

Structuration de l'analyse sémique

En dépit des notions de *dimension*, de *domaine* et de *taxème*, le grand problème de l'analyse sémique paraît être de disposer de règles, de repères permettant de construire les sémèmes de façon à peu près homogène et systématique.

Robert Martin a bien montré le caractère souvent approximatif de l'analyse sémique.

Les approximations de l'analyse sémique selon Robert Martin

Une analyse des définitions données dans les dictionnaires fait apparaître une multiplicité de techniques utilisées pour donner le sens des mots, différences de techniques qui peuvent révéler des différences de contenu.

Robert Martin distingue ainsi deux grandes catégories de définitions : les *définitions métalinguistiques*, qui consistent dans un commentaire sur le signe linguistique, (ex.: définitions commençant par des formules du type "se dit de...", "marque...", "exprime...", "En parlant de...") et les *définitions paraphrastiques* qui portent sur les contenus, sur les désignations, qui constituent la grande majorité des définitions.

Les définitions paraphrastiques, contrairement aux définitions métalinguistiques, permettent sans trop de difficulté la substitution dans la phrase du mot de référence par sa paraphrase. Néanmoins, la paraphrase dissimule des formes logiques variées. Ainsi :

- *l'hyponymie* définit un terme par rapport à un autre de niveau de généralité plus élevé, tel que ce qui est vrai pour le plus général est également vrai pour le plus spécifique. On appelle hyperonyme d'un vocable *D* un vocable *d* tel que la substitution de *d* à *D* dans *p*, sans autre modification, conduit à une phrase *q* telle que *p* implique *q* ($p \Rightarrow q$). Ainsi le verbe *aguicher* se définit par rapport au verbe *provoquer*.
- La *synonymie* définit un terme par un autre dont la signification est strictement ou presque identique.
- *L'antonymie* définit un terme par son contraire.
- La *conjonction* ou *intersection* définit un terme à partir de deux autres termes, ou plus. *Voler* résulte à la fois du fait de se soutenir dans l'air au moyen d'ailes et d'y effectuer un déplacement. De même *circuler* est équivalent à *aller et venir*, la notion d'orientation étant absente de la définition du verbe *circuler*.
- La *définition comme partie* (et non comme élément) d'un ensemble. Ex. *bras* comme partie du corps, *manche* comme partie d'un habit,...
- La *définition par énumération*, que l'on pourrait qualifier d'extensionnelle. En fait, il s'agit d'un type de définition par hyponymie consistant énumérer les termes se

trouvant dans la dépendance de l'hyperonyme. Ainsi les membres du corps humain comportent entre autres les pieds et les mains. Et les pieds et les mains sont des membres du corps humain.

- La *définition dérivationnelle* se fonde sur l'analyse morphologique. Ainsi la *jovialité* est dérivée par nominalisation de l'adjectif *jovial*. L'adjectif *juridictionnel* est dérivé du nom *juridiction*.

Nous ne sommes pas certain que l'inventaire de Robert Martin soit tout à fait exhaustif. On pourrait y ajouter une distinction entre des définitions à caractère descriptif telles que toutes celles qui viennent d'être énoncées et des *définitions fonctionnelles*, définissant un objet non par ses caractéristiques mais par son usage. Un *tournevis* sera ainsi un outil servant à visser une vis dans un matériau, ce qui permet d'englober dans une même catégorie le tournevis manuel et le tournevis électrique.

De cette profusion de techniques ou d'usages définitoires, on peut déduire assurément que les traits définitoires tels que l'on pourrait les extraire des définitions des dictionnaires ne permettent pas de construire directement un lexique structural assis sur des traits sémantiques stables et universellement admis.

Faut-il partager le jugement que portait en 1972 (p. 184) Georges Mounin sur l'état de la recherche sémantique?

"On peut affirmer que le point qui reste le plus faible est l'analyse du signifié de l'unité lexicale minimale : *monème* (ou mot dans certaines théories). Tout le monde convient, après Hjelmslev, que cette analyse est la condition première de toute construction d'un dictionnaire structural du lexique, ou de toute structuration du lexique sous une autre forme. Mais les "atomes opérationnels sémantiques" de Ceccato, les marqueurs et les différenciateurs sémantiques de Katz et Fodor, les "unités minimales de sens" des sémanticiens soviétiques ne sont rien d'autre et rien de plus que les traits sémantiquement distinctifs de Bloomfield à Prieto, que les figures de contenu de Hjelmslev. Ces atomes de contenu ne dépassent guère le genre prochain et la différence spécifique des définitions d'Aristote (jument =cheval + femelle, etc.); et surtout les procédures esquissées pour les extraire n'ont guère vraiment fait de progrès linguistiquement parlant. Or, ce sont des critères linguistiques qu'il faudrait trouver pour cette analyse de base."

Si l'on définit *l'analyticité* comme étant la propriété d'un énoncé exprimant une relation qui est vraie pour tout locuteur, indépendamment de la situation, en tout lieu et en tout temps, pour R. Martin, l'analyticité n'est pas strictement déterminable. "La définition, du fait de la diversité de ses formes et la variabilité de ses contenus, confère à l'analyticité un caractère imprécis; cela revient à dire que les conditions de vérité constituent des sous-ensemble flous, que les signes linguistiques sont le lieu de propriétés inégalement pertinentes,..."

Ainsi, même si l'on parvient à mettre un peu d'ordre dans la diversité des formes définitoires recensées, ce que nous croyons possible, il y a une part d'arbitraire dans le fait de retenir un contenu plutôt qu'un autre pour un terme donné. Le choix n'est sans doute pas illégitime, et les auteurs de dictionnaires sont bien obligés de le faire. Sans que soit mis en question la légitimité des dictionnaires, la pertinence des choix effectués sera toujours relative.

Entre autres sources du flou du sens du signe linguistique, il y a les hésitations des linguistes eux-mêmes à identifier de manière certaine les oppositions entre vocables, et Martin de citer l'exemple de *émeute*, *soulèvement* et *insurrection*. "En général l'émeute paraît plus spontanée, plus localisée¹. Mais ailleurs elle est vue sans plus comme un "soulèvement populaire"². Ici *insurrection* et *émeute* sont définies par *soulèvement*³. Ailleurs, le soulèvement apparaît comme un "début de révolte"⁴,...

Autre difficulté sur le chemin de la construction d'un dictionnaire sémique ou structural : le fait que les dictionnaires sont, par structure, voués dans les définitions qu'ils donnent à la *circularité*. Comment pourrait-il en être autrement puisque les termes du dictionnaire sont définis par d'autres termes qui sont eux-mêmes définis dans le dictionnaire à l'aide des autres termes du dictionnaire ?

Seul le choix de vocables primitifs, donnés a priori, permettrait de sortir du cercle. Ainsi en partant de l'exemple des mots les plus généraux se rapportant à la perception du son : entendre, ouïe, audible, son, auditif, acoustique, oreille, R. Martin propose de ne retenir comme termes primitifs ou noèmes que *perception*, *capacité*, *organe ou partie du corps*, *son* et *non harmonique*, à partir desquels les autres peuvent être définis, le noème *son* occupant une place centrale dans le dispositif. Par le choix du noème *son*, dit R. Martin, "le graphe à boucle est devenu arborescence. La voie s'ouvre ainsi à une axiomatisation lexicale." (1992 p. 98)

Toutefois R. Martin émet des réserves sur la légitimité de la traduction des noèmes par des vocables du lexique. En effet, le noème *son* ne correspond pas au mot *son*. Dans tous ses emplois en langue, le mot *son* va prendre des significations différentes du noème. En changeant de sens en fonction de son contexte d'énonciation, le mot *son* cesse d'être un noème. "Dans tous ses emplois le mot *son* a une valeur résultative et suppose une médiation instrumentale : ainsi le cristal n'a pas de son en soi; en parlant du son du cristal, on suppose qu'il a été heurté (légèrement!); c'est ce heurt qui le rend sonore. On fait du bruit, on ne fait pas de son. Mais quelque chose rend un son ou on en tire un son. Si donc les noèmes étaient (au moins partiellement) des universaux, de toute manière il n'en irait pas de même des vocables dont, par facilité, on leur attribue la forme, et dont les signifiés sont toujours plus complexes."(opus cit.)

B. Pottier rejoint tout à fait l'observation de R. Martin. En effet, pour B. Pottier, le noème n'appartient pas au domaine de la linguistique mais relève du niveau conceptuel.

"Le noème est un trait de sens posé indépendamment de toute langue naturelle. Il est absolu (et non relatif à un ensemble) et son existence est décidée par l'analyste."

"La *noémique* est l'étude de l'ensemble des éléments conceptuels et de leurs relations, considérée comme un instrument d'analyse nécessaire pour décrire le fonctionnement de base de la sémantique des langues naturelles." (Pottier 1992 p. 67-68).

F. Rastier pose à son tour le problème en des termes presque identiques :

¹ Littré

² DFC

³ Robert

⁴ Littré

"Le caractère circulaire des définitions ne serait regrettable que si l'on voulait constituer la sémantique des langues en une axiomatique qui briserait cette circularité, en négligeant que la relation de définition instaure une équivalence (de modulo conventionnel) mais non une identité." (1994 p. 50)

Nous ne pouvons qu'être d'accord avec F. Rastier en ce que le processus de communication ne nécessite aucunement la construction d'un échafaudage conceptuel supérieur au niveau linguistique, le sens n'existant pas en soi, mais résultant, selon la tradition saussurienne, de la confrontation des signes linguistiques.

Par contre, et nous espérons le démontrer, le traitement informatique impose de sortir de la circularité et de poser les fondements conceptuels de la structuration du lexique. Toute compréhension automatique implique que l'on puisse établir l'équivalence de sens entre des formulations formellement différentes, mais identiques ou équivalentes par leur contenu. Or, il est évident que deux termes synonymes qui seraient définis l'un par l'autre ne peuvent avoir le même sémème, ce qui rend impossible la constatation de leur identité ou de leur équivalence même relative.

Nous rejoignons une préoccupation exprimée vingt ans plus tôt par Georges Mounin lorsqu'il invitait à poursuivre la recherche en sémantique structurale :

"Tenter de déceler la véritable structuration du lexique, et celle des signifiés d'une langue, est donc une entreprise raisonnable; et raisonnable aussi parce que tout, depuis l'apprentissage du lexique par l'enfant jusqu'à l'analyse saussurienne des valeurs, nous suggère que les mots ne sont pas des unités isolées. Mais on peut se demander légitimement si les tentatives structuralistes connues jusqu'ici ont abouti à fournir des résultats aussi solidement acquis que dans les autres domaines linguistiques - et répondre que la structuration du lexique, et moins encore celle de la sémantique, n'ont pas livré leur secret.

"On a cherché ici à vérifier si le lexique d'une langue possède une structuration discernable dans la mesure où il décalquerait une structuration d'un autre ordre, introduite par la praxis humaine dans l'expérience qu'elle se fait du monde non linguistique. Le lexique alors ne serait pas structurable en vertu de ses propriétés strictement linguistiques, pour des raisons propres à la linguistique (sauf dans la zone rarement complète des séries dérivationnelles), mais uniquement parce qu'il serait le reflet plus ou moins exact d'une autre, ou de beaucoup d'autres structures non linguistiques."(1972, p. 161)

Il semble donc qu'il y ait un double accord : sur l'impossibilité d'un système axiomatique unique; et sur l'utilité de concevoir des systèmes noématiques.

"Dans la langue tout est mouvance, tout est flou. Et, du fait même, les axiomatisations possibles sont en grand nombre - aussi artificielles les unes que les autres, ce qui ne signifie d'aucune façon qu'elles soient dénuées d'intérêt, car chacune éclaire le réel sous un certain angle. Mais aucune n'en reflète la véritable complexité." (R. Martin 1992, p. 99)

Cette remarque laisse entière la question de la méthode pour structurer le lexique en évitant un relativisme débridé.

R. Martin nous ouvre plusieurs voies.

La première concerne la mise au jour d'universaux d'expérience. R. Martin se garde bien de parler d'universaux tout court. Il observe à juste titre que certaines notions peuvent apparaître a priori comme des universaux, mais que rien en fait n'en garantit l'universalité. Ainsi la notion de *perception* correspond plus à une abstraction qu'à une expérience élémentaire. On perçoit un son, une image, une odeur, un saveur, mais rien ne dit que la conscience de la perception, notion abstraite, soit elle-même universelle.

D'où l'idée des *universaux d'expérience*. "Primitifs d'une autre nature, ils ne visent pas, en tant que tels, l'axiomatisation d'un système sémantique. Ils viennent plutôt de l'idée que certaines données du monde, physiques, physiologiques, anthropo-culturelles, exercent sur la vie des hommes une si forte contrainte qu'il est impensable qu'elles ne laissent aucune trace dans la langue. Et ces traces, du fait même, ont toute chance d'être des universaux." (1992, p. 101)

Ainsi, comme exemple, en prenant *entendre* comme primitif du sous-système de la perception des sons, R. Martin obtient les définitions suivantes qui constituent une arborescence échappant à la circularité :

audible : que l'on peut entendre

son : ce qui est audible

acoustique : qui se rapporte au son

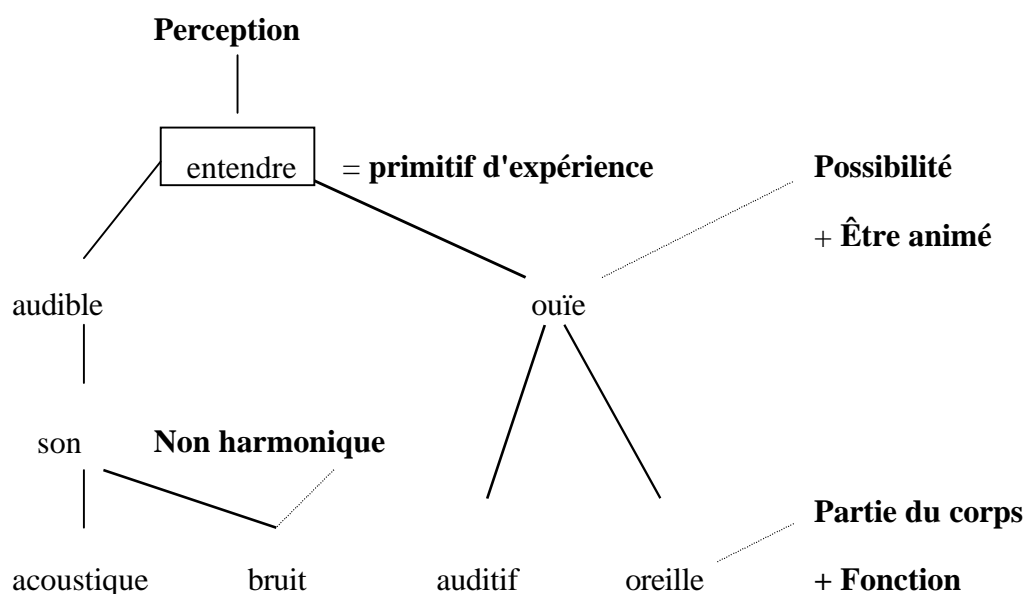
bruit : son + non harmonique

ouïe : capacité (possibilité + être animé) d'entendre

oreille : organe (partie du corps + fonction) de l'ouïe

auditif : qui se rapporte à l'ouïe

La représentation schématique est la suivante :



Ce résultat tangible appelle diverses questions.

Notons d'abord qu'il n'est pas exhaustif. On pourrait ajouter *écouter, auditoire, audition, auditorium, écoute, ...* Mais cela n'altère pas le principe de la méthode employée.

On sera attentif en second lieu à la nature des relations qui unissent les divers éléments ainsi identifiés.

La relation hyperonymique ou hyponymique est très minoritaire. On la trouve entre *son* et *bruit*, indirectement entre *oreille* et *organe*, mais précisément *organe* est étranger au sous-système de perception des sons. On la trouverait entre *entendre* et *écouter* si *écouter* était inclus dans le schéma.

La relation *partie de* est absente.

Les dérivés syntaxiques sont bien représentés : *ouïe, auditif, audible*, etc. Mais les dérivés syntaxiques sont porteurs de relations spécifiques que l'on peut retrouver entre deux vocables sans qu'il y ait dérivation. Ainsi, entre *ouïe* et *auditif* nous avons la relation "se rapporte à", que l'on retrouve entre *son* et *acoustique* alors qu'il n'y a pas entre ces deux termes dérivation syntaxique.

Nous en déduisons que la dérivation syntaxique n'a pas sa place dans la description d'un sous-système sémantique. Par contre, on pourrait utilement rechercher une classification des types de relations qui sont établies entre les divers éléments d'un sous-système sémantique.

Une autre question que l'on peut se poser est de savoir s'il y a équivalence entre le sous-système sémantique ainsi identifié et la notion de taxème telle que B. Pottier et F. Rastier se sont attachés à la définir. Nous reviendrons sur cette question plus loin.

Par ailleurs, au-delà de la question de savoir comment échapper à la circularité, se pose la question de savoir comment il est possible de limiter le nombre de traits sémantiques

entrant dans la constitution des sémèmes. R. Martin nous met sur la voie par une analyse extrêmement fine des contenus définitionnels.

R. Martin opère une première grande distinction entre les *définitions conventionnelles* et les *définitions naturelles*. (1992, p. 67 et s.)

Les définitions naturelles concernent les objets naturels que sont les vocables du langage ordinaire.

Les définitions conventionnelles sont des définitions a priori ou a posteriori.

Les définitions conventionnelles *a priori* déterminent, au moment même de la dénomination d'un objet, les caractéristiques qu'on lui assigne. Les définitions mathématiques, logiques, métalinguistiques sont de ce type.

Les définitions conventionnelles *a posteriori* délimitent conventionnellement le sens, par nature vague, des mots du langage ordinaire lorsque ceux-ci sont voués à un usage technique. Les définitions juridiques et plus généralement les définitions normatives sont de ce type.

Qu'elles soient a priori ou a posteriori, les définitions conventionnelles relèvent d'une activité prescriptive. Elles échappent au jugement de vérité dans la mesure où elles sont vraies par définition. On dit qu'elles sont "analytiquement vraies", c'est-à-dire vraies en vertu de leur sens. Elles échappent aussi à l'évolution du temps, et restent vraies tant qu'elles n'ont pas été remplacées par d'autres définitions. Elles sont dans leur formulation strictement limitées au sens qu'elles posent, aux nécessités de la définition.

A l'inverse, la *définition naturelle* vise à saisir le contenu des mots. Elle a une vocation descriptive et non prescriptive ou stipulatoire. Elle est plus ou moins juste et évolue avec l'objet qu'elle essaie de cerner. Elle est plus ou moins détaillée dans les propriétés qu'elle recense.

A cet égard, R. Martin distingue la *définition minimale* qui ne retient que les propriétés qui sont strictement nécessaires à l'identification de l'objet désigné, à la pertinence des traits recensés, et la *définition stéréotypique* qui vise à donner au-delà de la pertinence linguistique, une représentation de l'objet dénommé suffisante pour en permettre l'identification effective.

Il est évident que pour les besoins d'un dictionnaire destiné au grand public, la définition stéréotypique est légitime, dans une démarche de modélisation des connaissances, la définition minimale devra être préférée.

Pour passer de la définition stéréotypique à la définition minimale, R. Martin, analyse les différents types de propriétés que l'on rencontre dans les définitions stéréotypiques :

- les propriétés universelles que sont satisfaites par tous les objets dénommés;
- les propriétés généralement vérifiés, c'est-à-dire satisfaites par la plupart des objets dénommés;
- les propriétés de nature symbolique, plus ou moins conventionnellement attachées aux objets dénommés.

- parmi les propriétés universelles, certaines sont des propriétés typiques, c'est-à-dire distinctives ou discriminatoires, et les propriétés universelles mais non typiques, c'est-à-dire satisfaites également par d'autres types d'objet.
- Par ailleurs, parmi les propriétés universelles, typiques ou non typiques, certaines sont reconnues par la plus grande généralité des locuteurs, d'autres ne le sont que par des publics plus restreints. Ce sont des propriétés *encyclopédiques*. Par exemple, pour un oiseau, on citera des propriétés comme celles-ci : que c'est un vertébré à sang chaud, à respiration pulmonaire, qu'il possède un jabot et un gésier et non un estomac, etc.

La sélection des propriétés à retenir à fin de modélisation des connaissances dépendra évidemment du but recherché. Le contexte peut imposer par exemple de sélectionner les propriétés universelles typiques et encyclopédiques ou un sous-ensemble de ces propriétés.

Toutefois, dans un traitement linguistique qui vise essentiellement à dégager le sens des énoncés et à comparer des énoncés entre eux, seuls sont à considérer les traits qui sont strictement indispensables à la correcte identification des lexies. On devrait donc se limiter aux propriétés universelles typiques. Ce qui n'enlève rien à l'intérêt d'une identification des savoirs encyclopédiques relatifs à un objet ou à un concept donné. Nous verrons au contraire, que ces savoirs encyclopédiques jouent un rôle déterminant dans les mécanismes de la compréhension, car ils sont à la base de la compétence des locuteurs. Dans le cas des textes normatifs, il s'agit de la compétence de celui auquel s'adresse le texte normatif, qu'il s'agisse du sujet de droit à qui la norme s'impose ou qu'il s'agisse des autorités tenues de veiller à son application ou encore des juridictions amenées à dire le droit, c'est-à-dire à donner l'interprétation « exacte ». Sujets de droit, autorités, et juridictions, chacun à son niveau interprète la norme en fonction de sa propre compétence, or cette compétence est directement déterminée par ce que Pottier appelle « l'environnement du message », lequel ne saurait se réduire aux définitions du dictionnaire.

Nous pensons qu'il existe un niveau de traitement où il est nécessaire de s'en tenir aux définitions minimales. Nous nous attachons donc pour l'instant aux moyens d'établir des définitions minimales nécessaires à une reconnaissance non ambiguë des termes du texte.

Pour constructive qu'elle soit, l'approche de R. Martin ne donne pas toutes les clés permettant de construire les définitions minimales c'est-à-dire les sémèmes.

Pottier (TAL, 1992) ouvre un certain nombre de pistes à travers les modèles sémantiques, les noèmes et nous le verrons plus loin, la théorie des voix.

Pour progresser dans cette voie, il est indispensable de s'arrêter préalablement sur la question de la catégorisation. Elle occupe une place centrale en psychologie des connaissances, mais aussi en linguistique au travers des théories qui ont impulsé des recherches importantes dans les années soixante dix et quatre-vingt, en particulier la théorie du prototype qui a profondément évolué au cours de la période au point que G. Kleiber (1990) a pu parler de théorie standard et de théorie étendue.