

Si tous les chemins mènent à Rome, ils ne se valent pas tous.

Le problème d'accès lexical en production

Zock, Michael

Laboratoire d'Informatique Fondamentale (LIF),

groupe TALEP

CNRS & Aix-Marseille Université

michael.zock@lif.univ-mrs.fr

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 cognitif savent énormément de choses concernant le mot recherché (sens, nombre de syllabes, origine, etc.), et que les mots avec lequel ils le confondent lui ressemblent étrangement (lettre ou son initial, catégorie syntaxique, champ sémantique, etc.).

Mon objectif (à long terme) 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. À cette fin, je prévois d'ajouter à un dictionnaire électronique existant un index d'association (collocations rencontrées dans un grand corpus). Autrement dit, je propose 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.

Ma démarche est fondée sur plusieurs hypothèses.

1° Les stratégies de recherche dans notre dictionnaire mental dépendent, bien entendu, de la représentation des mots dans notre cerveau. Hélas, on n'a toujours pas une carte précise de cette organisation. Quant à la recherche on pourrait dire qu'elle s'opère essentiellement sur deux axes :

- Le premier décrit le passage des idées à leurs expression (idées, forme, sons). Cette vision représente l'ordre naturel des 'choses' : partant du *sens* on va vers l'*expression* (forme sonore ou graphique du mot) en passant par les *concepts lexicaux* (lemmes dans la théorie de Levelt).
- Le deuxième axe est plus proche de ce qu'on peut considérer comme une forme d'organisation de mots. Il représente leur usage (fréquent/typique) dans le discours. C'est un graphe de co-occurrences ou d'associations.

Il y a donc deux idées complémentaires : (a) l'expression des idées au sens restreint (passage des concepts aux mots) et (b) les rôles que ces idées (concepts/mots) peuvent jouer dans le cadre d'une phrase (discours, contextes possibles des mots). Ce contexte précise d'ailleurs souvent le sens des mots.

Si le 1er axe représente la voie naturelle en production, voie empruntée pratiquement en toutes circonstances (plan A), le 2ème axe (voie associative) est la voie de rechange (plan-B), utilisée en cas d'échec du plan A. Le premier processus est automatique (rapide et inconscient), tandis que le second est contrôlé, donc lent et accessible à notre conscience. C'est lui qui m'intéresse,

car il reflète la situation dans laquelle un auteur se trouve lorsqu'il fait appel à un dictionnaire ou thésaurus.

2° Le dictionnaire mental est un vaste réseau dont les *noeuds* sont des concepts ou mots (lemmes ou expressions) et les *liens* essentiellement des *associations*. Etant donné que tout est lié, tout peut être trouvé, du moins en principe : il suffit de suivre des bons liens. Chercher un mot consisterait donc d'entrer dans ce réseau, puis de suivre les liens pour (re)trouver le terme faisant obstruction.

3° Le dictionnaire mental est à la fois un dictionnaire et une encyclopédie. Etant donné que les mots sont utilisés pour coder des connaissances du monde, ces dernières peuvent être sollicitées pour nous aider à retrouver le mot recherché (ainsi le terme 'baguette' pourrait-il être obtenu à partir de 'restaurant chinois' ou à partir de 'type de couvert'). Tout nous fera penser à quelque chose, tout est associé à quelque chose. De ce fait, tout est susceptible d'être évoqué par un terme lié, fût-il indirect (chaîne associative; recherche à plusieurs pas).

4° Les informations permettant d'effectuer ce type de navigation (*atlas sémantique*) se trouvent non seulement dans notre cerveau, mais aussi dans nos productions (manifestation linguistiques : phrases, textes). Comme ces traces constituent une forme d'extériorisation de l'*organisation* des idées (concepts/mots) dans notre cerveau, on peut s'en servir pour créer un modèle analogue. Ceci donnera un atlas ou une *carte sémantique* permettant alors aux auteurs de s'orienter pour trouver la forme du mot qui leur est (momentanément) inaccessible.

Voici mon ambition. L'objectif de cet exposé est de montrer comment construire une telle ressource, comment s'en servir, quelles sont les difficultés de construction et quelles sont les possibilités qu'offre un telle ressource.