (actor)*, *hindi* and *Mumbai*.

3 Evaluation for the long Tailed Network

With AKI, more 15000 games were played creating more than original 80000 relations that were not part of the network beforehand. Also, around 1500 new terms have been introduced. We evaluated the impact of long tail relations in two contexts: 1) the evolution over time of the retrieving capability of AKI and 2) under a WSD task.

3.1 Performances as a Tip of the Tongue Tool

The performance of the AKI tool in properly guessing terms is found to be around 75% with an evaluation undertaken during around 18 months. That is to say 11545 out of the 15895 game sessions played ended by AKI successfully guessing the target term of the player. On a smaller scale (3000 games) , we proposed the very same games where people were supposed to replace AKI in order to try to find target terms from clues. The global performance of people was only 48%. This is very interesting especially considering that the clues given to people, were exactly those given by people to AKI. It can be interpreted that the system is better at guessing from clues given by people, than people to guess their own clues. The question that remains is to know if achieving 75% success rate is enough for a useful Tip of the Tongue Tool?

In fact, when used as a tool, people tend to give frontal (more straightforward) clues and were not willingly trying to tick the system. In this case, actual performances are much higher than 75%. Nevertheless, these facts have been collected from people that were using AKI as a TOT tool and a large scale planned experiment might be quite difficult to set up.

Another question, left open, is whether 75% of success rate is by itself a limit. Certainly, we cannot expect to achieve a 100% rate, considering new incoming words over time.

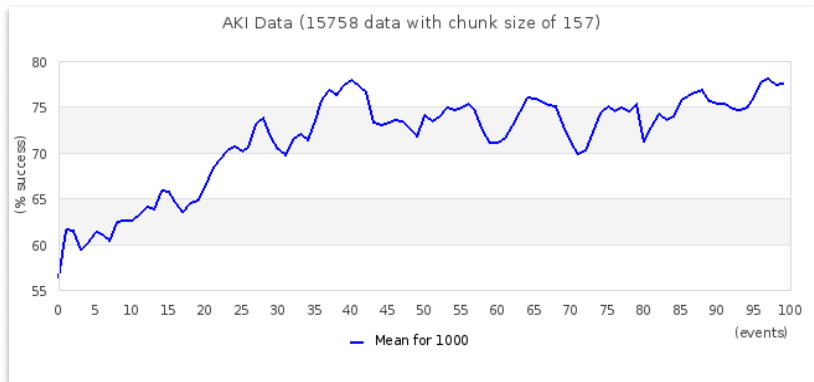


Figure 4: Evolution of AKI success over time. The x-axis is the number of games (by segment of 157 games). The curve shows the success rate at guessing the proper term that a user has in mind from the clues he/she giving to the system.

3.2 Performances for WSD

The purpose of the evaluation on WSD was to assess the impact of the network in the case of WSD when viewing this task as a guessing problem almost identical to the guessing game presented above. A full presentation of various WSD techniques is beyond the scope of this paper, the interested reader can refer to (Navigli, 2009) and for a more general account of graph-based NLP to (Mihalcea and Radev, 2011).

We selected, from the French version of Wikipedia, a set of 250 sentences containing polysemous words (restricted to common nouns) that are going to be used as target words to be disambiguated. The number of different target nouns was 48 (there was 5.20 sentence per target word has a mean). Not all meanings for each target word were represented, but we ensure that at least two meanings were proposed for each of them. We then asked to French native speakers to be volunteers for enumerating typed clues that seemed relevant for guessing the proper usage of the polysemous target words.

For example, the word *mine* in French has, amongst others, the meanings of: *appearance*, *explosive device*, *mineral exploitation (coalmine, gold mine)*, and the *graphite part of a pencil*. In the following sentences, they have been asked to select the proper meaning above all to produce clues (as given in below). The figures correspond to the number of time this clues has been given by volunteers.

(1) La première **mine** antipersonnel, hautement explosive et dotée d'un détonateur mécanique moderne fut employée par les troupes confédérés. (Eng. The first antipersonnel mine, highly explosive and provided with a mechanical detonator was used by confederate troops.)

Target term : *mine* > *charge explosive (mine as explosive)*

:carac antipersonnel (4) bataille (2)	:carac explosive (4) :patientof employer	:part détonateur (3) :part détonateur mécanique
--	---	--

(2) Une **mine** est un gisement exploité de matériaux. (Eng. A mine is a field exploited for materials.)

Target term : mine > gisement (mine as field)

:isa gisement (3)	:carac exploité (2) exploité	:locof matériaux (3) matériaux(2)
-------------------	---------------------------------	--------------------------------------

(3) On trouve la trace dès la très haute antiquité de l'exploitation des **mines** d'argent du Laurion. (Eng. We find evidence since antiquity of exploitation of the Laurion silver mines.)

Target term : mine > gisement (mine as field)

exploitation (5)	argent (3) :locof argent	Laurion (2) :loc Laurion
------------------	-----------------------------	-----------------------------

(4) La **mine** noire est réalisée à partir d'un mélange de graphite en poudre combiné à un mélange de kaolin et de bentonite. (Eng. The black mine is realized from a mixture of graphite powder combined to a mixture of kaolin and bentonite.)

Target term : mine > dessiner (mine as drawing/pencil)

:carac noire (3) noire (2) noir (2)	graphite (4) :subst graphite	kaolin (3) bentonite
--	---------------------------------	-------------------------

First, some few remarks are worthy. The annotators were free to choose their clues (and the type if any) but only from the words present in the sentences. They were not asked the type of the clues to follow any syntactic/semantic constraints present in the sentence. This last point could be discussed, but this constraint was felt as too complicated to the majority of volunteers. The clues could be given in the form occurring in the text or in a lemmatized version (like *noire/noir*). Terms of clues could be given several time with different types. Multiword terms could be used as long they are present in the text and known to the system, that is to say, existing in the lexical network).

Prior to the experiment, we made a large number of people to plays with AKI in taboo mode for the target words. Those players were not those who volunteering for producing clues, and they were not aware of the global experiment nor the sentences we would be using as test corpus.

The evaluation experiment was conducted has follow. For a given target word, the *initial lexical signature* was composed of its word senses (usages) with an equal weight equal to 1. The learning mechanism (adding new relations to the network) was disabled. All clues were given at the same time, reading the proposal made by the system only after.

The obtained figures are the following when considering all clues (typed or not) :

	Random		Belly only no weight		Belly + tail no weight		Belly only		Belly + tail	
	count	%	count	%	count	%	count	%	count	%
OK	69	27,6	158	63	176	70	195	78	245	98
NOK	181	72,4	92	37	74	30	29,6	12	5	2

The *OK* line refers to when the system has found the proper meaning/usage, and the *NOK* when the system proposed any other inadequate usage. The *Random* column refers to a totally random choice amongst senses. Columns with the mention *belly only* refers to when only relations concerning the target terms and belonging to the belly are considered. In that case, we ignore all relations of the tail. Column with *no weight* means that weights are ignored (they are all equal to 1). The mention of *belly + tail* means that all relations in the lexical network are taken into account.

We made also a comparison of the performances with ignoring the type of the clue. For example, the set of clues of sentence (1) given above is reduced to :

antipersonnel (4) bataille (2)	explosive (4) employer	détonateur (3) détonateur mécanique
-----------------------------------	---------------------------	--

The obtained figures are the following when clue types are ignored :

	belly only no type		Belly + tail no type	
	count	%	count	%
OK	165	66	223	89,2
NOK	85	34	27	10,8

As we said earlier, this experiment doesn't mean to prove anything as a new WSD approach but rather to assess the impact of the contents of the lexical network with a very simple approach. The experiment, although slightly reminiscent of (Véronis and Ide, 1990), is by itself far too limited (a very small set of terms and sentences and only limited to nouns) to pretend to have any insight in general large scope WSD. Nevertheless, the obtained results seem to show that relations belonging to the tail have a positive effect in guessing what could be the proper meaning in the context of a sentence. Moreover, the explicit use of strength (weights) for relations does improve the overall performance. Ignoring types for clues does reduce performance but to a less extend than ignoring weights. This can be explained by the large proportion of specific relations that are also existing as associated ideas (the basic relations without particular type in JDM).

A large scale experiment would be desirable, especially including verbs. A fully automatic handling of the process, that is to say not asking people to produce the clues, is also certainly a way to go, but at this stage the lack of a French analyzer able to produce the proper typed clues remains an obstacle. In any case, asking people to produce clues for WSD is by itself interesting for assessing the relative usefulness of the various relation types. Annotating this way a large collection of sentences may be worth the effort.

Conclusion

The lexical network (JDM) created under the JeuxDeMots project is large scaled and has a wide coverage. From this network, we have conceived a prototype that can be viewed both as a game and as a Tip of the Tongue (TOT) tool, and whose purpose is to increase the number of low weight relations, thus making the JDM lexical network *long tailed*. We have in this paper considered the long tail property as a global property of the edge weights, and not the frequency distribution of terms nor the distribution of relation number linking terms. Globally, for a given term the cumulated weight of the first 20% stronger relations is equivalent to the 80% remaining. Depending on terms and of their lexical richness and usage, the long tail can start in a range from 10% to 30% of the cumulated relation weight. Under the game process with intersection by pairs, the construction of dense long tail can be slow (in an inverse quadratic way), because they are not “frontal” ones and users do not spontaneously think to them. We saw in this paper how a TOT game, used in a taboo mode, can help create such “indirect” relations in a more efficient way, while retaining the principle of typed and weighted relations. Beside presenting the approach for increasing the long tail, the second objective of this paper was to try to assess if this work had any usefulness. We evaluated the impact of the long tail in two different contexts. First, it does help the retrieval process of a TOT software as evaluated in a quite large number of occurrences (more than 15000 plays) over more than a year time span. Secondly, the long tailed of typed and weighed relations seems to have a positive effect on a WSD task.

References

- von Ahn L., Dabbish L. (2004). *Labelling Images with a Computer Game*. ACM Conference on Human Factors in Computing Systems (CHI). pp. 319-326.
- von Ahn L., Dabbish L. (2008) *Designing Games With a Purpose*. Communication of the ACM, 51(8):58–67 August 2008.
- Anderson, Ch. (2004) *The Long Tail*. Wired 12:10, October 2004.
- Boyd-Graber J., Fellbaum C., Osherson D., and Schapire R. (2006) *Adding Dense, Weighted Connections to WordNet*. In *Proceedings of the Thirds International WordNet Conference*. Masaryk University Brno, 2006, 9 p.
- Budanitsky A., Hirst G. (2001) *Semantic distance in WordNet: an Experimental, Application-oriented Evaluation of Five Measures*. Proceedings Workshop WordNet and Other Lexical Resources. The North American Chapter of the Association for Computation Linguistics (NAACL), Pittsburgh, PA, 2001.
- Chamberlain, J., Poesio, M., and Kruschwitz, U. (2008). *Phrase Detectives: A Web-based Collaborative Annotation Game*. In Proceedings of the International Conference on Semantic Systems (I-Semantics'08), Graz.

- Chklovski, T. and Gil, Y. (2005) *Improving the design of intelligent acquisition interfaces for collecting world knowledge from web contributors*. In Proceedings of K-CAP '05, pages 35–42.
- Collins A, Quillian M.R. (1969). *Retrieval time from semantic memory*. Journal of verbal learning and verbal behaviour, 8(2), pp. 240-248.
- Fellbaum C. (ed.): *WordNet, an Electronic Lexical Database*, MIT Press, (1998)
- Gaume, B. (2008) *Mapping the form of meaning in Small Worlds*. in Journal of Intelligent Systems, 15 p.
- Lafourcade M., (2007) *Making people play for Lexical Acquisition*. In Proc. SNLP 2007, 7th Symposium on Natural Language Processing. Pattaya, Thailande, 13-15 December 2007, 8 p.
- Lafourcade M., Joubert A. (2010). *Computing trees of named word usages from a crowdsourced lexical network*. *Investigationes Linguisticae*, volume XXI, pp. 39-56.
- Lenat D. (1995) *CYC: A large-scale investment in knowledge infrastructure*. *Communications of the ACM*, 38(11):33–38, 1995.
- Lieberman H., Smith D.A., Teeters A. (2007). *Common Consensus: a web-based game for collecting commonsense goals*. International Conference on Intelligent User Interfaces (IUI'07). Hawaiï, USA.
- 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 2003).
- Mihalcea R. and Radev D. (2011), *Graph-based Natural Language Processing and Information Retrieval*, Cambridge University Press, 2011.
- 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*, 3 (4), pp. 235-244.
- Morris J., Hirst G. (2004) *Non-classical lexical semantic relations* Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics, 46-51, 2004.
- Navigli R. (2009) *Word Sense Disambiguation: a Survey*. *ACM Computing Surveys*, 41(2), ACM Press, 2009, pp. 1-69.
- Navigli R., 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, pp. 216-225.
- Polguère A. (2006). *Structural properties of Lexical Systems : Monolingual and Multilingual Perspectives*. Proceedings of the Workshop on Multilingual Language Resources and Interoperability (Coling/ACL), Sydney, pp. 50-59.
- Sagot B. et Fiser D. (2008) *Construction d'un WordNet libre du français à partir de ressources multilingues*. Dans les actes de TALN 2008, Avignon, France, 2008.

- Sigman M, Cecchi GA. (2002) *Global organization of the WordNet lexicon*. Proc Natl Acad Sci USA. 2002;99:1742–1747.
- Véronis J., & Ide N. (1990). *Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries*. In Proceedings of 13th International Conference on Computational Linguistics (COLING'90), vol. 2, pp. 389-394. Helsinki.
- Lafourcade, M., Zampa, V. (2009) *PtiClic : a game for vocabulary assessment combining JeuxDeMots and LSA*. In proc of CICLing (Conference on Intelligent text processing and Computational Linguistics). Mexico : Marsh 1-7 , 10 p.
- Zesch, T.& I. Gurevych, I. (2009) *Wisdom of crowds versus wisdom of linguists measuring the semantic relatedness of words*. Natural Language Engineering, Cambridge University Press, pp 25–59, 2009.
- Zipf G K. (1965) *The Psycho-Biology of Language*. Cambridge, MA: MIT Press; 1965.
- Zock M., Ferret O. and Schwab D. (2010), *Deliberate word access : an intuition, a roadmap and some preliminary empirical results*, in International Journal of Speech Technology, Volume 13, Number 4, December 2010, Springer Verlag.