

L'ordinateur assistant l'être humain à trouver le mot que celui-ci a sur le bout de la langue

Michael Zock

Groupe Langues, Information et Représentations
LIMSI-C.N.R.S., B.P.133, 91403 Orsay
zock@limsi.fr

Résumé : Tout le monde a déjà rencontré le problème suivant : on cherche un mot (ou le nom d'une personne) que l'on connaît, sans être en mesure d'y accéder à temps. Les travaux des psychologues ont montré que les personnes se trouvant dans cet état savent énormément de choses sur le mot recherché (sens, nombre de syllabes, etc.), et que les mots avec lequel ils le confondent lui ressemblent étrangement (lettre ou son initial, catégorie syntaxique, champ sémantique, etc.).

L'objectif de notre travail est de réaliser un programme tirant bénéfice de cet état de faits pour assister un locuteur ou rédacteur à (re)trouver le mot qu'il a sur le bout de la langue. A cette fin, nous prévoyons d'ajouter à un dictionnaire électronique existant un index d'association (collocations rencontrées dans un grand corpus). Autrement dit, nous proposons de construire un dictionnaire analogue à celui des êtres humains, qui, outre les informations conventionnelles (définition, forme écrite, informations grammaticales) contiendrait des liens (associations), permettant de naviguer entre les idées (concepts) et leurs expressions (mots). Un tel dictionnaire permettrait donc l'accès à l'information recherchée soit par la forme (lexicale : analyse), soit par le sens (concepts : production), soit par les deux.

1 Introduction

Il est évident qu'un *dictionnaire* est un composant essentiel pour tout système de traitement de la langue (naturelle ou artificielle) qu'il s'agisse d'*analyse* ou de *production*. Il est également clair, qu'un dictionnaire restera d'un intérêt limité si l'information recherchée (mot, expression ou informations associées) n'est pas accessible facilement. Mon objectif ici est de clarifier ce qu'il faut entendre par *accès facile* (ou navigation naturelle) et comment construire cette aide concrètement (implémentation). Je me place du point de vue du locuteur ou rédacteur, grands absents des préoccupations des constructeurs des dictionnaires.¹

2 Pertinence des travaux issues de la génération de texte :

Travaillant dans le domaine du GAT (génération automatique du langage) on est tenté de regarder ce que ces chercheurs ont à proposer. Hélas, la réponse est « rien » pour ce qui nous concerne ici. Pourtant ce n'est pas faute d'avoir essayé. En fait, de très nombreux travaux ont été consacrés au choix lexical,² mais comme nous le verrons, le problème se

¹ Bien sûr il y a des *thesauri* (Roget, 1852 ; Péchoin, 1992), des *dictionnaires analogiques* (Boissière, 1862 ; Robert, 1993), le *Language Activator* de Longman (1993) et WordNet, mais ce sont là plutôt des exceptions. La plupart des dictionnaires ont été faits dans l'optique du récepteur. Cela dit, même si les ouvrages (ou ressources) mentionnés ici, font office d'exception, ils sont encore bien perfectibles pour ce qui est de la navigation (accès à l'information).

² Pour d'excellents résumés de ces 15 dernières années voir Robin (1990) ; Stede (1995), Wanner (1996), ou Cumming (1986) pour des travaux antérieurs. De nombreuses solutions ont été proposées pour tenir compte de différents aspects. Des aspects conceptuels ont été modélisés par des *réseaux de discrimination* (Goldman, 1975 ; Danlos, 1987), des *taxinomies* (Novak, 1987), des *dictionnaires d'énoncés* (Hovy, 1988), des couples *concept-patron* (Jacobs, 1987), des *prototypes* (Nirenburg and Nirenburg, 1988), ou des techniques d'*appariement* (Nogier and Zock, 1992, Horacek 1996). Les connaissances pragmatiques ont été traitées par Hovy (1989), Batemann and Paris (1989), Reiter (1991), Wanner and Maier (1991). Enfin, un très grand nombre de chercheurs se sont penchés sur les aspects linguistiques : Heid & Raab,

pose très différemment en GAT, en génération spontanée et en génération différée, assistée par ordinateur (ou d'autres ressources externes : dictionnaire, encyclopédie, thésaurus).

Voyons donc un instant une des techniques utilisées en GAT : l'appariement. Utilisant un formalisme identique, les graphes conceptuels, pour représenter la sémantique des mots et celle du message, Nogier & Zock (1991) ont montré que la technique de recouvrement pouvait être utilisée pour trouver des mots correspondant aux idées. La lexicalisation se faisant par « consommation » successive du message jusqu'à consommation de ce dernier. L'exemple ci-dessous (Zock, 1996), permet d'illustrer la méthode. Supposons que l'idée à exprimer soit la suivante: « *un homme, dont le métier est d'attraper et de vendre des poissons, se dirige à une certaine vitesse sur une surface terrestre en une certaine direction* » (la direction n'étant pas encore spécifiée, le message est incomplet)».

En regardant la figure 1, on constate que le graphe initial peut être réécrit successivement par « pêcheur » et « se précipiter », dans la mesure où ces mots expriment précisément les parties respectifs du graphe « *un homme, dont le métier est d'attraper et de vendre des poissons* » et « *se diriger à une certaine vitesse en une certaine direction* ».

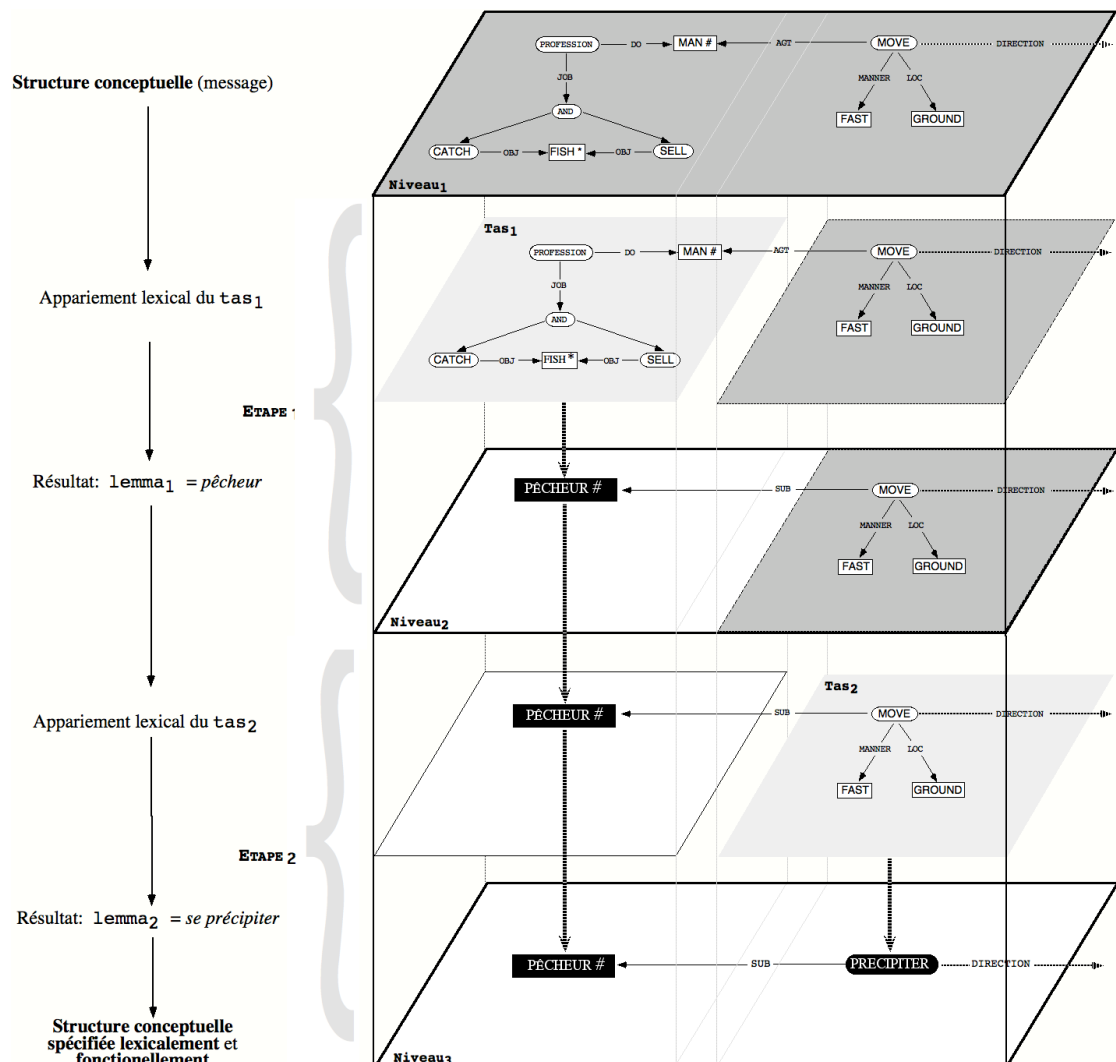


Figure 1 Mise en mots (lexicalisation) par projection (superposition) du graphe lexical (en gris clair) sur le graphe message (gris foncé).

(1989), Wanner & Batemann (1990) et Smadja & McKeown (1992) ont traité des phénomènes de *collocation*, tandis que Danlos (1987) et Jordanskaja & al. (1991) se sont appuyés sur les travaux de très grands lexicographes : Gross (1975) and Mel'cuk (1991).

Si la méthode fonctionne, elle n'est pas sans poser certains problèmes,³ et comme toutes les autres méthodes utilisées dans la communauté GAT, elle a un défaut majeur par rapport au problème qui nous intéresse ici : elle suppose que les mots stockés en mémoire sont accessibles. Bien entendu, si cela vaut généralement pour l'ordinateur, il n'en est pas toujours ainsi pour l'être humain. Qui n'a pas encore connu cet état d'amnésie momentanée, consistant à ne pouvoir accéder à un mot qu'on connaît pourtant bel et bien ? Pour comprendre où trouver la solution, il est utile de s'arrêter un instant et observer l'être humain en train de « lutter avec son cerveau ».

3 Le problème du locuteur

La situation typique du locuteur consiste à « planifier » un message pour chercher ensuite les mots correspondants. Différents cas de figure peuvent se présenter, selon la qualité de l'entrée (spécificité du message à communiquer : partielle/intégrale) et celle de la sortie (qualité du résultat de la recherche lexicale) :

- on trouve un mot, mais celui-ci n'est pas le terme recherché;⁴
- on ne trouve pas le mot, mais on se souvient de beaucoup de détails le concernant : le *champ sémantique*, le *nombre de syllabes*, le *début* et la *fin* du mot, le *genre grammatical*, etc. ;⁵
- on ne trouve rien, la seule chose qui vous vient à l'esprit c'est le sens, une partie du sens (élément de la définition) ou un lien avec d'autres concepts ou mots;

Tous ces cas peuvent être exploités à condition d'admettre que la recherche puisse être initiée à partir d'entrées imparfaites (mot mal orthographié, sens approximatif, etc.). Avant de poursuivre cette idée, voyons un instant ce qui a pu causer notre amnésie locale, les lapsus.

En parlant nous commettons tous des erreurs. Personne n'échappe à cela. Comme les erreurs sont régulières (Fromkin, 1973; Aitchinson, 1987), il est intéressant de les analyser, car elles nous révèlent bien des choses sur l'organisation et le fonctionnement du *dictionnaire mental*.

MOT PRODUIT	MOT CIBLE	TYPE D'ERREUR
1. La route était <i>glassante</i>	<i>glacée</i>	sémantique : compétition entre <i>glacé</i> et <i>glissant</i>
2. Je me suis coupé les <i>doigts</i>	<i>ongles</i>	sémantique : axe paradigmatique
3. Il avait une <i>liaison</i> gauche	<i>lésion</i>	inversion phonologique (axe syntagmatique) ⁶
4. La semaine <i>dernière</i> sera le <i>prochain</i> cours.	<i>prochaine</i>	inversion sémantique : axe syntagmatique
5. La route était <i>grissante</i>	<i>glissante</i>	inversion phonologique (axe paradigmatique) ⁷
6. Une <i>feuille</i> de <i>faye</i>	<i>feuille-payé</i>	inversion phonologique (axe syntagmatique)

³ Par exemple, que faire s'il n'y a pas de mot pour exprimer une partie du graphe, ou si le graphe est sous-spécifié, ce qui est souvent le cas en discours spontané.

⁴ Comme le mot trouvé est généralement lié au mot cible (synonyme, hyperonyme, antonyme), cela permet d'initier la navigation pour s'approcher progressivement du mot cible.

⁵ Les études portant sur les problèmes du manque de mot révèlent que les locuteurs savent souvent énormément de choses sur le mot en question : le *domaine*, une partie du *sens*, *nombre de syllabes*, la *nature* de certaines *syllabes*, *catégorie lexicale*, et parfois même le *genre grammatical* (Brown et McNeill, 1966 ; Burke et al. 1991 ; Vigliocco et al., 1997). Ces travaux ont également montré, que les personnes savent reconnaître immédiatement, et à coup sûr, le mot en question lorsqu'on le leur présente.

⁶ On doit à Rabelais le fameux « Femme *folle* à la *messe* » qui pourrait être expliqué soit par une permutation de *phonèmes* soit celle de *lexèmes*.

⁷ Cette confusion est due au fait que /l/ et /r/ sont phonologiquement très proches. D'ailleurs, dans certaines langues, comme le japonais, par exemple, il n'y a pas de différence. Aussi les Japonais ont-ils souvent du mal à les prononcer *correctement*.

- | | | |
|--|-----------------|-----------------------------------|
| 7. Il a pré endu ne rien savoir | <i>pré</i> endu | anticipation; compétition /t/-/d/ |
| 8. Si vous voulez vraiment savoir | <i>vous le</i> | anticipation |

Lorsqu'un locuteur substitue un mot par un autre, il a tendance à produire un mot *séman- tiquement proche* mais *phonologiquement éloigné* (2 : *doigt-ongle*), ou *phonologiquement proche*, mais *sémaniquement distant* (3 : *liaison* vs. *lésion*). En revanche, on ne rencontre pour ainsi dire jamais d'erreurs dues à une confusion simultanée des deux niveaux.

Des erreurs attestées comme celles ci-dessus révèlent plusieurs choses : les mots sont stockés de façon bi-modale, par le *sens* et par la *forme phonologique* ; le choix des *catégories syntaxiques* (classe lexicale : nom, verbe, adj., etc.) s'effectue avant celui des *lexèmes*, qui eux précèdent la détermination de la forme phonologique. Le *choix des lexèmes* et le *traitement morphologique* (ajout de terminaisons) constituent donc deux phases distinctes, se succédant dans le temps (Butterworth, 1989 ; Levelt et al. 1999). A noter également que les confusions portent généralement sur des mots de la même catégorie lexicale, et qu'en cas d'inversion, les mots confondus prennent la marque du mot source. Ainsi dans l'exemple 4 « *dernier* » (choisi malencontreusement à la place de « *prochain* ») prend la marque du féminin (la semaine *dernière*) et non pas celle du masculin.

Après toutes ces observations on peut se demander si l'on ne pourrait pas exploiter certains de ces faits. Nous présenterons ici deux directions pour aider le locuteur à trouver le mot qu'il a sur le bout de la langue.

Comme déjà indiqué, il y a au moins deux cas de figure : le locuteur trouve un candidat, formellement ou sémantiquement proche du *mot source*. Si l'on veut tenir compte de ce fait, cela veut dire, qu'on peut partir soit de la *forme* linguistique (graphique/phonétique) soit du sens : concepts (définition) ou mots (mot ayant un rapport avec le mot cible).

3.1 Accès par la forme :

Nous présentons deux méthodes, l'une qui part de la forme écrite, l'autre qui part de l'oral, ou qui traite l'écrit comme si c'était de l'oral (pour plus de détails voir Zock & Fournier, 2001)

3.1.1 L'accès par la forme écrite

Notre argument concernant l'accès est donc basé sur l'hypothèse suivante : ce qui nous empêche d'accéder au mot qu'on a sur le bout de la langue c'est une certaine ressemblance formelle entre le mot source (celui qu'on produit) et le mot cible. C'est elle qui nous aveugle, exerçant un effet quasi-hypnotique, nous empêchant d'accéder au mot auquel on pense (mot cible). Etant donné que cette ressemblance formelle est souvent due à une inversion de syllabes ou à l'homophonie, on peut se servir du même mécanisme (en l'inversant) pour (re)-trouver le mot qu'on avait sur le bout de la langue : on permute des phonèmes ou des syllabes, puis on choisit parmi les candidats. Etant donné que le locuteur sait ce qu'il veut dire, il sait reconnaître le mot convenable, pourvu qu'on le lui présente.

L'algorithme est en gros le suivant (Fournier et Letellier, 1990) : on prend en entrée le mot source, erroné certes, mais raisonnablement proche du mot cible (par exemple, "**déléguer**", "**réléguer**" ou "**redegler**" à la place de "**déréglér**"), on permute les éléments le composant (graphèmes/phonèmes), puis on vérifie dans le dictionnaire s'il y a un candidat. S'il n'y a qu'un seul on le présente, sinon (s'il y en a plusieurs) on décide de la priorité suite à un

filtrage sémantique (voir figure-2). S'il n'y a aucun candidat il faut élargir l'espace de recherche en considérant des éléments graphiquement ou phonétiquement proches des éléments du *mot source*. Par exemple, on peut remplacer le /p/ par un /b/ ou un /m/.

<div style="border: 1px solid black; padding: 5px;"> <p>Input: redegler</p> <p>Output: déréglér grêlerez réglerez reléguer préréglé prérègle</p> <p>Number of proposals</p> <p><input type="checkbox"/> Based on spelling: 14</p> <p><input type="checkbox"/> Based on sound: 0</p> <p><input type="checkbox"/> Both: 15</p> </div>	<div style="border: 1px solid black; padding: 5px;"> <p>Input: hatantion</p> <p>Output: attention attentions</p> <p>Number of proposals</p> <p><input type="checkbox"/> Based on spelling: 0</p> <p><input checked="" type="checkbox"/> Based on sound: 2</p> <p><input type="checkbox"/> Both: 0</p> </div>
<p>Figure 2a : Recherche par la <i>forme écrite</i></p>	<p>Figure 2b : Recherche par le <i>son</i></p>

3.1.2 L'accès par le son

Comme il existe des erreurs résultent non seulement d'inversion de phonèmes ou des syllabes, mais également d'une mauvaise transcription phonème-graphème,⁸ nous utilisons un traducteur graphème -> phonème pour produire une clef insensible à la manière de transcrire la sonorité. Si le mot donné est *hippOTHénuse* pour *hypOTénuse*, ou **hatantion** à la place d'**attention** (figure 2b), le système n'aura aucun mal à trouver la bonne réponse, étant donné que le Phonécode de chaque couple est le même. Bien sûr, les deux méthodes peuvent être combinées. Ceci peut s'avérer utile dans le cas d'un lapsus comme *topo* à la place de *poteau*, par exemple. Dans ce cas on convertit les graphèmes en phonèmes (/topo/-/poto/), puis on effectue les différentes combinaisons jusqu'à trouver un candidat.

Il y a un cas où notre méthode ne fonctionne pas. Si jamais l'utilisateur insérait un caractère, produisant de ce fait accidentellement un mot possible, mais exprimant autre chose que ce qu'il a voulu dire (*masson* au lieu de *maison*), le système ne saura le deviner. En fait, il stopperait trop tôt la recherche, considérant qu'il a trouvé le bon candidat : le phonécode suggérant *maçon* qui est effectivement plausible, mais faux en l'occurrence.

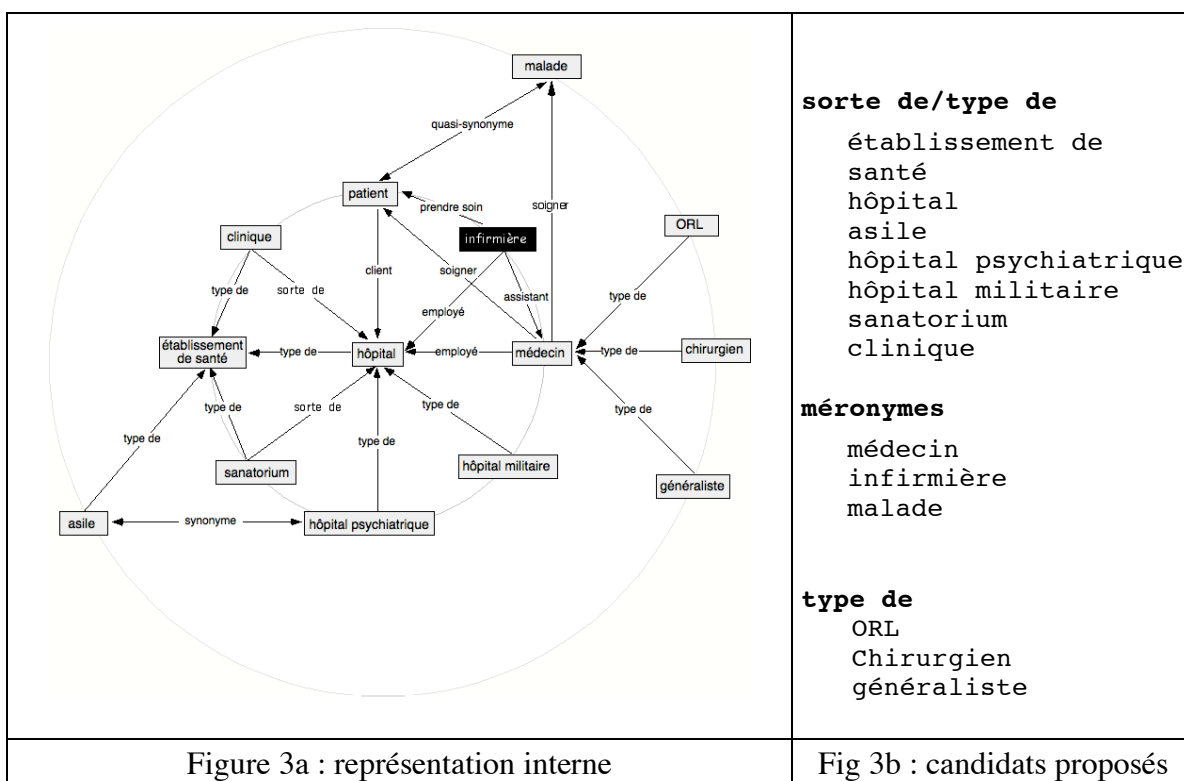
3.2 Accès par le sens : navigation dans un hyperespace conceptuo-lexical

Il est possible que le locuteur n'arrive pas à trouver de candidat lexical. Les seuls éléments sur lesquels il peut s'appuyer sont toute, ou partie d'une *définition*, une *idée* en rapport avec le mot cible (une *association*, donc un autre mot ne faisant pas partie du sens), ou une combinaison de ces facteurs. Nous envisageons plusieurs méthodes. Nous n'esquissions qu'une : la navigation/propagation dans un réseau lexical.

Supposons qu'on cherche le mot *infirmière* (*mot cible*), alors que le seul mot qui nous vienne à l'esprit (*mot source*) soit *hôpital*. Le système est censé, sinon trouver immédiatement le mot recherché (*mot cible*), aider le locuteur à s'en approcher en lui

⁸ C'est le problème de l'homophonie : voir, les nombreuses formes possibles pour transcrire le son /vin/ → vin, vins, vingt, vaincs, vint, etc.

proposant des candidats, dont au moins un aurait un rapport plus ou moins direct avec le mot recherché. Le système construira donc en interne un réseau lexico-sémantique ayant comme noyau le mot *hôpital* (mot *source*) et comme satellite les différents mots ayant un rapport avec celui-ci, par exemple, les *employés/méronymes* (médecin, infirmière), des *sous-types* (clinique, sanatorium), etc. Ce processus est récursif : les satellites peuvent à leur tour servir de nouveau point de départ, et comme le locuteur sait normalement quel concept (ou quel mot) est en rapport avec le mot cible, il devrait tôt au tard le rencontrer. A noter que le graphe est la *représentation interne* du système (figure 3a), l'utilisateur ne voit que des paquets de mots candidats, paquets qui sont créés grâce aux liens connus par le système (figure 3b).



Bien entendu, pour pouvoir fonctionner selon le manière esquissée ici (navigation par association) il faut indexer les mots en conséquent. A cette fin nous envisageons d'utiliser un extracteur de collocation qu'on lancerait sur un vaste corpus, corpus représentant les connaissances du citoyen moyen. Pour plus de détails, voir Zock & Bilac (2004).

L'intuition selon laquelle le dictionnaire mental (d'aucuns préféreraient parler d'encyclopédie) serait un réseau, dont les *noeuds* sont des mots (et/ou des concepts), et les *liens* essentiellement des associations ne date pas d'hier. Pourtant, bien que cette intuition ne soit pas nouvelle et bien qu'elle soit partagée par de nombreux chercheurs⁹, il n'y a à notre connaissance aucun inventaire exhaustif (ou de classification) quant à la nature de ces

⁹ En effet, cette intuition se trouve déjà chez Aristote (« De memoria et reminiscencia »), puis chez des philosophes (Locke, Hume) et physiologistes anglais (James et Stuart Mills), des psychologues (Galton, 1880 ; Freud, 1901 ; Jung & Ricklin, 1906) et des psycholinguistes (Deese, 1965 ; Jenkins, 1970). Enfin, cette idée est sous-jacente à WORDNET (Miller,1990), aux travaux connexionistes (Stemberger, 1985 ; Dell, 1986), aux hypertextes et au web (Bush, 1945 ; Nelson, 1967). Pour des synthèses en psycholinguistique voir (Hörmann, 1972 ; chapitres 6-10), pour des références plus récentes voir (Spitzer, 1999). Enfin, pour des travaux portant sur des réseaux sémantiques, voir (Sowa, 1992).

liens¹⁰. Or, connaître leur nature est l'une des conditions préliminaires pour indexer un tel dictionnaire, et c'est là un de nos objectifs dans les mois à venir.

4 Conclusion

En matière de dictionnaire, deux choses sont fondamentales : quelle information y mettre et comment y accéder. Or, ce deuxième point, généralement laissé de côté, devrait nous éclaircir également sur le contenu des dictionnaires (que mettre dedans), car les deux **sont** liés.

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 accéder à l'information. A l'heure actuelle il y a un fossé énorme entre les dictionnaires papiers, les dictionnaires électroniques et le dictionnaire mental. Pourtant, 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 à l'information. Si les dictionnaires traditionnels sont passifs et assez limités en termes d'accès, les dictionnaires électroniques ont un potentiel considérable, susceptibles de présenter rapidement et sous des formes diverses l'information recherchée. Les méthodes 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).

5 Références

- Aitchison J. Words in the mind: An introduction to the mental lexicon. Oxford: Blackwell. 1987
- Aristote (350 avant JC) De memoria et reminiscentia. In Parva Naturalia, Vrin
- Bateman J. & E. Hovy: Computers and text generation: principles and uses. In C. Butler (Ed.), Computers and texts: an applied perspective. Blackwell. Oxford, 1991
- Bateman, J. and Paris C. L., Phrasing a text in terms the user can understand. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, Detroit, 1989.
- Boissière, P. (1862) Dictionnaire analogique de la langue française : répertoire complet des mots par les idées et des idées par les mots, Paris
- Brown, R and Mc Neill, D. (1966). The *tip of the tongue* phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5, 325-337
- Burke, D.M., D.G. MacKay, J.S. Worthley and E. Wade 1991 «On the Tip of the Tongue: What Causes Word Finding Failures in Young and Older Adults?», *Journal of Memory and Language* 30, 542-579.
- Bush, V (1945) As we may think. *The Atlantic Monthly*; Volume 176, No. 1; pages 101-108
- Butterworth, B. (ed.): Sentence Production. vol.1. Academic Press. New York, 1980
- Butterworth, B. (1989) Lexical Acces in Speech Production. In, W. Marslen-Tayler (Ed.).
- Cahill, L. & M. Reape (1999). Lexicalisation in applied NLG systems. Brighton, ITRI: 9.
- Collins, A. & Quillian, L (1969) Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247
- Cumming, S. (1986). *The Lexicon in Text Generation*, ISI: 86-168.
- Cutler, A. 1982. (Ed) *Slips of the Tongue and Language Production*. Amsterdam: Mouton

¹⁰ A cet égard, le travail de Mel'cuk (1992) est peut-etre encore l'une des meilleurs pistes. Pour d'autres propositions, veuillez consulter Zock & Bilac (2004).

- Danlos L. The linguistic basis of text generation. Cambridge: Cambridge University Press. 1987
- Deese J. (1965) The structure of associations in language and thought. Baltimore
- Deese J. Thought into Speech. Prentice Hall. Englewood Cliffs. NJ., 1984
- Dell, G. S., Chang, F., and Griffin, Z. M. (1999), "Connectionist Models of Language Production: Lexical Access and Grammatical Encoding," *Cognitive Science*, **23/4**, pp. 517-542.
- Elhadad, M. (1992). Using argumentation to control lexical choice: a functional unification implementation. New York, Graduate School of Arts and Science, Columbia University.
- Fournier J-P. & S. Letellier. (1990) PIC: a Parallel Intelligent Corrector. Artificial Intelligence Application & Neural Networks AINN'90, pp 38-41, Zürich
- Freud, S. (1901) Psychopathologie de la vie quotidienne. Paris : Payot, 1997
- Fromkin V. (ed.): Errors in linguistic performance: Slips of the tongue, ear, pen and hand. New York: Academic Press, 1980
- Fromkin V. Speech Production. In J. Berko, Gleason & N. Bernstein Ratner (eds.) Psycholinguistics. Fort Worth, TX: Harcourt, Brace, Jovanovich. 1993
- Galton, F. (1880). Psychometric experiments. *Brain*, 2, 149-162.
- Goldman, N., Conceptual generation. In R. C. Schank (ed.), *Conceptual Information Processing*, New York, American Elsevier, 1975.
- Gross, M., Méthodes en syntaxe. Paris, Hermann, 1975.
- Heid, U. & Raab, S., Collocations in Multilingual Generation, In *Proceedings of the EACL*, Manchester, 1989.
- Horacek, H. (1996). "On expressing metonymic relations in multiple languages." Special Issue on Lexical Choice, *Machine Translation* 11(1-3): 109-158.
- Hörmann H. (1972) Introduction à la psycholinguistique. Paris, Larousse
- Hovy, E., Generating Language with a Phrasal Lexicon. In D. Mc Donald and L. Bolc (eds), *Natural Language Generation Systems*, New York, Springer Verlag, 1988.
- Hovy, E., Pragmatics and natural language generation. *Artificial Intelligence* 43, 1989, pp. 153-97.
- Jacobs, P., Knowledge intensive natural language generation. *Artificial Intelligence*, 33, North Holland, 1987.
- Jenkins, J.J. (1970). The 1952 Minnesota word association norms. In: L. Postman, G. Keppel (eds.): *Norms of Word Association*. New York: Academic Press, 1-38.
- Jordanskaja, L., Kittredge, R. and Polguère, A., Lexical selection and paraphrase in a meaning-text generation model. In Paris, W. Swartout and W. Mann (eds), *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, Dordrecht, Kluwer Academic Publishers, 1991.
- Jung, C.G., Ricklin, F. (1906). Experimentelle Untersuchungen über Assoziationen Gesunder. In: C.G. Jung (ed.): *Diagnostische Assoziationsstudien*. Leipzig: Barth, 7-145.
- Kempen, G. & Huijbers, P. (1983) Lexicalization Process in Sentence Production and Naming: indirect Election of Words, *Cognition*, 14.
- Levelt, W. & Schriefers, H. (1987) Stages of Lexical Access. In, Kempen, G. (Ed.)
- Levelt, W. (1992). "Accessing Words in Speech Production: Stages, Processes and Representations." *Cognition* 42: 1-22.
- Levelt, W. Lexical Access in Speech Production (ed.). Blackwell Publishers. Cambridge, 1993
- Levelt, W. Speaking: From intention to articulation. Cambridge. MA: MIT Press, 1989
- Longman Language Activator (1993) The world's first production dictionary, Longman, London
- Marslen-Taylor, W. (Ed.) (1979) Lexical Representation and Process, Bradford book, MIT Press, Cambridge, Mass.
- McDonald, D.D. (1988) On the Place of Words in the Generation Process, *4th International Workshop on Natural Language Generation*, Catalina Island.
- Mel'cuk, I., A. Clas & A. Polguère (1992) Introduction à la lexicologie explicative et combinatoire, Ed. Duculot, Paris
- Mel'cuk, I. et al., Dictionnaire explicatif et combinatoire du français contemporain. Montréal, Presses de l'université de Montréal, 1991.
- Miller, G.A., ed. (1990). WordNet: An On-Line Lexical Database. *International Journal of Lexicography*, 3(4).
- Nelson, T. (1967) Xanadu Projet hypertextuel, <http://xanadu.com/>

- Nioclov, N. (1999) Approximated text generation from non-hierarchical representation in a declarative framework. Phd dissertation, Edinburgh
- Nirenburg, S. & Nirenburg, I., A framework for lexical selection in natural language generation, *COLING-88*, Budapest, 1988.
- Nogier, J.F. & Zock, M., Lexical choice by pattern matching. In *Knowledge Based Systems*, 5 (3), 1992.
- Novak, H.J. (1987) Strategies for generating coherent descriptions of object motions in time-varying imagery. In G. Kempen (ed.)
- Palermo, D., Jenkins, J. (1964). *Word Association Norms*. Minneapolis, MN: University of Minnesota Press.
- Péchoin, D. (1992) (ed.) *Thésaurus Des idées aux mots, des mots aux idées*, Larousse, Paris
- Polguère, A. (2000). A "Natural" Lexicalization Model for Language Generation. Symposium on Natural Language Processing (SNLP'2000), Chiangmai.
- Reiter, E., A new model for lexical choice for nouns. In Special Issue on Natural Language Generation, *Computational Intelligence*, 7(4), 1991.
- Robert, P. (1993) *Dictionnaire alphabétique et analogique de la Langue Française*, Paris
- Robin, J. (1990) A Survey of Lexical Choice in Natural Language Generation, *Technical Report CUCS 040-90*, Dept. of Computer Science, University of Columbia
- Roelofs, A. (1992). "A spreading-activation theory of lemma retrieval in speaking." Levelt, W. (ed.) Special issue on the lexicon, *Cognition*, 42: 107-142.
- Roget, P. (1852) *Thesaurus of English Words and Phrases*, Longman, London
- Smadja, F. and K. McKeown, K., Using collocations for language generation. In Special Issue on Natural Language Generation, *Computational Intelligence*, 7(4), 1992
- Sowa, J. (1992) "Semantic networks," *Encyclopedia of Artificial Intelligence*, edited by S. C. Shapiro, Wiley, New York
- Stede, M. (1995). "Lexicalization in Natural Language Generation: a survey,." *Artificial Intelligence Review* 8. pp. 309-336
- Spitzer, M. (1999). *The mind within the net : models of learning, thinking and acting*. A Bradford book. MIT Press, Cambridge
- Stemberger, J. P. (1985) "An interactive activation model of language production." In A. W. Ellis [ed] *Progress in the Psychology of Language*, Vol. 1, 143-186. Erlbaum.
- Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, 8, 314-317.
- Wanner, L. (1996). Lexical Choice in Text Generation and Machine Translation. Special Issue on Lexical Choice, Machine Translation. L. W. (ed.). Dordrecht, Kluwer Academic Publishers. 11: 3-35.
- Wanner, L. and Bateman, J., Lexical Cooccurrence Relations in Text Generation. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, Cambridge, Mass., 1990.
- Wanner, L. and Maier, E., Lexical choice as an integrated component of situated text planning. In *Proceedings of the Third European Workshop on Natural Language Generation*, Judenstein, Austria, 1991.
- Ward, N. (1988). *Issues in Word Choice*. COLING-88, Budapest.
- Zock, M. 1996. "The Power of Words in Message Planning". *16th International Conference on Computational Linguistics*, Copenhagen
- Zock, M. & J.P. Fournier (2001) How can computers help the writer/speaker experiencing the Tip of the Tongue Problem? Proc. of RANLP, Tzigov Chark, Bulgarie, pp. 300-302
- Zock, M. & S. Bilac (2004). *Word lookup on the basis of associations : from an idea to a roadmap*. Proc. of Coling workshop : Enhancing and using dictionaries, Geneva, pp. 29-35