TOTAKI: a help for lexical access on the TOT problem

Mathieu Lafourcade, Alain Joubert

LIRMM, Université Montpellier II
{joubert,lafourcade}@lirmm.fr

Abstract

The JDM lexical network has been built thanks to on-line games the main of which, JeuxDeMots (JDM), was launched in 2007. It is currently a large lexical network, in constant evolution, containing more than 310 000 terms connected by more than 6.5 million relations. The riddle game Totaki (Tip Of the Tongue with Automated Knowledge Inferences), the initial version of which was elaborated with Michael Zock, was launched in a first version in 2010. The initial aim of this project is to cross validate the JDM lexical network. Totaki uses this lexical network to make proposals from user given clues, and in case of failure players can supply new information, hence enriching the network. Endogenous processes of inference, by deduction, induction, abduction, also allow to find new information not directly available in the network and hence lead to a densification of the network. The assumption about the validation is that if Totaki is able to guess proper terms from user clues, then the lexical network contains appropriate relations between words. Currently, Totaki achieves a 75% success rate, to be compared to less than 50% if the guessing is done by human users. One serious application of Totaki is to be viewed as a tool for lexical access and a possible remedy for the tip of the tongue problem.

The Wikipedia encyclopaedia, built in a collaborative way, represents a very important volume of knowledge (about 1.5 million articles in its French version). The idea developed in this chapter consists in benefiting from Wikipedia to enrich the JDM network and evaluate the impact on Totaki performance. Instead of relying only on the JDM network, Totaki also makes use of information extracted from Wikipedia. The overall process is then both endogenous and exogenous.

In a first part, we shall remind the reader the basic principles of a lexical network, then the aims and the underlying principles of the Totaki game. We shall see on examples Totaki may be used as a game to evaluate and enrich the JDM network, but also it may be considered as a tool for the Tip Of the Tongue problem; partial syntactic or morphologic information may be added to semantic information to help the user. In a second part, we shall show the results of the evaluation of the JDM network, results we obtained playing Totaki. We shall clarify the process allowing the introduction in the Totaki game of data extracted from Wikipedia as a complement in the information from the JDM network, and we shall briefly present the results provided by the first experiments.

1. Introduction

The JDM lexico-semantic network has been built thanks to on-line games (Games With A Purpose or GWAPs) the main of which, JeuxDeMots (JDM), was launched in 2007 (Lafourcade, 2007). It is a large lexical network, in constant evolution, containing currently more than 310 000 terms connected by more than 6.5 million relations. The riddle game Totaki (Tip Of the Tongue with Automated Knowledge Inferences), the initial version of which was elaborated with Michael Zock, was launched in its first version in 2010 (Lafourcade et al., 2011). The initial purpose of this project was to cross validate the JDM lexical network. In particular, we wanted to answer the question: *"Is the lexical network complete enough with terms and relations (between terms) it contains?"*. With *"enough"*, we mean data into the network are sufficient in number and quality to allow in a satisfying way the realization of classical tasks in NLP, such as textual semantic analysis with lexical disambiguation. It is obvious that, in an exhaustive way, such a network can never be complete, only because of the permanent evolution of the language and the linguistic data[1]. Totaki uses the data of the network to allow the system to develop its proposals from clues supplied by the players, but in case of failure of the system the players can supply new information, thus enriching the network. Recently, endogenous processes, working by

[1] In particular, new terms (e.g.: *obamania* or *to vapote*) or new meanings of already existing terms (e.g.: *tablet*) regularly arise.

deduction, induction and abduction (Zarrouk et al., 2013), also allowed a densification of the network: approximately 1 million new relations were so inferred.

In a first part, we shall remind the reader the basic principles of a lexical network and these of Totaki game. We shall show that Totaki, initially designed to estimate the JDM network, can be an interesting solution to the Tip Of the Tongue (TOT) problem, but also allows an enrichment of this network, in particular of the long tail of its relations. The second part of this chapter will be dedicated to the results of the evaluation of the JDM network, obtained thanks to Totaki. Then, we shall clarify a new process (still unpublished) allowing the introduction in the Totaki game of data extracted from Wikipedia as a complement of the information from the JDM network. This process will be illustrated by an example, before presenting the first results of this experiment.

2. Lexical Networks and Totaki

2.1. Structure of a Lexical Network

The structure of a lexical network, like the one we are building, is composed of nodes and links between nodes, as it was initially introduced in the end of 1960s by (Collins and Quillian, 1969) and more recently clarified by (Polguere, 2006). A node (a vertex) of the network refers to a term (or a multiple word expression or any textual segment), usually in its canonical form (lemma). The links between nodes are typed and are interpreted as a possible relation holding between the two terms. Some of these relations correspond to lexical functions, some of which have been made explicit by (Mel'čuk et al., 1995), others are semantically motivated like hypernym, hyponym, agent, patient …

More formally, a lexical network is a graph structure composed of nodes (vertices) and links.

- A node is a 3-tuple : <label, type, weight>
- A link is a 4-tuple <start-node, type, end-node, weight>

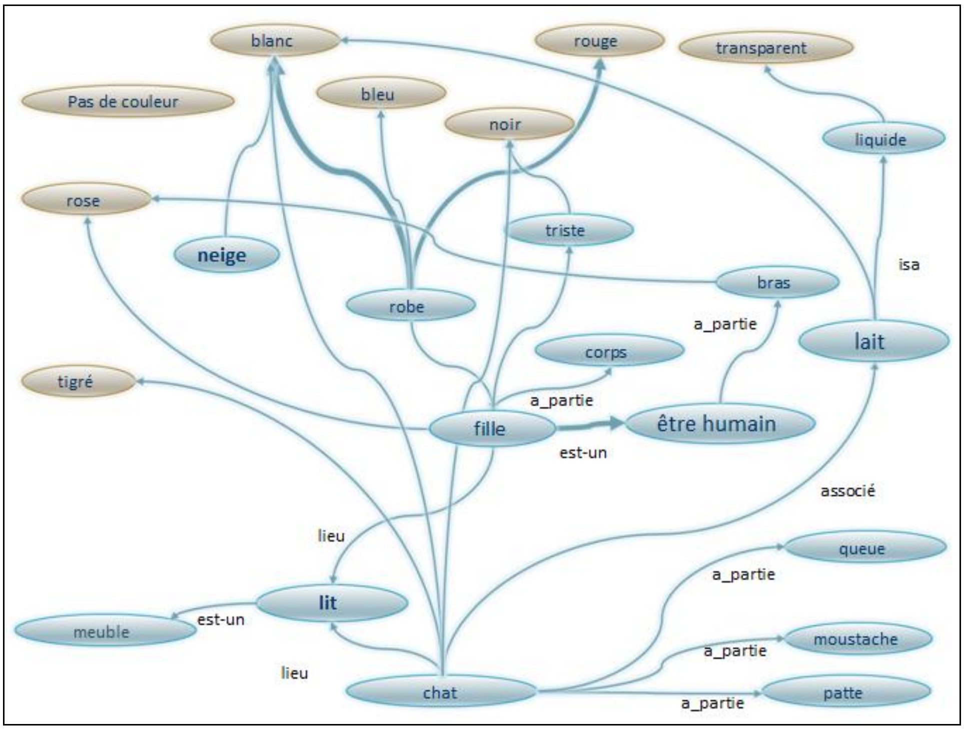

Figure 1: An example (partial) of a lexical network. For sake of clarity, the relation weights are not represented here. Only nodes corresponding to terms are displayed.

The label is simply the string holding the term. The type is an encoding referring to the information holding by the node. For instance a node can be a term or a Part of Speech (POS) like :Noun, :Verb. The link type refers to the relation considered, for instance: is_a, synonym, part_of. A node weight refers to the number of times a term has been used by players. Similarly, the weight of a relation refers to the strength of the relation. Figure 1 shows a very small example of the kind of lexical network we are dealing with.

2.2. Presentation of the JDM network

The interest in and the feasibility of on-line games for acquisition of high quality lexical resources have clearly been established by (Lafourcade and Joubert, 2013, published by Gala and Zock). The JDM network, which is in constant evolution, has been built by means of several games:

- JeuxDeMots: it is the very first game of the project, launched in July 2007, and its purpose is the constitution of the JDM network, from an already existing base of 150 000 terms. In a JDM game, two players, anonymously and in an asynchronous way, propose typed associations for a term randomly picked up in the base. JDM allowed to acquire approximately 2 million relations[2], as well as to increase the base to more than 310 000 terms.
- PtiClic: this game was launched in 2008. Contrary to JDM, it is a “closed” game in which the player has to associate terms which are displayed on the screen. These terms come from the JDM network, but also from a voluminous corpus analyzed with LSA.
- Games with choice such as AskIt or LikeIt: these games were most recently launched, from 2010. They suggest the user he answers “true” or “false”, “like” or “unlike”, on simple statements about lexical data. They are very fast games, the user generally can give his answer within a few seconds. These games allow to obtain polarities about the terms of the network.
- Another game, called Tierxical, ask the player to sort (and bet) three proposals among a set of nine proposals which are associated to a term and a directive.

Diko is a tool allowing to display of the information contained in the network. For a target term, the screen presents in a clear way relations it is concerned with. In the example reproduced in figures 2 and 3, the target term corresponds to a well known researcher (unfortunately not very well lexicalized in the network).

Figure 2: Diko screen for the term *Michael Zock.*

[2] Here, it is about relations acquired thanks to JDM game. The main part of the other relations present in the network was acquired by deduction, induction, abduction processes (about 1 million relations) or using data from Wikipedia (about 3.5 million relations).

Figure 3: Diko screen for the term *Mickael Zock*. The reader may notice that players have some difficulties spelling correctly *Michael Zock*'s first name, because when his first name is spelt with a *k* (what is not the correct spelling) the term is strongly lexicalized, almost as well as when its first name is correctly written (with a *h*).

Diko, besides being a visualisation tool for the JDM network, also allows contributions from interested and experimented players. Indeed, some relations turn out to be poorly playable: they are relations either too specific (e.g.: *magn* indicating an intensification of the target term[3]), or scarcely lexicalized, that is for which there are only very few possible answers (e.g.: *instrument* > *action*[4]). A significant number of players, mainly among those who spent a lot of hours playing JDM, wished to become contributors to inform specific terms or more difficult relations, and thus weakly lexicalized (e.g.: the *parasitology* domain was mainly lexicalized this way): it allows the players, so become contributors, to take part on domains which interest them more particularly or on which they have specific knowledge. Currently, approximately 1 million contributed relations are awaiting validation. These validations are manually realized by expert validators: when several contributors proposed the same relation, this one is posted first and foremost for the expert validator; at the moment, there is no automatic validation. A process of automatic validation is currently been studied; it leans on the notion of minimal vote: when a number of contributors will have proposed the same relation, with a very strong proportion of corresponding votes, then this relation could be automatically validated, with a type indicating that it results from contributions and with a relatively low weight.

2.3. Validation of the JDM network

The basic principles of the design of Totaki, developed in association with Michael Zock, were presented by (Lafourcade and al., 2011). The initial aim was to obtain a qualitative evaluation of the JDM network, also thanks to an on-line game. The motivation to obtain this evaluation from a game relies on the idea that the number of participants would be much more important with a game than with a classical approach based on the simple voluntary service. With the aim of such an evaluation, the question of the comprehensiveness of our network settles in a more practical way: *"for a given term, are the typed and weighted relations it possesses with the other terms of the network sufficient to determine it in a unique way?"*. If the answer is positive, any term can be found by means of a reduced number of typed terms (which are clues).

---

[3] e.g. : *magn (fever) = high fever*
[4] e.g. : *scissors* → *cut*

The principle of the Totaki game consists in making guess a word to the system by proposing it clues. Following each clue given by the player, the system proposes a possible answer term. The clues can be typed (e.g.: *is_a* for hyperonymy or *syn* for the synonymy), otherwise we consider them as simply associated to the word to be found. The type of a clue corresponds to one of the types of relations existing in the network (e.g.: *is_a animal*). The answers proposed by the system result from typed and weighted relations stored in the JDM network: Totaki proposes the term which is the most strongly connected to the clues given by the player and not already previously given neither by the player nor by Totaki itself in the same game. The process Totaki relies on was exposed in (Joubert and Lafourcade, 2012). After each clue supplied by the user, the system computes the intersection of all the terms associated to the previously given clues with all the terms associated to this last clue. If no term may be proposed by the system (in case of an empty intersection), then the system computes the union of these sets: naturally, in it fall-back position, the precision is strikingly less good.

If, after several clues, Totaki does not find the term which tried to make it guess the player, Totaki admits defeat. The player supplies the target term he thought, so realizing an enrichment of the network: the relations *clue → target term* are added in the JDM network, typed “Totaki” in order to make the distinction with the relations resulting from other games, but also the target term is added if this one was until then unknown by the network.

In its initial version, Totaki was based only on semantic data contained in the JDM network, that is on the relations between terms, their types and their weights. A more recent version introduces the exploitation of morphological and phonetic notions, by allowing the user to specify:

- the length of the target term (e.g. *:long = 4* in order to select only terms of 4 characters long)
- elements of spelling, even phonetics (e.g. *:reg_tion* for terms containing 1 character followed by the characters *tion*, or *:reg%tion* for terms of any number of characters followed by characters *tion*).

Figures 4, 5 and 6 show an example of a game partially played using these last possibilities. Figure 7 shows other examples of clues allowing to obtain the same target term.

Vos indices	Mes réponses
TALN	grammaire
:reg Mi%	Mickael Zock

Figure 4: example of the beginning of a game: part of the screen showing the clues supplied by the user, as well as the answers Totaki made; the user looks for a term connected to the clue *NLP* (Totaki proposed the term *grammar*) and beginning with the characters *Mi*: guess which researcher was proposed by Totaki!

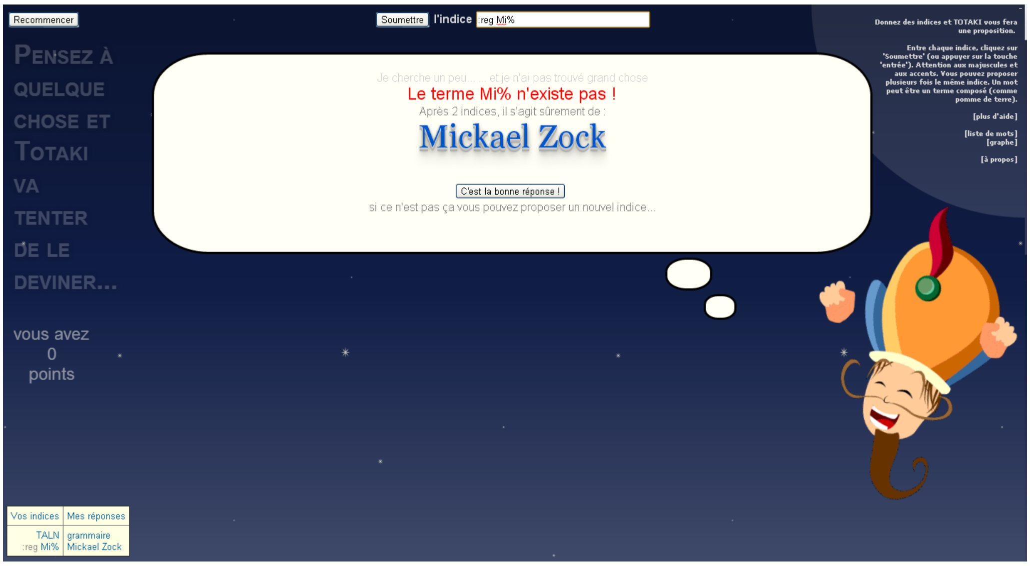

Figure 5: example of the beginning of a game in which the user asked terms connected with *NLP* and beginning with the characters *Mi*. Look at Totaki answer!

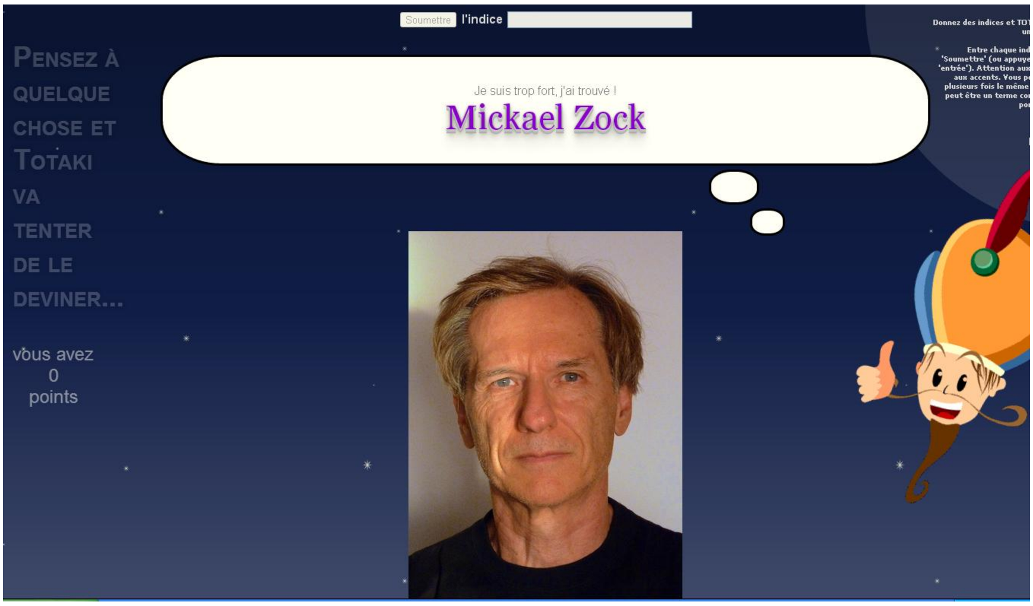

Figure 6: Screen obtained after the user validates Totaki proposal.

Vos indices	Mes réponses
TALN	grammaire
Marseille	Mickael Zock

Vos indices	Mes réponses
LIF	Mickael Zock

Vos indices	Mes réponses
Marseille	bouillabaisse
Taiwan	lieu
Grenoble	Michael Zock

Vos indices	Mes réponses
sympathique	gentil
faire de la recherche	Michael Zock

Figure 7: Several examples of set of clues allowing Totaki to find *Michael Zock* (not always correctly spelt!). The last example (with *nice* and *doing research* as clues) is, in our mind, the most characteristic one to find Michael.

Another semantic learning system, Never-Ending Language Learning (NELL), developed by Tom Mitchell's team (Carlson et al., 2010), regularly parse web pages looking for semantic relationships between information it already knows and what it finds through its search process; thus, it makes new connections in a manner that is intended to mimic the way humans learn new information. NELL, such as IBM's Watson (Ferrucci et al., 2010), aims at to be able to develop means of answering questions posed by users in natural language with no human intervention in the process.

2.4. Help for the TOT problem

Totaki can also be used as a help for the TOT problem, as clearly analyzed by Michael Zock in (Lafourcade and al., 2011). Indeed, in the case of a real TOT problem, the user does not find spontaneously a term which nevertheless he knows very well[5]. He knows many elements about this term, but he does not manage to have access on it. The user generally knows semantic information about this term (its context, as well as its lexical field), morphological information (its lexical category: name, verb, adjective …, his genre: male, feminine), phonological information (intonative outline, approximate number of syllables). He is then going to supply a series of targeted clues, possibly semantically typed. One of the real examples of the TOT problem is reproduced in figure 8: the user was not finding the term *saltcellar*, he supplied as first clue the term *salt*, the proposal made by Totaki was *sugar*, because *sugar* was the most strongly connected term with *salt*, then the user supplied the clue *bowl*, the proposal of Totaki was then *salt cellar*, which was the term the user looked for.

Vos indices	Mes réponses
sel	mer
récipient	salière

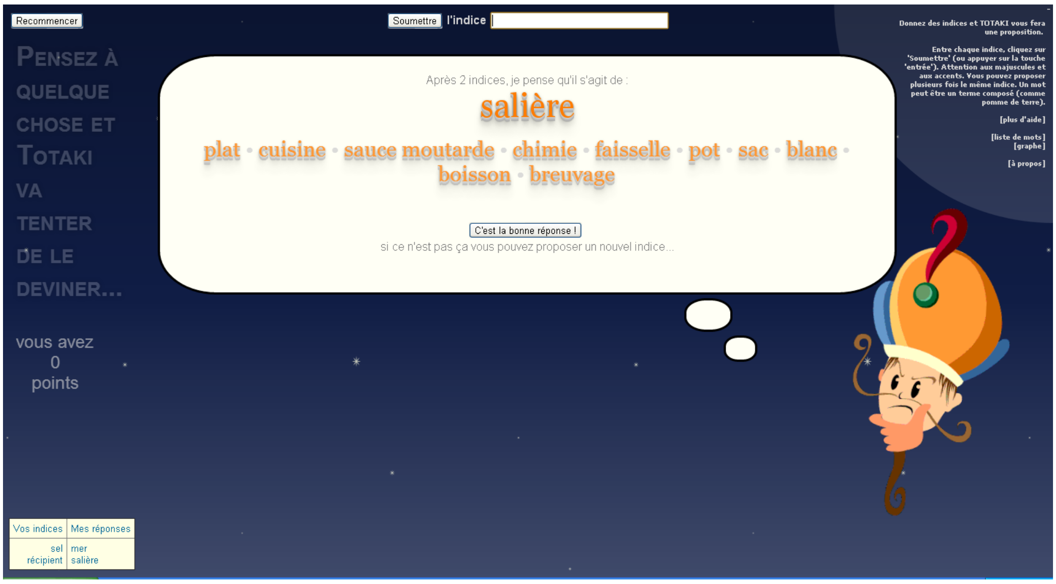

Figure 8: extract from the screen, then the complete screen of an example of real case where Totaki allowed a user to find the term that he had on the tip of the tongue, namely *saltcellar* from the clues *salt* and *bowl*. It is to notice that when we did this example again the first proposal of Totaki was not the term *sugar*, but the term *sea*, what clearly shows the evolution of our network: between the real game of the user and the reproduction which we made for this paper (a few weeks), the term the most strongly connected with *salt* was not any more *sugar*, but *sea*.

---

[5] The TOT problem has been studied by a lot of authors. One of the most recent analysis (Zock and Schwab, 2013) also supplies very promising elements of answer.

A second example of Totaki used as tool for a TOT problem is reproduced in figure 9. The user not finding any more the term *washbasin* supplied the clues *latin* (remembering that there is a joke in Latin with the target term), then the clues *white*, *hard* and *bathroom*.

Vos indices	Mes réponses
latin	langue
blanc	noir
dur	plastique
salle de bain	lavabo

Figure 9: extract from the screen of another example of help for a TOT problem in which Totaki found the target term.

The first version of Totaki allowed the exploitation of the only semantic information at the disposal of a user looking for a term. Even if the latter possessed morphological or phonological information, these could not be taken into account. As seen previously, the current version of Totaki allows a partial exploitation of this information.

2.5. Enrichment of the JDM network thanks to Totaki

When a player discovers Totaki, most of his games are played to verify the scale of the knowledge of the network, beginning generally with common terms with frontal indications (e.g.: *feline*, *mouse* to make guess *cat*). Very quickly, the players try to test the limits of the system and are going to propose thus either games on common terms but with side clues (e.g.: *booted*, *angora* for *cat*), or games on infrequent terms generally with frontal clues (e.g.: *cat*, *tale* for *Puss in Boots*). In both cases, the players try "to trap" the system: there is learning of new relations (in the first case), but also of new terms often connected to the current events (in the second case). For example, the set of the names of Pokemons (*Pikachu*, *Dracaufeu …*) was mainly constituted thanks to Totaki. So, Totaki allows enrichment on terms directly chosen by the users.

However, these new relations or these new terms introduced into the JDM network thanks to Totaki are initially supplied only by a single player. It would be possible that, faithfully or not, the player introduces erroneous information. It is the reason why the relations introduced by Totaki into the network are differently typed ("Aki" type) and weighted with a low value (equal to 10, while the weight of a relation introduced by a single couple of players into JDM values 50).

2.6. Enrichment of the long tail

Due to the concept of the JDM game, most of the relations in the network are “direct” or “frontal” (e.g.: *dog* → *animal*, of “associated idea” type and the weight of which is currently more than 820); they are the ones which were the most spontaneously given by the users. However, certain relations are “indirect” or “lateral” (e.g.: *dog → sit down, don't move, slaver, bring back* … also of “associated idea” type and all of them with a current weight less or equal to 60). These last relations constitute the *long tail*. For most of the terms present in the network, the major part of the relations, in number, is in the long tail. Currently, the distribution of the relations for a given term follows a power law (more exactly, a Zipf law): so, the cumulated weight of the 80 to 85 % weaker relations is similar to the weight of the 15 to 20 % stronger relations. In a classic Totaki game, the user can supply the clues he wishes; during a Totaki game in taboo mode, we are going to make guess to the system a target term, forbidding the user to supply as clues ten terms the most strongly connected with this target term in the network, that is forbidding the strongest "frontal" indications. For the example *dog*, the ten forbidden terms in taboo mode are: *cat*, *animal*, *Snowy*, *Idefix*, *to bark*, *poodle*, *niche*, *wolf*, *bone* and *she-dog*. The user is thus obliged to supply clues less strongly connected with the target term and thus belonging to the long tail. The principle reminds the one of the Taboo board game. This process, presented by (Lafourcade and Joubert, 2012) inevitably increases the recall.

3. Evaluation of the JDM network …

3.1. … thanks to Totaki

The initial idea of the design of Totaki is that the success rate to find a target term from clues is covariant to the rate of comprehensiveness of the JDM network: the more our network will be complete, the more the success rate of Totaki will be important. Totaki so allows us to have an empirical evaluation of the network JDM.

The evaluation, just like the learning, is made only according to what the players informed. As already mentioned, Totaki can be envisaged as a game or a tool for the TOT problem. A priori, our software does not know how to make the distinction between both uses. Indeed, after a single game and if AKI finds the solution, we cannot know a priori if the user knew or if he looked for the target term. On the other hand, if in a relatively short lapse of time (within a few minutes) the same term is played several times, we can make the hypothesis that it is about a use of game (at least from the second game) where the user tries to make find the target term by Totaki, generally proposing more and more side different clues. In both cases, game or TOT tool, the target terms are mainly terms of average or low frequency. Indeed, play to find a frequent word does not present a big interest, and generally we do not look for a common term thanks to a TOT tool. Figure 10 shows the evolution in time of the ratio between the number of games Totaki won (games the user indicated that the software found the target term) and the number of played games. The result we obtained seems to stabilize towards a value of the order of 75%. It should be noted that this value is probably only an underestimate of the comprehensiveness of the JDM network (on the played vocabulary), because, as we have already indicated it, a large number players look for the limits of Totaki.

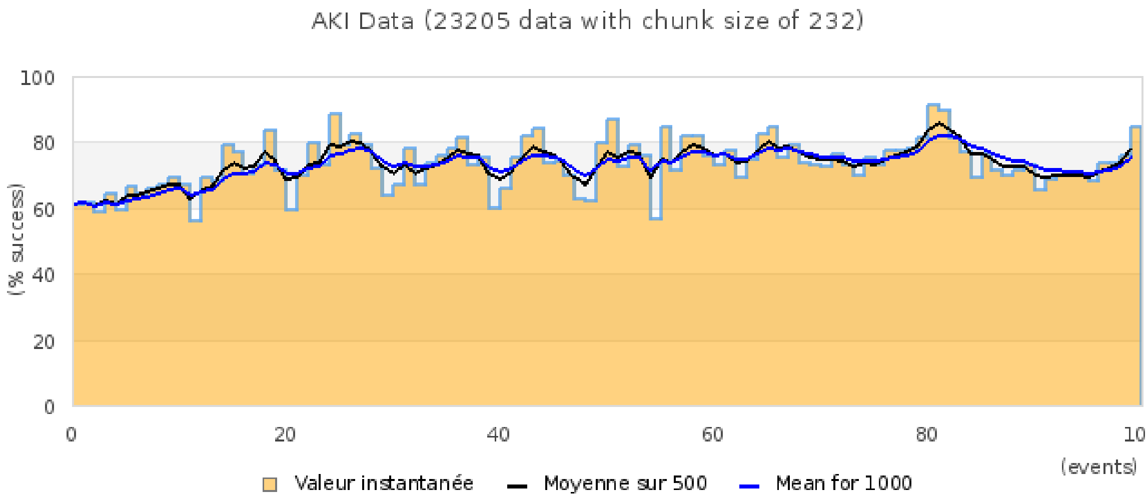

Figure 10: evolution in time of the success rate of Totaki. We notice a light progress of the averages since 60% to values which can exceed 85%, but generally in the order of 70 to 80%: the network is thus complete in approximately 75%! It would seem that the light reduction in the success rate in last thousands of played games results from the fact that a largest number of players learnt new terms to Totaki. Histogram (immediate value) is drawn with a chunk size of 1% of the number of played games; we also give mean values for 500 and for 1000.

3.2. … and Wikipédia

3.2.1. Principle

The Wikipedia encyclopedia, constructed in a collaborative way, represents a very important volume of knowledge (about 1.5 million articles in its French version). The idea we develop in this section consists in benefiting from Wikipedia to enrich the JDM network. Instead of leaning only on the data of the JDM network, Totaki will also use information extracted from Wikipedia. Thus, it is here about an endogenous and exogenous process.

During a game of Totaki, every clue the player supplies serves as entry to Wikipedia. The software selects, on the Wikipedia page of this entry, the different links, except the links of numerical type[6]. In the JDM network, all the relations *clue* → *link* are then added, with the "Wiki" type and a low valuable weight 10, unless these relations already existed (whatever is their type). If the term corresponding to this link does not exist in the JDM network, it is added to it, in particular to be able to add the correspondent "Wiki" relation *clue* → *link*. Naturally, for the target term of this game of Totaki, the similar relations (between this target term and the corresponding links on the Wikipedia page) are also added to the JDM network. When the player confirms the proposal from Totaki, if Totaki had found this proposal thanks to "Wiki" relations, these relations are then added as "AKI" relations, also with a weight of 10.

In order to not slowing down the Totaki games, it was decided to make a scan on all the terms of the JDM network and thus to create all the "Wiki" relations *clue* → *link*. So, when a player proposes a clue, the system will not need to scan the corresponding Wikipedia page.

Wikipedia contains much more information than the JDM network; figures 11 and 12 clearly show this difference on an example: while the Wikipedia page relative to *Gérard Philippe* draws up a relatively exhaustive list of its movies, the corresponding Diko screen only gives a few of them. On the other hand, a number of Wikipedia links, relative to general terms, corresponds to noise (that is the reason why we chose a low value weight of 10 for "Wiki" relations). Figure 13, extracted from the Wikipedia page of *Gérard Philippe*, contains links which are absolutely not characteristic for him, such as the terms *war* or *baccalaureate*.

Filmographie [modifier | modifier le code]

- 1944 : *Les Petites du quai aux fleurs* de Marc Allégret : *Jérôme Hardy*
- 1945 : *Le Pays sans étoiles* de Georges Lacombe : *Simon Le Gouge et Frédéric Talacayud, clerc de notaire*
- 1945 : *La Boîte aux rêves* d'Yves Allégret et Jean Choux : *une silhouette dans le film*
- 1945 : *Schéma d'une identification*, court métrage inédit d'Alain Resnais : *Le viveur en smoking*
- 1946 : *L'Idiot* de Georges Lampin : *Le prince Muychkine, naïf d'esprit*
- 1946 : *Ouvert pour cause d'inventaire*, court métrage inédit d'Alain Resnais (semble perdu)
- 1947 : *Le Diable au corps* de Claude Autant-Lara : *François Jaubert*
- 1947 : *La Chartreuse de Parme* de Christian-Jaque : *Le marquis Fabrice Del Dongo*
- 1947 : *Les Drames du bois de Boulogne*, court métrage de Jacques Loew : *commentaire*
- 1948 : *Une si jolie petite plage* d'Yves Allégret : *Pierre Monet, le voyageur*
- 1948 : *Tous les chemins mènent à Rome* de Jean Boyer : *Gabriel Pégase, géomètre*
- 1949 : *La Beauté du diable* de René Clair : *Henri et Faust, jeune*
- 1949 : *Visite à Picasso*, court métrage documentaire de Paul Haesaerts : *récitant du film*
- 1950 : *Souvenirs perdus* de Christian-Jaque : *Gérard de Narcay, le fou évadé de l'asile*
- 1950 : *Juliette ou la Clé des songes* de Marcel Carné : *Michel, le rêveur qui*
- 1951 : *Les Sept Péchés capitaux*, sketch *Le Huitième péché* de Georges Lacombe : *Le bonimenteur et le peintre*
- 1951 : *Fêtes galantes : Le peintre Watteau*, court métrage, documentaire de Jean Aurel : *commentaire*
- 1952 : *Les Belles de nuit* de René Clair : *Claude, obscur compositeur de musique*
- 1952 : *Les Amants de la villa Borghèse*, sketch *La Rupture* de Gianni Franciolini : *Le beau jeune homme sur le banc*
- 1953 : *Les Orgueilleux* d'Yves Allégret : *Georges, ancien médecin alcoolique*
- 1953 : *Monsieur Ripois* de René Clément : *Mr André Ripois, mari de Catherine*
- 1953 : *Si Versailles m'était conté...* de Sacha Guitry : *D'Artagnan*
- 1954 : *Le Rouge et le Noir* de Claude Autant-Lara : *Julien Sorel*
- 1954 : *Forêt sacrée*, court métrage, documentaire de Pierre-Dominique Gaisseau : *commentaire*
- 1955 : *Sur les rivages le l'Ambre*, court métrage, documentaire de Jerzy Kalin : *commentaire*
- 1955 : *Les Grandes Manœuvres* de René Clair : *Armand de La Verne, lieutenant des dragons*
- 1955 : *Si Paris nous était conté* de Sacha Guitry : *Le chanteur des rues*
- 1955 : *La Meilleure Part* d'Yves Allégret : *Philippe Perrin, l'ingénieur sur le barrage*
- 1956 : *Les Aventures de Till l'Espiègle* de Gérard Philipe et Joris Ivens : *Till l'Espiègle*

Figure 11: part of the Wikipedia page relative to *Gérard Philippe*, showing a part of his filmography: the links are the colored terms (red when the Wikipedia page does not still exist).

[6] Years (e.g.: *1984*) are not taken into account; on the other hand, dates (e.g.: *September 11th*) are.

Figure 12: Diko screen about *Gérard Philippe*: only a few of his movies are listed.

Biographie [modifier | modifier le code]

Profession	Acteur
Films notables	*L'Idiot* *Le Diable au corps* *La Chartreuse de Parme* *La Beauté du diable* *Fanfan la Tulipe* *Les Orgueilleux* *Monsieur Ripois* *Le Rouge et le Noir* *Les Grandes Manœuvres* *Les Liaisons dangereuses*

Enfance et jeunesse [modifier | modifier le code]

Il voit le jour à Cannes (Alpes-Maritimes), dans une famille aisée, fils de Marcel Philip (1893-1973), avocat qui possède un cabinet de contentieux juridique, puis sera administrateur-gérant du Parc Palace Hotel de Grasse, et de Marie Villette. Son frère aîné se prénomme Jean. Le passé de Marcel Philip (ancien membre des ligues fascistes, admirateur de Doriot, hébergeur de l'état-major mussolinien dans le sud-est de la France en sa qualité d'administrateur-gérant du Parc Palace Hotel)[2], devient après la guerre le grand drame de la famille.

Il suit toute sa scolarité au lycée de l'Institut Stanislas de Cannes tenu par les marianistes où il est bon élève. Il y obtient, au début de la guerre, son baccalauréat.

Son père le destine à une carrière de juriste, mais, rencontrant de nombreux artistes réfugiés sur la Côte d'Azur (en zone libre) depuis 1939, il décide de devenir comédien. Sa mère le soutient dans ce choix. Il ajoute un « e » à son nom pour obtenir treize lettres avec son nom et son prénom, chiffre porte-bonheur selon celle-ci.

La guerre, les débuts d'acteur [modifier | modifier le code]

En 1941, le réalisateur Marc Allégret lui fait passer une audition, en compagnie de son amie Danièle Delorme, et l'envoie prendre les cours d'art dramatique de Jean Wall et Jean Huet à Cannes. Le comédien Claude Dauphin le fait jouer au théâtre à partir de 1942 avec *Une grande fille toute simple* d'André Roussin au casino de Nice.

En 1942, Marc Allégret lui fait jouer une silhouette dans son film *La Boîte aux rêves*, réalisé par son frère Yves. En novembre de la même année, la zone libre est occupée par l'armée allemande.

En 1943, la famille Philip s'installe rue de Paradis, dans le 10e arrondissement de Paris, où Gérard s'inscrit au Conservatoire national supérieur d'art dramatique, suit les cours de Denis d'Inès puis de Georges Le Roy et obtient le second prix de comédie. Il participe à la Libération de Paris en 1944 en faisant partie de la résistance française (FFI) alors que son père est un collaborateur notoire.

Il libérera [3] notamment l'Hôtel de Ville de Paris en août 1944 en compagnie de 30 personnes sous les ordres de Roger Stéphane.

Figure 13: part of the Wikipedia page relative to *Gérard Philippe*, showing a part of his bibliography; links such as *war* or *baccalaureate* are absolutely not characteristic for *Gérard Philippe*.

3.2.2. Evolutionary aspect

The JDM network, just like the encyclopaedia Wikipedia, is in a Never Ending Learning context. In order to take into account the evolutionary character of Wikipedia, the "Wiki" relations are created with a deadline date (currently arbitrarily chosen 1 month). When a "Wiki" relation is called, it is really used only if its deadline date is not overtaken.

With the aim of minimizing the "Wiki" importance of the relations *target_term* → *link*, in particular compared with the "AKI" relations, their weight is decreased of 1 for each Totaki game concerning this target term. When the weight of a "Wiki" relation reaches the 0 value, this relation is deleted from the JDM network. The next time this relation will be found thanks to Wikipedia, it will be created with a weight equal to 10. This process allows to take into account the evolutionary character of the Wikipedia articles.

In the section 3.2.1, we mentioned that a scan on all the terms of the JDM network, with scan of the correspondent Wikipedia pages, had been realized. To take into account the evolutionary character of the JDM network and of Wikipedia, such a process must be regularly made (currently every 30 days, according to the deadline period of the "Wiki" relations).

3.2.3. Interest for Totaki game

For the terms poorly lexicalized in the JDM network but suitably informed in Wikipedia, (for example named entities), this process allows to supply them with relations. These relations have a low weight: if the clues the user supplies are correctly lexicalized, these relations will not be used by Totaki; on the other hand, if the clues are weakly (even not) lexicalized, then these relations allow Totaki to make proposals, while the first experiments showed that it would have been able not to be able to make them. For the strongly lexicalized terms, there are relatively few modifications: most of the relations *term → link* already existed in the network.

This process, just like the use of Totaki in taboo mode (section 2.6), should allow to enrich the long tail, and thus increases the recall.

3.2.4. First results

In the current state of our experiments, the JDM network contains about 3 500 000 "Wiki" relations making the total over 6 500 000 relations. Let us remind these relations have a low valuable weight (10 or less), while a relation introduced via JDM by a single couple of players has a higher weight (equal to 50). This translates the ascendancy we want to give to these "initial" JDM relations with regard to those obtained thanks to Wikipedia. The use of Wikipedia so allows to improve the capacities of the JDM network. So, for example, thanks to the Wikipedia links, the clue *Mouilleron-en-Pareds* allows Totaki to find *Georges Clemenceau* (it is its home town). Let us indicate however that, without the use of Wikipedia, Totaki would have been able to find *Georges Clemenceau* with "side" clues such as *tiger* and *victory*.

Also let us indicate that these relations obtained thanks to Wikipedia are simply typed "Wiki", while a lot of types exist for the "initial" JDM relations: this "Wiki" type is comparable to a free association (or *associated_idea*) with a very low weight.

3.3. Use of the contributions awaiting validation

Approximately 1 million relations proposed by the contributors via Diko are currently awaiting validation. Thus, they are not present, even temporarily, in the JDM network. Why to ignore such a mass of information which by experiment we found very relevant in almost all of the cases? The basic idea thus consists in using these relations, by allocating them a low weight, much lower than that of the relations of the JDM network. As a consequence, the relations of the JDM network have priority compared with these pseudo-relations, in the same way as the "Wiki" relations. These pseudo-relations, informed as we previously saw on often very specific and little lexicalized terms, can allow Totaki to provide answers, in particular in cases it could not be able to supply them.

4. Conclusion and perspectives

Totaki on-line game was initially designed in association with Michael Zock with the aim of estimating the validity of the JDM network. From the beginning of this design, a duality appeared: Totaki can be considered as a game (just like JDM), but also as a tool for the TOT problem. In this second approach, Michael Zock's contribution was determining.

Thanks to the help of players looking ceaselessly for its limits, Totaki allows to enrich in a consequent way the JDM network, mainly by allowing the acquisition of "side" relations, as well as that of new terms.

The use by Totaki of links from Wikipedia pages in the search for proposals, from the clues supplied by the players, as well as the use of the contributions of certain players (proposed, but not yet validated relations) allowed an important densification of the JDM network, mainly for terms which the games of the JDM project had relatively weakly lexicalized.

References

Carlson A., Betteridge J., Kisiel B., Settles B., Hruschka E.R., Mitchell T.M.: *Toward an Architecture for never-Ending Language Learning*, In Proc. of the Conference on Artificial Intelligence (AAAI), 8 p (2010)

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

Ferrucci D., Brown E., Chu-Carroll J., Fan J., Gondek D., Kalyanpur A., Lally A., Murdock W., Nyberg E., Prager J., Schlaefer N., Welty C. : *Building Watson: An Overview of the DeepQA Project*, AI Magazine, 31 (3), pp. 59-79 (2010)

Joubert A. and Lafourcade M.: *A new dynamic approach for lexical networks Evaluation*, In Proc of the 8th edition of Language Resources and Evaluation Conference (LREC 2012), Istanbul, May 2012, 5 p, ELRA Ed.

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

Lafourcade M., Joubert A., Schwab D. et Zock M.: *Évaluation et consolidation d'un réseau lexical grâce à un assistant ludique pour le " mot sur le bout de la langue "* In Proc of TALN'11, Montpellier, France, 27 juin-1er juillet 2011, pp. 295-306

Lafourcade M. and Joubert A.: *Bénéfices et limites de l'acquisition lexicale dans l'expérience JeuxDeMots*, In Ressources Lexicales, Gala N. and Zock M. (eds), Dec. 2013, pp. 187-216

Polguère A.: *Structural properties of Lexical Systems: Monolingual and Multilingual Perspectives,* In: Workshop on Multilingual Language Resources and Interoperability (COLING/ACL 2006), Sydney, pp. 50-59 (2006)

Mel'čuk I.A., Clas A., Polguère A.: *Introduction à la lexicologie explicative et combinatoire,* Editions Duculot AUPELF-UREF (1995)

Zarrouk M., Lafourcade M. and Joubert A.: *Inductive and deductive inferences in a Crowdsourced Lexical-Semantic Network.* In proc of the 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013), Hissar, Bulgaria, September 7-13, 2013, 6 p

Zock M. and Schwab D.: *L'index, une ressource vitale pour guider les auteurs à trouver le mot bloqué sur le bout de la langue*, In Ressources Lexicales, Gala N. and Zock M. (eds), Dec. 2013, pp. 313-354

Annexes

Other examples of Totaki games

Vos indices	Mes réponses
nourriture	manger
:good	viande

Vos indices	Mes réponses
nourriture	manger
:bad	junk food

Vos indices	Mes réponses
sentiment	amour
:bad	haine

Vos indices	Mes réponses
:isa animal	tigre
:good	chat
peluche	ours

Vos indices	Mes réponses
:isa animal	tigre
WWF	panda

Vos indices	Mes réponses
:isa animal	tigre
:long = 13	poisson rouge