Capitalisation d'une ressource en or : le dictionnaire

Michael Zock

Capitalisation d'une ressource en or : le dictionnaire

Michael Zock

Laboratoire d'Informatique Fondamentale (LIF) – CNRS
michael.zock@lif.univ-mrs.fr

Résumé

Notre objectif est de montrer comment l'ajout de certaines fonctionnalités à un dictionnaire électronique existant pourrait aider des êtres humains à apprendre ou à traiter la langue. Ces extensions concernent l'accès aux mots, leur mémorisation et l'acquisition d'automatismes pour produire des structures syntaxiques fondamentales d'une langue. Pour atteindre ce dernier objectif, nous proposons un générateur de phrases paramétrable, basé sur la notion de formulaires : l'utilisateur indique son *intention de communication* et les *mots* à utiliser (apprendre), et le programme construit les phrases correspondantes. Quant à l'aide à l'accès lexical, nous proposons d'ajouter, à un dictionnaire électronique existant, un index basé sur la notion d'association. L'index est construit à partir des mots co-occurrents d'un corpus qui, lui est supposé représenter les connaissances du monde de l'homme de la rue. La recherche lexicale se fait par navigation : partant d'un mot ou d'une idée, on s'approche progressivement du candidat lexical recherché.

Mots-clés : extensions aux dictionnaires électroniques, générateur d'exercices linguistiques, mémorisation de mots, acquisition de réflexes linguistiques, outils d'aide à la navigation dans un espace lexico-conceptuel, index basé sur la notion d'association.

Abstract

The goal of this paper is to explore extensions to electronic dictionaries. Adding certain functions could considerably extend the range of tasks for which they could provide support. Putting the needed information at the distance of a mouse click would allow for *active* reading. This would require tight coupling of the dictionary with a text editor: all the information in the dictionary should be accessible via a mouseclick. Dictionaries combined with a flashcard system and an exercise generator could support the *memorization* and *automation* of words and syntactic structures. Finally, structuring the dictionary in a way akin to the human mind (associative network) could help the writer to find new ideas, and if needed, the word he is looking for. In sum, rather than considering the dictionary just as another component of the process of language production or comprehension chain, we consider it as the single most important resource, provided that one knows how to use it.

Keywords: extensions to electronic dictionaries, exerciceexercise generator, memorisation of words, acquisition of linguistic skills (habits), tools for assisting navigation in a conceptual-lexical network, association-based index.

1. La problématique

L'objectif de ce papier est triple : (a) montrer comment au prix de quelques extensions à un dictionnaire électronique existant, on pourrait aider des êtres humains à apprendre ou à traiter la langue ; (b) convaincre les chercheurs du monde du TAL qu'il vaut mieux construire des petits modules, mais ouverts que des systèmes complets et fermés ; (c) informer les étudiants en langues de l'existence de certains outils disponibles sur le web pour envisager l'apprentissage différemment.

Si les méthodes de langue ont sûrement des qualités, elles ont également toutes au moins un défaut : elles sont fermées. Or, les besoins des apprenants sont variables, souvent

imprévisibles et de toute façon sujets à des changements. Aucune méthode ne pourrait jamais convenir à tout le monde à tout moment.

Il y a maintes manières d'apprendre une langue, celle de l'école est artificielle qu'on le veuille ou non. Pourtant, il y a d'autres manières d'apprendre bien plus naturelles, en l'occurrence l'apprentissage incident. En effet, c'est beaucoup plus naturel de lire un journal ou d'écouter des nouvelles que de faire des exercices comme on vous le demande à l'école. Pourtant, en se livrant à ce type d'activités extra scolaires on apprend également, mais le résultat est un effet de bord et non pas l'objectif premier[1]. Enfin, si on apprend pour la (sur)vie, c'est qu'on apprend également pendant toute la vie, surtout à notre époque où tout change si rapidement. C'est ici que le TAL pourrait rendre d'énormes services, en assistant la formation continue (ou l'auto-formation) et l'apprentissage incident.

Dans ce qui suit nous allons montrer comment on pourrait assister le lecteur en intégrant les dictionnaires à un éditeur de texte, si bien qu'en cliquant sur un mot on verrait apparaître les informations contenues dans les différentes rubriques (traduction, informations grammaticales), etc. Pour aider l'étudiant à acquérir une certaine aisance verbale (fluidité) nous présenterons deux outils dont les buts respectifs sont la *mémorisation* de mots et l'acquisition d'*automatismes* concernant les structures fondamentales d'une langue. Enfin, pour aider les rédacteurs à trouver les mots cherchés, nous présentons une méthode pour les mettre sur la bonne voie. En particulier, nous proposons d'ajouter à un dictionnaire électronique existant un index basé sur la notion d'association. L'index étant construit à partir de mots co-occurrents issus d'un corpus représentatif des connaissances du monde du citoyen moyen. Un dictionnaire enrichi de ce type d'information permettrait donc d'initier la recherche avec des mots quelconques pour s'approcher progressivement du candidat idéal.

2. FonctionalitésFonctionnalités assistant le *récepteur :* mettre l'information recherchée au bout des doigts

Étant habitués à des langues comme l'anglais, l'allemand ou l'espagnol, nous avons tendance à oublier qu'il y a des langues dont l'écriture est très différente. Outre le problème de la morphologie (les entrées des dictionnaires sont généralement des lemmes et non pas la forme fléchie), il y aura donc en plus celui du déchiffrage de caractères. Supposons qu'on veuille apprendre une langue dont l'écriture ne se fait pas en termes de caractères latins (grec, russe, coréen, japonais, chinois). Comment consulter le dictionnaire dans ce cas là? là ? C'est pratiquement impossible, pourtant il existe une solution relativement simple et pour la plupart des cas satisfaisante.

Prenons comme exemple un texte en japonais でんわほどこですか. L'apprenant est confronté ici à au moins deux problèmes : celui d'identification des frontières lexicales (les mots ne sont pas forcément séparés par un blanc), et celui de la conversion des symboles (kana) en phonèmes/graphèmes (で → de)[2]. Ces deux informations sont bien entendu capitales. Ainsi, on aimerait voir apparaître sur écran la prononciation correspondante à la chaîne de caractères でんわ, à savoir, denwa. Même si cela ne nous dit pas pour autant que ce mot signifie 'téléphone', on aurait pu l'apprendre ensuite, car, grâce à la translittération, on pourrait

[1] Bien entendu, il ne s'agit pas de bannir les cours de langue. Le point ici est plutôt de montrer qu'on peut construire, à coup raisonnable, des béquilles fort utiles pour l'utilisateur de la langue, qu'il soit déjà expert ou encore apprenant.

[2] La situation est encore un peu différente dans une langue comme le chinois qui utilise des idéogrammes.

consulter un dictionnaire (à condition que les entrées soient en lettres romanes), ou demander à une personne sachant parler cette langue.

Trouver des informations dans un dictionnaire n'est donc pas toujours chose aisée, d'abord cela demande souvent beaucoup de temps, et il peut y avoir des problèmes liés à l'écriture et à la morphologie. Pourtant, les problèmes mentionnés pourraient trouver une solution simple et satisfaisante pour bon nombre de cas. Il suffirait d'ajouter au dictionnaire un translittérateur[3] et un lemmatiseur[4]. En intégrant ce type de fonctionnalité dans des applications courantes (traitement de texte, butineur), on pourrait désormais lire et comprendre des textes écrits dans une langue étrangère (*lecture active*). Il suffirait alors de cliquer sur un mot pour voir apparaître un menu "pop up" permettant au lecteur de choisir parmi les informations celle qui l'intéresse à cet instant (traduction, définition, information grammaticale, etc.). Par exemple, on pourrait imaginer l'interface suivante (cf. figure 1) : la traduction concerne ici le mot « shite », forme conjuguée de l'entrée lexicale « suru » (verbe à l'infinitif)

À noter qu'on commence à voir des programmes capables de faire cela, comme par exemple, GLOSSER (Nerbonne et Smit, 1996), un prototype développé au GETA par Mathieu Lafourcade, et les produits de la société *Transparent Language*. Hélas, tous ces systèmes sont fermés, et il n'y actuellement aucun moyen d'y ajouter d'autres fonctionnalités. Enfin, les urls suivantes méritent considération, surtout pour ceux souhaitant apprendre à lire en chinois ou en japonais : (http://language.tiu.ac.jp/tools_e.html ; http://www.rikai.com/perl/HomePage.pl?Language=Ja; http://www.popjisyo.com/ ; http://www.popjisyo.com/WebHint/Portal_e.aspx ; http://www.newsinchinese.com/. La même remarque vaut pour les sites suivants : http://www.animelab.com/anime.manga/translate ; http://www.animelab.com/anime.manga/dictionary/ ; http://www.eloquentsw.com/livedictionary.html, mais cette fois-ci pour des problèmes liés aux dictionnaires.

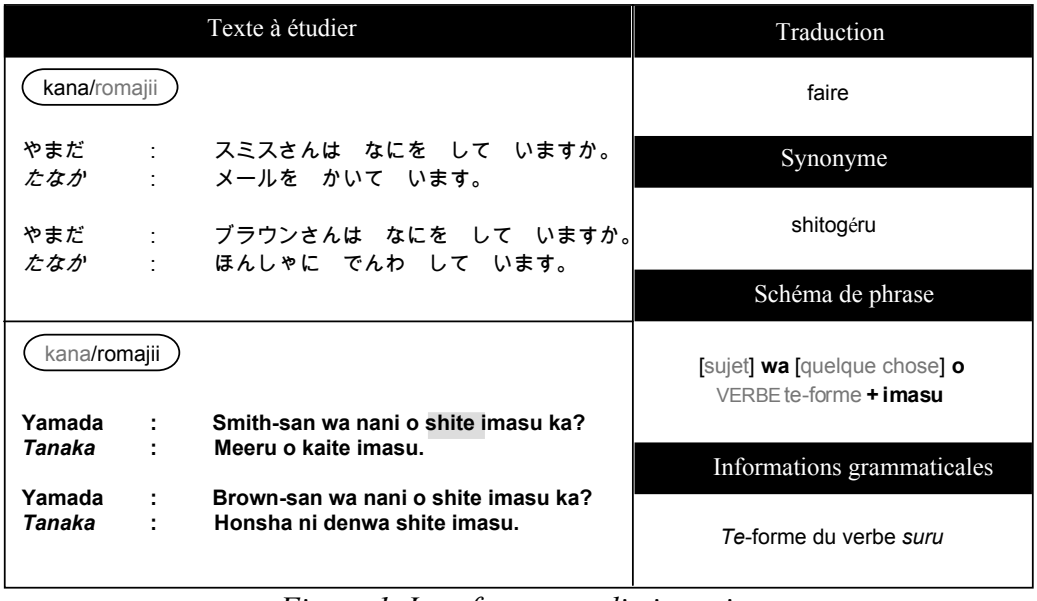

Figure 1. Interface texte-dictionnaire

[3] La translittération est l'opération consistant à transcrire les graphèmes d'un alphabet ou d'un syllabaire (comme le japonais) dans les graphèmes d'un autre système d'écriture (généralement un alphabet), de telle sorte qu'à un même graphème (ou suite de *graphèmes*) de la langue de départ corresponde toujours un même graphème (ou suite de graphèmes) du système d'écriture d'arrivée, et ce indépendamment de la prononciation (http://fr.wikipedia.org/wiki/Translitération).

[4] Un lemmatiseur est un programme qui permet de passer d'un mot portant des marques de flexion (pluriel, forme conjuguée d'un verbe...) à sa forme de référence (entrée lexicale, lemme) ou inversement.

3. Fonctionnalités assistant les *producteurs* de langue

Les besoins sont bien entendu différents selon que l''on est expert ou étudiant, en train d'apprendre une langue. Commençons par ce dernier.

3.1. Assister l'étudiant de langues

Pour pouvoir s'exprimer dans une langue, il faut avoir non seulement énormément de connaissances, mais également posséder un savoir-faire non négligeable. Ainsi, un locuteur doit-il pouvoir trouver le mot exprimant sa pensée[5], l'insérer au bon endroit de la phrase, l'adapter morphologiquement, tout en continuant à planifier l'idée suivante, et tout ceci en un très court laps de temps. Si jamais une de ces étapes tarde ou échoue, on assiste à des lapsus, bafouillages, sons de remplissage, ou, des pauses plus ou moins prononcées, pouvant aller jusqu'au silence total. L'apprentissage du *vocabulaire* (mémorisation) et des *structures syntaxiques*, tout comme l'*acquisition* d'*automatismes* est donc indispensable pour pouvoir produire des phrases à un débit *normal*.

*3.1.1. Apprentissage de vocabulaire : **mémorisation** de mots*

Dire que l'apprentissage de mots est fondamental est trivial, ce qui l'est moins c'est de dire comment, car, même appris, les mots semblent avoir la fâcheuse tendance de vouloir s'échapper de notre mémoire.

Se basant sur les travaux des psychologues étudiant la mémoire, Leitner (1972) a proposé une application intéressante. Le principe est simple. L'auteur propose de ranger dans une boîte à cinq compartiments des cartes contenant d'un côté la question (par exemple, un *mot à traduire*) et de l'autre la réponse (*mot traduit*). Toutes les cartes se trouvent dans le premier compartiment au début de l'apprentissage, pour passer successivement au compartiment suivant (ou précédent), tout dépend de la qualité de la réponse (bonne/mauvaise). La leçon est considérée acquise lorsque toutes les cartes sont dans le dernier compartiment. L'idée sous-jacente à cette méthode est simple : consacrer un maximum de temps aux éléments récalcitrants. L'idée ne date pas d'hier, en fait elle est connue sous le nom de la *loi de Jost*, selon laquelle un apprentissage fractionné et espacé dans le temps est plus efficace qu'un apprentissage regroupé (Kekenbosch, 1991 : 7-13).

Bien sûr une version informatique n'a de sens que si elle apporte quelque chose par rapport à la version papier. Or, c'est clairement le cas. La souplesse et la puissance informatique sont au rendez-vous. Les données sont faciles à acquérir, à échanger et à mettre à jour. Par exemple, on peut imaginer la construction de ponts mnésiques (associations) entre la question et la réponse. L'ergonomie, l'ouverture, la gestion des performances (décompte automatique) sont toutes des facteurs apportant un confort indéniable à l'utilisateur. Quant aux paramètres de présentation des données (couleur, taille, nombre de présentations, vitesse de défilement, etc.) il y a de très nombreuses possibilités pour adapter l'outil à son goût et ses besoins. Enfin, on peut même imaginer la mise au point de stratégies pour tester leurs avantages respectifs : *précision* vs *rapidité*.

[5] Ce qui veut dire, qu'il doit chercher dans un stock énorme (les chiffres avancés varient selon les auteurs entre 30 à 60 000 mots.) un élément particulier. La performance est impressionnante, équivalent à la consultation d'un dictionnaire comme *Le Grand Robert* trois fois par seconde, et ceci pendant plusieurs heures.

*3.1.2. **Mémorisation** et **automatisation des** structures fondamentales **d'une langue***

Posséder un grand vocabulaire (même actif) ne signifie pas encore savoir produire des phrases. Pour cela il faut savoir (et savoir-faire) bien plus de choses. Il y a différentes manières de produire une phrase : (a) on peut la produire pièce par pièce (mot par mot) en recourant à une grammaire formelle ;; (b) on peut faire appel à des fragments de taille variable, allant d'expressions toutes faites jusqu'à la phrase ;; (c) on peut utiliser des schémas de phrase que l''on remplit ensuite avec des données lexicales.

Cette dernière solution est une sorte d'heureux compromis entre la génération incrémentale et celle consistant à réutiliser des pièces toutes faites. La première étant basée sur une grammaire, donc souple et puissante, mais lente et assez complexe, tandis que la seconde est rigide, mais rapide et extrêmement simple. Ce n'est donc pas étonnant de constater que la méthode représentant le meilleur compromis soit celle retenue dans l'apprentissage naturel d'une langue et dans l'enseignement des langues.

Nous allons esquisser ici comment, partant d'un dictionnaire, on peut construire ce type de générateur de phrase. En fait, comme nous allons voir, ce n'est pas seulement un générateur de phrases, mais un générateur d'exercices. L'idée est toute simple. On indexe l'ensemble des structures à apprendre en termes de buts, puis on demande à l'utilisateur d'en choisir un et de remplir les variables de la structure correspondant au but avec des données lexicales. Ceci étant fait, le système a tout ce qu'il faut pour construire des phrases[6] (Zock et Quint, 2003).

Prenons un exemple. Supposons qu'on veuille exprimer le but suivant : *définition (animal)*. Pour cela il y a plusieurs schémas en français, par exemple : (a) «un **X** est un **Y** qui **ACTION**» (b) «un **X** est une espèce de **Y** vivant en **Z**», où **X**, **Y**, **Z** et **ACTION** sont des variables (**X** : nom d'animal; **Y** : hyperonyme ; **Y** : lieu) pour lesquelles il faudrait préciser la valeur lexicale. Ceci étant fait, le système pourrait alors produire des phrases du type:type : (a) un **perroquet** est un **oiseau** qui **parle**, ou (b) un **koala** est une espèce de **marsupial** vivant en **Australie**.

Prenons un autre exemple. Supposons qu'on veuille apprendre en japonais l'équivalent du français « *où est* **x** ? », but ou intention qu'on pourrait communiquer soit en choisissant dans un menu, soit en ayant recours à un langage de requête : `lieu (x), x` étant l'objet pour lequel on demande la localisation.

Connaissant désormais l'*intention* de communication, le système présenterait alors la ou les structures correspondantes, en l'occurrence « x-*wa doko desu ka* », attendant qu'on lui précise la ou les valeurs de « x ». Supposons qu'elles soient « *arrêt de taxi*, *banque* et *hôpital* ». Le système aura donc désormais tout ce qu'il faut pour produire les phrases contenant ces éléments, mais, comme il s'agit d'un exercice dont le but est d'aider l'étudiant à produire ces phrases, on incite ce dernier à essayer d'abord lui-même, avant que le système ne produise sa version. Ainsi, le système présente une amorce, attendant que l'élève l'insère au bon endroit de la structure correspondant à son intention. C'est en effectuant les opérations requises tout en comparant les résultats (la sienne et celle du système) que l'étudiant apprend.

Utilisateur *:*	*Lieu x ?*	(intention de communication)
Système :	x wa doko desu ka?	(structure correspondante)
	x = ?	(demande de précision de valeur)
Utilisateur :	x : *arrêt de taxi*, banque, hôpital	(valeurs lexicales)

[6] Bien sûr, ce type de générateur est d'autant plus limité que le système n'a pas de composant morphologique. Certes, on pourrait l'inclure, mais le point ici était justement de montrer qu'on pourrait créer un générateur de phrase (ou d'exercice) avec un minimum d'informations. Celles-ci se trouvent pratiquement toutes dans le dictionnaire. Il n'y a que les buts qui n'y figurent pas.

Système :	*Takushii-noriba* wa doko desu ka?	(insertion dans la structure de la phrase)
Système :	*banque*	(amorce)
Utilisateur :	*Ginko* wa doko desu ka?	(réponse étudiante)
Système :	Ginko wa doko desu ka?	(confirmation système)
Système :	hôpital	(amorce suivante)
		etc.

3.1.3. Discussion

Le type d'exercice que nous venons de proposer existe depuis fort longtemps. Ce sont les fameux pattern-drills (*exercices structuraux* en français), très en vogue pendant les années 50 et 60, à l'époque où les laboratoires de langue et les méthodes inspirées des idées behavioristes (notamment les méthodes audio-orales) avaient le vent en poupe. Pourtant, cette technique avait également ses détracteurs.

Si le behaviorisme de Skinner (1968) a néanmoins tant inspiré le monde éducatif, malgré les très nombreuses critiques de la part de *linguistes* (Chomsky, 1959), de *didacticiens* (Besse 1975) et de *psychologues* (Chastain, 1969; Le Rouzo, 1975), c'est qu'on trouve à sa base deux principes fondamentaux de l'apprentissage : celui de la ***rétroaction***, information concernant la qualité d'une réaction (réponse) à un problème (stimulus) et celui de la ***répétition***[7]. S'ajoute à cela un troisième élément, celui de la ***structure***, nommée jadis Gestalt, ou, selon les périodes, patron (pattern), schéma, cadre, frame. Ce que l''on cherchait à capter par ces termes, c'étaitest l'idée, que derrière une masse de formes variables il y a un invariant, la structure sous-jacente. Vu la généralité et la complémentarité de ces principes, il n'est donc pas étonnant de les trouver à la base de certaines théories comme le structuralisme ou l'apprentissage, ou encore à la base de certaines pratiques didactiques comme l'enseignement programmé (Pocztar, 1971) ou les *exercices structuraux*.

Certes, cette forme d'exercices n'est pas une panacée, mais utilisée à bon escient elle peut s'avérer utile, ne serait-ce que pour mémoriser et automatiser les mots dans le contexte de la phrase. Ainsi faisant, elle libère l'apprenant des aspects élémentaires de la langue (éléments mécaniques et de bas niveau) pour lui permettre d'accéder aux niveaux supérieurs, ceux des idées (planification de messages). Aussi, qu'on le veuille ou non, la répétition des formes (structures) est le prix à payer pour acquérir la maîtrise d'une activité aussi complexe que la production du langage.

Cependant, comme tous les praticiens le savent, les exercices structuraux souffrent de certaines faiblesses évidentes. Ils sont rigides et ils engendrent rapidement une certaine lassitude, ce qui est partiellement lié au caractère fermé des supports (livre, magnéto). Tout doit être prévu, et rien ne peut être changé après impression et/ou enregistrement. Or, ceci a complètement changé avec l'arrivée des ordinateurs. Désormais on peut changer les données à tout moment, pour les adapter en fonction des besoins du « client ». Or, ceux-ci varient non seulement d'une personne à l'autre, mais aussi intra-individuellement. Nos besoins changent à tout moment, d'où l'intérêt de construire des outils ouverts, adaptables en fonction des besoins du moment.

Enfin, il y a d'autres manières d'apprendre ces structures. Nous avons montré ici une façon parmi d'autres pour construire sa base. La méthode est interactive et fait appel à un générateur

[7] Que l'apprentissage (mémorisation), l'adresse (habileté) et la perfection demandent de l'exercice est connu depuis fort longtemps (practice makes perfect), si bien que les débuts de la psychologie expérimentale coïncident pratiquement avec leur étude. En effet, Ebbinghaus a étudié dès 1885 le rôle de la répétition (nature, espacement, etc.) dans l'apprentissage. Pour une revue de la question, voir Hilgard et Bower (1975).

(rudimentaire, certes). Cependant, on pourrait également constituer ce type de base en fouillant un corpus comme le web. Bien sûr, un débutant ne connaît pas forcément les schémas réalisant une intention de communication, mais le système les connaît (à condition de le lui avoir appris). Ayant fourni au système votre intention de communication, celui-ci peut vous indiquer le ou les schémas permettant sa réalisation. Désormais on pourrait donc lancer une recherche en prenant ce schéma comme filtre et fouiller le web pour trouver des instances (exemples).

3.2. Assister le rédacteur à trouver le mot recherché

La production du langage peut être vue comme une forme de réécriture de sens ou d'idées par des mots. L'hypothèse sous-jacente étant que les idées[8] précèdent les mots. Ayant à l'esprit un sens il y a plusieurs cas de figure lors de la consultation du dictionnaire. En particulier, on peut distinguer la qualité d'entrée (ce que le locuteur sait à cet instant précis : sens, et/ou forme) et la qualité de la sortie, proximité *formelle* et *sémantique* entre le mot source (MS, celui qu'il est capable de produire) et le mot cible (MC), le mot recherché. Autrement dit, l'information fournie au moment de la requête peut être très variable (sens, mots), tout comme la distance entre le MS et le MC qui peut être plus ou moins grande : (a) le MS et le MC peuvent avoir des rapports de sens ("chaud-froid"; "jaune-banane"; "fruit-banane", "dog-chien", etc.) sans qu'il n'y ait de rapport formel ; (b) ils peuvent avoir des rapports formels, sans avoir de rapports sémantiques (vin vs vingt) ;; (c) ils peuvent avoir et des rapports formels et des rapports sémantiques (chat-rat) ;; (d) le MS peut être formellement proche du MC (reléguer vs dé*léguer*).;

Seul le premier cas de figure nous intéresse ici. La réalisation d'une méthode d'accès par la forme a été décrite dans (Zock et Fournier, 2001). Nous ne nous y attardons donc pas ici. Nos efforts actuels sont concentrés sur l'accès par le sens ou plutôt sur l'accès par les mots liés à l'idée à exprimer (carnaval-Brésil). En clair, nous nous intéressons aux rapports associatifs, l'hypothèse étant, que le dictionnaire mental est un vaste réseau dont les mots sont les noeuds et les liens, les associations.[9] L'accès au mot s'effectuera par navigation. On entre dans le réseau en donnant un mot (MS) proche du MC pour recevoir en sortie tous les mots associés à ce dernier. En choisissant parmi ces éléments un candidat prometteur pour le soumettre à nouveau on s'approche progressivement du mot cible.

Prenons un exemple. Supposons qu'on cherche le mot ***infirmière*** (*MC*), alors que le seul mot qui nous vienne à l'esprit (*MS*) est ***hôpital***. Le système prendra alors celui-ci comme noyau pour présenter tous les mots (satellites) ayant un rapport direct avec lui, par exemple, les *employés/méronymes* (médecin, infirmière), des *sous-types* (clinique, sanatorium), etc. C'est à l'utilisateur de décider dans quelle direction continuer la recherche, car il sait généralement quel mot de ceux présentés par le système est le plus prometteur. Celui-ci deviendra donc à son tour le noyau (point de départ), susceptible de produire d'autres candidats. Le tout s'arrête lorsqu'on a trouvé le mot recherché.

À noter que le graphe de la figure 2a n'est qu'une *représentation interne* du système, l'utilisateur ne voit successivement que desque des paquets de mots, paquets qui sont créés

[8] Bien qu'en absence d'un support matériel, c'est moins trivial qu'il n'y paraît. Voir tous les débats tournant autour des questions concernant les rapports langage-pensée.

[9] Cette hypothèse est à nuancer cependant, selon qu'on parle d'une simulation des fonctionnalités de la mémoire mentale (comme nous le faisons ici) ou du cerveau lui-même. En lisant les travaux de certains psychologues, et en regardant leurs simulations on a l'impression que le cerveau est plutôt une usine "fabriquant" des mots qu'un lieu de stockage de mots (Zock, 2005).

grâce aux liens connus par le système (figure 2b). C'est justement pour éviter de noyer l'utilisateur dans un amas d'information qu'on a recours à ces liens qui serviront alors comme critères de sélection ou aides de navigation.

Bien entendu, pour pouvoir fonctionner selon la manière esquissée ici (navigation par association) il faut d'abord créer le réseau, typer les liens et déterminer leur poids. Nous avons commencé ce travail (Zock et Ferret, (2006) en utilisant l'extracteur de collocation de Ferret sur un corpus un peu particulier, le journal *Le Monde*. Bien entendu, le choix du corpus est très important dans la mesure où il est censé représenter les connaissances du monde de l'homme de la rue. Qui plus est, les poids devraient être pondérés en fonction du thème. Ce dernier point n'est pas un problème simple. Enfin, si l'intuition selon laquelle le dictionnaire mental (d'aucuns préféreraient parler d'encyclopédie) serait un vaste réseau, dont les *noeuds* sont des mots (et/ou des concepts), et les *liens* essentiellement des associations ne date pas d'hier, et si elle est partagée par de nombreux chercheurs[10], il n'y a à notre connaissance aucun inventaire exhaustif (ou de classification) de ces liens. Or, connaître leur nature est l'une des conditions préliminaires pour indexer un tel dictionnaire, et c'est là un de nos objectifs des années à venir. Pour une feuille de route, voir (Zock et Bilac, 2004).

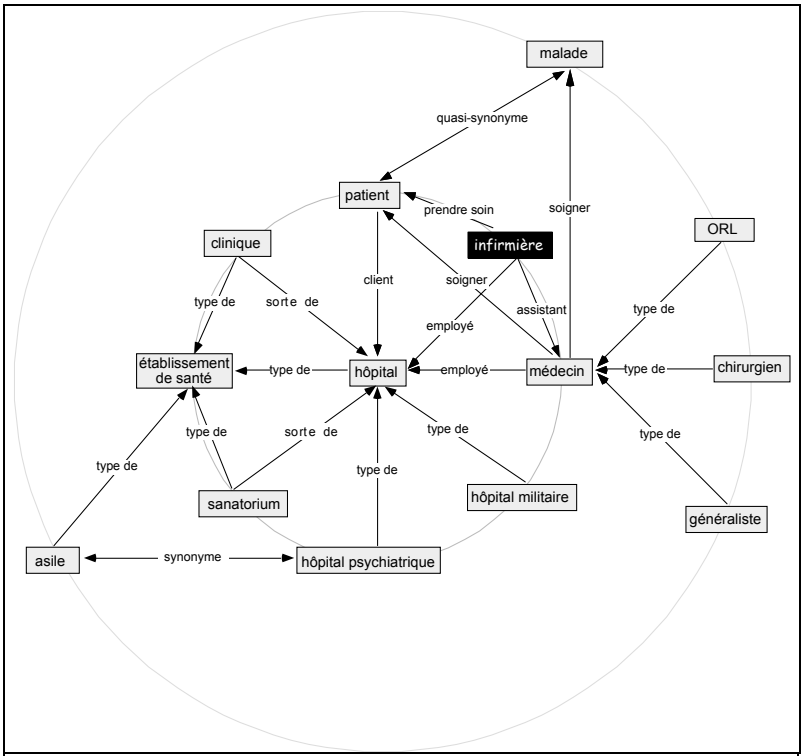

Figure 2a. Représentation interne

sorte de/type d'
établissement de santé
hôpital
asile
hôpital psychiatrique
hôpital militaire
sanatorium
clinique

méronymes
médecin
infirmière
malade

type de
ORL
Chirurgien
généraliste

Figure 2b. Candidats proposés

4. Conclusions

Les composants esquissés ci-dessus tournent tous autour de l'idée d'une meilleure utilisation des ressources lexicales, l'accent étant mis sur la *réutilisation* (capitalisation) des informations stockées dans le dictionnaire et sur l'aide à la navigation. Il est clair qu'un dictionnaire est un composant fondamental pour tout système de traitement de la langue. Mais un dictionnaire ne vaut que par l'information qu'il contient et par les moyens qu'il offre pour

[10] En effet, cette intuition se trouve déjà chez Aristote (« De memoria et reminiscentia »), chez des *psychologues* (Galton, 1880) et *psycholinguistes* (Deese, 1965). Enfin, cette idée est sous-jacente à WORDNET (Miller, 1990). Pour des synthèses en psycholinguistique voir (Hörmann, 1972 ; chapitres 6-10).

accéder à l'information. À l'heure actuelle, il y a un fossé énorme entre les dictionnaires papier, les dictionnaires électroniques et le dictionnaire mental. L'architecture particulière de ce dernier lui confère un énorme pouvoir en termes d'organisation et de souplesse d'accès. Contrairement à une hiérarchie avec une seule voie d'accès, dans ce réseau hautement interconnecté il y a presque toujours un moyen d'accéder à l'information recherchée. De ce fait, le dictionnaire mental constitue un excellent modèle en termes de stockage et d'accès d'informations. Si les dictionnaires traditionnels sont passifs et assez limités en termes d'accès, les dictionnaires électroniques ont un potentiel considérable, susceptible de présenter rapidement et sous des formes diverses l'information recherchée. Les idées présentées ici sont une première tentative allant dans ce sens, mais il est clair, que beaucoup de travail reste encore à faire, notamment au niveau des liens (associations).

Références

BESSE H. (1975) « De la pratique aux théories des exercices structuraux ». In *Études de Linguistique Appliquée* 20 : 8-30.

CHASTAIN K. (1969). « The audio-lingual habit learning theory vs. the code-cognitif learning theory ». In *IRAL* 7, 2 : 97-107.

CHOMSKY N. (1959). « Critique de *Verbal Behavior* de B.F. Skinner », dans *Language* 35 : 26-58.

DEESE J. (1965). *The structure of associations in language and thought*. Baltimore.

GALTON F. (1880). « Psychometric experiments ». In *Brain* 2 : 149-162.

HILGARD E., BOWER G. (1975). *Theories of learning*. Englewood Cliffs, N.J.

HÖRMANN H. (1972). *Introduction à la psycholinguistique*. Larousse, Paris.

KEKENBOSCH C. (1991). *La mémoire et le langage*. Nathan Université, Paris.

LE ROUZO M.L. (1975). « Y a-t-il une justification psychologique à la pratique des exercices structuraux ? ». In *Études de Linguistique Appliquée* 20 : 37-51.

LEITNER, S. (1972). *So lernt man lernen.* (Voici comment-on apprend). Heider, Freiburg.

MILLER G.A. (éd.) (1990). « WordNet: An On-Line Lexical Database ». In *International Journal of Lexicography* 3(4).

NERBONNE J., SMIT P. (1996). « GLOSSER: in Support of Reading ». In *COLING '96*, Copenhagen : 830-835.

POCZTAR J. (1971). « En enseignement programmé, quoi de nouveau? nouveau ? ». In Revue française de pédagogie 15 : 5-14. (voir aussi du même auteur: *Théories et pratique de l'enseignement programmé*).

SKINNER B.F. (1968). *La révolution scientifique de l'enseignement*. Dessart, Bruxelles.

SPOLSKY B. (1966). « A Psycholinguistic Critique of Programmed Foreign Language Instruction ». In *IRAL* 4, 2.

ZOCK M. (2005). « Le dictionnaire mental, modèle des dictionnaires de demain? demain ? ». In *Revue Française de Linguistique Appliquée*, Vol. X, 2005-2.

ZOCK M., FERRET O. (2006). « Enhancing electronic dictionaries with an index based on associations » (en préparation).

ZOCK M., QUINT J. (2004). « Why have them work for peanuts, when it is so easy to provide reward? Motivations for converting a dictionary into a drill tutor ». In *Papillon*, *5th workshop on Multilingual Lexical Databases*, Grenoble.

ZOCK M., BILAC S. (2004). « Word lookup on the basis of associations :associations: from an idea to a roadmap ». In *Proceedings of Coling workshop : Enhancing and using dictionaries*, Genève : 29-34.

ZOCK M., FOURNIER J.-P. (2001). « How can computers help the writer/speaker experiencing the Tip-of-the-Tongue Problem ? Problem? » In *Proceedings of RANLP*, Tzigov Chark : 300-302.