

Texte publié dans *Revue d'intelligence artificielle*, Volume 17, n°5-6/2003, Paris: Lavoisier, 909-921.

Distance sémantique et degrés d'analogie dans les matrices analytiques définitoires

Amr Helmy IBRAHIM

Professeur de linguistique à l'Université de Franche-Comté
5, rue Louis Léon Lepoutre 94130 Nogent-sur-Marne
amr.ibrahim1@libertysurf.fr

RÉSUMÉ. Partant d'une comparaison entre le premier grand dictionnaire français de synonymes de Lafaye qui distingue les entrées lexicales sémantiquement proches en opposant des co-occurrences différentes au sein de phrases où ces entrées sont habituellement produites et se pose ainsi en précurseur des analyses du Lexique-grammaire initiées par Maurice Gross et le premier dictionnaire analogique de la langue française de Boissière qui permet de trouver des mots inconnus mais en écrasant les différences grammaticales et sémantiques des entrées lexicalement proches, l'auteur propose de calculer le degré d'analogie entre deux entrées par le biais d'une description analytique définitoire se référant à la matrice dont sont issus les énoncés définitoires des deux entrées comparées.

ABSTRACT. Starting with a comparison between the first big French thesaurus of Lafaye which distinguishes between two lexically close items by opposing different co-occurrences within the sentences in which these items are usually uttered, thus pioneering the Lexicon-grammar analyses initiated by Maurice Gross, and the first big French analogical dictionary of Boissière which indeed helps finding unknown words but with the cost of squeezing grammatical and semantic differences between lexically close items, the author suggests a calculus of the degree of analogy between two items through an analytic and defining description referring to the matrix from which stemmed the defining matrixes of the two compared items.

MOTS-CLÉS : analogie – Lexique-grammaire – matrices analytiques définitoires – verbes supports – synonymes.

KEYWORDS : analytic defining matrixes - analogy – Lexicon-grammar – light verbs – synonyms.

Dédicant aux *Messieurs de l'Académie Française* l'édition de 1857 de son *Dictionnaire des synonymes de la langue française* auquel les immortels ont décerné une première fois en 1853 le *Prix de linguistique*, Pierre-Benjamin Lafaye leur écrit: "*Il restait après vous quelque chose à faire pour l'exacte intelligence et l'emploi éclairé de notre langue classique: à cette belle littérature qui fait l'orgueil de la France et l'admiration du monde il manquait un livre consacré à la distinction des termes en apparence équivalents, livre méthodiquement conçu et fait d'une seule main, dans lequel tous les travaux partiels du même genre fussent non pas seulement rassemblés, mais fondus en un tout. J'ai mis mon ambition, j'ai employé ma vie presque entière à remplir cette lacune*".

L'édition de 1858 sera à nouveau primée et le *Dictionnaire des synonymes* de Lafaye sera longtemps réédité, revu, corrigé et augmenté jusqu'à atteindre dans les dernières éditions de Louis Hachette, par exemple celle de 1865, 1531 pages de petits caractères.

Les 83 pages d'introduction de cette édition¹ nous expliquent en fait qu'il n'existe pas de *synonymes* mais seulement différentes formes d'*analogies* dont la conscience et la description systématique avec ce que cela suppose de typologies et de classements ouvrent la voie à une *définition* authentique du sens des mots. En effet, si Lafaye juge nécessaire d'adopter une nouvelle méthode dans l'élaboration des dictionnaires de synonymes c'est qu'aussi bien les dictionnaires courants que les anciens dictionnaires de synonymes n'apprennent rien, selon lui, à qui les consulte puisque leurs définitions et leurs équivalences reposent toutes sur un double mirage à savoir qu'un mot puisse servir de définition à un autre et que deux mots puissent être sémantiquement identiques. "*Posant en principe qu'il ne saurait y avoir de synonymes parfaits, surtout dans la langue usuelle d'un peuple avancé en civilisation*" Lafaye estime qu' "*il n'y a jamais identité de signification entre les mots réputés synonymes*" mais "*le même rapport que les variétés d'une même couleur principale*" et précise: "*Au premier coup d'œil et à distance, ils semblent tous se confondre, tant les nuances qui les séparent sont légères. Mais, en y regardant de près, on aperçoit ce qu'il y a de particulier dans chacune de ces nuances, ou, pour parler sans figure, on s'aperçoit que chaque mot est marqué de traits distinctifs qui le rendent seul propre à exprimer dans certaines circonstances l'idée générale qu'ils représentent tous*". Un peu plus loin il souligne: "*Il ne suffit pas de définitions irréprochables pour mettre en état de discerner toujours et sûrement la propriété des termes; il faut de plus en rapprochant les définitions de ceux dont le sens se touche, faire ressortir leurs nuances distinctives, et pour cela ce n'est pas trop la plupart du temps d'une longue comparaison où on les*

¹ 1865, Paris: Librairie de L. Hachette et C^{ie}, Boulevard Saint-Germain, n° 77 (Deuxième édition, revue, corrigée et suivie d'un supplément).

oppose les uns aux autres sous toutes les faces, au moyen de phrases faites à dessein ou d'exemples empruntés aux écrivains les plus considérables" (X à XII).

Qu'on en juge sur pièces. Pour distinguer *se vêtir* de *s'habiller* Lafaye cite ce passage du *Cours de littérature ancienne et moderne* (1799) de Jean-François de La Harpe : "*Par un progrès ultérieur et révolutionnaire, les femmes en sont venues à s'habiller sans se vêtir, grâce aux tissus légers qui, en dessinant les formes de leur sexe, ne refusent aux yeux que la nudité absolue*". Dans son introduction il commente : "...à l'époque de la Révolution l'impudence des mœurs fut telle, que les femmes en vinrent à s'habiller sans se vêtir; expression admirable, mais dont la justesse parfaite doit échapper à bien des lecteurs, à tous ceux qui s'en rapportent aux dictionnaires pour les sens des mots: les dictionnaires définissent s'habiller par se vêtir, et se vêtir par s'habiller". A l'article *Vêtu, revêtu, habillé, affublé, fagoté* (p. 1023) il écrit notamment: "*On est vêtu de ce qu'on porte d'ordinaire pour le besoin et la commodité, on est revêtu de ce qu'on porte par-dessus le vêtement comme un insigne, une marque d'honneur ou de dignité (...)* Habillé, du latin *habitus extérieur*, a rapport à la forme, à l'air, à la manière dont on est mis, au point de vue du goût, de la mode, ou relativement à certaines circonstances (...). Vêtu ou revêtu, **on est couvert**, habillé, **on est ajusté ou mis de telle façon**". (...) *Un homme bien vêtu, c'est-à-dire commodément, de façon à braver les intempéries, peut être mal habillé, c'est-à-dire sans élégance, comme on ne s'habille plus ou comme on ne doit pas s'habiller dans le cas où il se trouve. (...) Il y a des personnes toujours mal habillées faute de soin ou de goût pour la toilette*"².

Autrement dit Lafaye rapporte la différence, la distance sémantique entre "*se vêtir*" et "*s'habiller*" au fait que l'on porte un vêtement mais que l'on se met un habit. Que dans un cas ce que l'on porte nous couvre ou nous recouvre alors que ce que l'on se met détermine notre mise, la manière dont nous sommes habillés plutôt que la fonction de ce dont nous nous vêtons. Et il est vrai que sans cette opposition, notre *Je n'ai rien à me mettre*³ n'aurait aucun sens, pas plus d'ailleurs que le *comme on ne s'habille plus* de Lafaye. Ils correspondent en effet très exactement pour le premier à *Je n'ai rien qui corresponde à ma mise habituelle, à la manière dont je me vois mis* et pour le second à *il y a une mise qui n'est plus de mise, une manière de se mettre des habits, de s'habiller qui ne correspond plus à aucune des manières connues de s'en mettre*.

C'est donc finalement au choix du verbe support⁴ actualisateur *porter* ou *mettre* qu'est corrélée la différence fondamentale entre le fait d'être *vêtu* ou d'être *habillé*. Mais

² (Caractères gras et mots soulignés sont de nous)

³ Lafaye ne donne pas cette expression.

⁴ *porter* ou *mettre* sont des verbes supports (V_{-sup}) dans des constructions du type (1) *Les vêtements qu'elle porte sont sales* ou (2) *Le papillon qu'il a mis ne lui va pas* dans la mesure où (1) est synonyme de (1') *Ses vêtements sont sales* et (2) de (2') *Son papillon ne lui va pas*. Autrement dit ces 2 verbes sont des V_{-sup} en (1) et (2) parce qu'ils y sont effaçables et n'y apparaissent que comme actualisateurs des prédicats nominaux *vêtements* et *papillon*. Pour un approfondissement de la problématique et une analyse plus complète des propriétés de ces verbes cf *Langages* 121, *Les supports* ainsi que Ibrahim 99, 00a, 00b, 01, 03a.

cette différence, cette *distance sémantique* est-elle mesurable? Elle devrait l'être si l'on parvient à la corréler à des variables précisément assignables et identifiables de la définition de chacun des deux termes. Ce qui n'est possible qu'à la condition expresse qu'il existe une grammaire régulière des définitions. C'est cette grammaire que manifestent les matrices analytiques définitoires que nous présenterons tout à l'heure.

A ce point de notre exposé on pourrait tout naturellement nous interrompre pour nous demander pourquoi si notre idée était déjà en germe dans une œuvre dont la première édition remonte à 1853, qui a eu pendant près de vingt ans une très large diffusion et obtenu une reconnaissance institutionnelle certaine, elle n'a pas suivi plus tôt le chemin que nous traçons. Et il serait légitime de penser qu'à défaut d'une réponse convaincante c'est soit que l'idée était mauvaise ou que le chemin ne mène nulle part.

Il y a en réalité à cela deux raisons de nature très différente.

La première, qui est celle qui peut venir le plus immédiatement à l'esprit d'un linguiste contemporain est qu'il n'existait pas à l'époque et qu'il n'existera pas pendant près d'un siècle de véritable travail sur les synonymies phrastiques, sur les équivalences possibles entre des séquences plus ou moins longues de discours. Il n'existait aucun travail systématique ni a fortiori formel sur les *paraphrases* mais aussi, surtout, sur les *transformations* et les *restructurations* de l'ensemble fini des constructions phrastiques d'une langue. Depuis les analyses transformationnelles de Zellig Sabetai Harris et surtout, depuis les descriptions effectuées dans la mouvance du *lexique-grammaire* de Maurice Gross, ce travail non seulement existe mais a, pour au moins une langue, le français, une couverture exceptionnelle.

La deuxième vient de l'histoire et de la fortune des démarches linguistiques partant de la constatation de l'existence de *différentes formes d'analogie entre les mots*. En effet, en 1862, paraît le premier dictionnaire dit *analogique* de l'histoire: le *Dictionnaire analogique de la langue française* de Jean-Baptiste Prudence Boissière. Cherchons y *vêtir* et *vêtement*⁵. On les trouve dans les petites colonnes du haut de la page avec un renvoi à l'article *Habillement*. Dans cet article nos mots ont certes leur place mais elle n'est pas définie ou plus exactement elle ne se définit que par la nature du sous-groupe de mots où elle se trouve. Sous-groupe où on ne trouvera aucun énoncé, aucune construction.

Bien que le dictionnaire analogique fournisse "*pour la première fois, un moyen commode de trouver les mots quand on a seulement l'idée des choses*" (Boissière – Préf. à la 3^e éd.: II) et que sa "*partie principale*" vise à "*conduire avec la plus grande facilité d'une idée quelconque, même conçue vaguement, à son expression précise et souvent à plusieurs expressions synonymiques entre lesquelles il sera permis de choisir*" (op. cit.: IV), enfin, bien que contrairement au dictionnaire habituel, qui "*n'est qu'un trésor fermé*

⁵ Éd. consultée: 3^{ème} éd., sans date, éd. par Auguste Boyer & Cie et impr. par Pierre Larousse (Biblioth. de la Sorbonne – cote LPf26) .

dont le possesseur ne connaît pas, ne peut jamais saisir les richesses" (op.cit.: VI) dans la mesure où il ne lui sert qu'à "vérifier un jugement porté d'avance", le dictionnaire analogique, à l'instar des anciens dictionnaires *français / latin* ou *français / grec* de *thème*, permet de chercher et de trouver des mots inconnus (op. cit., Intr.: IV), il ne parvient à ce résultat qu'en écrasant les différences grammaticales et sémantiques entre les *synonymes*. Boissière pense en effet que "Toute langue vivante forme en réalité deux langues, l'une usuelle et connue des plus ignorants, l'autre complète et renfermant, outre les mots compris déjà dans la première, une foule de termes qui n'entrent dans l'intelligence et dans la mémoire qu'à mesure que les connaissances s'élèvent au-dessus du niveau commun" et il précise: "Les termes de la langue usuelle s'apprennent dès les plus jeunes années, non dans des livres ni par les leçons d'un professeur, mais par les rapports de toute nature qui s'établissent entre l'enfant et tous ceux qui l'entourent; on n'a pas besoin d'expliquer les mots à l'enfant, on les emploie devant lui et il en devine le sens par une sorte d'intuition dont il ne se rend nullement compte, mais qui s'impose à lui par le fait même des circonstances extérieures. Il n'y a pas un enfant de trois ans qui ne comprenne parfaitement le mot fois, par exemple, et ce mot n'a jamais pu être clairement défini, ne le sera probablement jamais. Quelques années de cet enseignement vivant qui se fait de lui-même en quelque sorte, ou plutôt qui vient de tout le monde et de toutes choses, amassent dans la mémoire naissante de l'enfant une somme de mots qui s'y gravent d'une manière ineffaçable et qui s'identifient tellement avec les choses mêmes, qu'ils en deviennent la forme propre, la couleur, le vêtement naturel, et que jamais celles-ci ne se présentent plus à sa pensée sans porter les mots avec elle; **il ne pense plus en quelque sorte que par les mots. Donc, par cela seul qu'une chose devient l'objet de notre pensée, elle se formule à notre esprit en termes usuels quelconques, et ce qui peut seul nous faire défaut, c'est le terme précis, spécial, appartenant à la langue non usuelle. Mais dans cette formule usuelle qui est le vêtement, la couleur inséparable de l'idée, il y a nécessairement quelques termes saillants dont l'importance efface en quelque sorte tous les autres, et qu'il est toujours aisé de reconnaître; (...) Environ deux mille mots usuels ont été choisis comme marquant ainsi la place où il faudra venir chercher l'expression de certaines idées**" (op. cit., Préf.: III – c'est nous qui soulignons).

Le nombre de *mots-usuels* que la tradition lexicographique baptisera ensuite *mots-centres* est intéressant. On n'est pas loin du nombre de mots désignant des notions présumées universelles que certaines sémantiques ont essayé de postuler. On reste par contre très en-dessous du vocabulaire réputé *fondamental* – 5 à 7.000 puis 8 à 10.000 mots – que les linguistes pédagogues d'une époque relativement proche de nous jugeaient indispensable à l'apprentissage d'une langue. On n'entrera pas ici dans une discussion sur la notion, citée plus haut, de *saillance* dont Boissière semble avoir été l'un des premiers, sinon le premier, à tirer parti dans l'élaboration d'un mode d'accès à la signification et qui joue aujourd'hui le rôle central que l'on sait dans toutes les linguistiques de la cohésion textuelle comme dans les tentatives de traitement automatique de l'anaphore mais il semble bien qu'on ne puisse pas déterminer et définir la saillance de la même manière selon que l'on table sur moins de 2.000 ou sur plus de

10.000 éléments saillants. Avec moins de deux mille on se trouve au seuil d'une forme de *généricité infragrammaticale* ou encore de *généricité qui ignore la grammaire* qui réduit la saillance à une perception exclusivement sémantique et holistique. Avec plus de 10.000 il devient possible d'y intégrer la grammaire avec ce qu'elle comporte de mécanismes d'appropriation qui font fusionner syntaxe et sémantique dans des entités spécifiques.

Mais l'intégration de la grammaire au lexique est une entreprise difficile, impossible à mener par un homme seul à partir d'un corpus de citations littéraires. La tradition lexicographique française ne s'y est d'ailleurs pas trompée. Boissière aura deux types très différents de continuateurs mais aucun des deux n'articulera sa méthode analogique à la méthode de Lafaye, intégrant, rappelons-le, une grammaire dont le traitement unifié de la syntaxe et de la sémantique respecte et en partie explique les relations d'appropriation entre les unités informatives du lexique. Paul Robert couplera la méthode analogique de Boissière à celle des dictionnaires traditionnels, magnifiée dans l'œuvre d'Émile Littré, en cherchant dans la sélection des citations littéraires la représentation de la spécificité du profil d'une entrée. Quant à la maison Larousse, imprimeur puis propriétaire des droits du Boissière, elle développera le dictionnaire analogique en tant que dictionnaire spécifique mais dans un sens révélateur. Si la première refonte du Boissière par Ch. Maquet en 1936 ne modifie pas vraiment l'équilibre du dictionnaire, garde le même nombre de *mots-centres* et ne change rien à la méthode, la deuxième refonte de Georges Niobey en 1979 introduit par endroits des *définitions* ainsi que des *séquences de constructions* plus ou moins longues notamment pour *préciser* le sens des verbes et des adjectifs. Du coup le nombre des *mots-centres* augmente de quelques centaines et "*Lorsque le mot-centre est polysémique, ce qui est le cas le plus fréquent, l'article est divisé en un certain nombre de paragraphes qui permettent de saisir et les principales acceptions et les différents emplois du mot.*". (Niobey et alii 79: VI). On n'est pas encore tout à fait dans le sillage de Lafaye mais on s'en rapproche. C'est qu'entre-temps est paru le premier grand dictionnaire de la langue française fait en grande partie par des linguistes⁶ et que l'essentiel des travaux du *Lexique-grammaire* de Maurice Gross commence à être connu du grand public et des lexicographes.

Dès lors donc que l'analogie et la synonymie, au sens large de ce terme, accèdent au statut d'outils descriptifs du lexique elles sont prises dans un double mouvement: d'un côté une détermination grammaticale des conditions d'équivalence ou de différenciation, de l'autre une organisation graduée par rapport à un chef de file, une saillance. Les deux mouvements ne génèrent pas la même méthodologie et ne présupposent pas les mêmes postulats sur le mode de construction du sens. D'un côté, on pose que les relations grammaticales sont le lieu d'intégration, de coalescence des formes et des significations et que les équivalences ou les ressemblances sémantiques sont le fait de restructurations, c'est-à-dire de redistributions formelles et qu'il ne peut y avoir de définition et partant d'explication qui ne soit pas consubstantielle à ces

⁶ Le *Grand Larousse de la langue française* en 6 volumes chez Larousse, (1971/1972)

restructurations. De l'autre, on postule un nombre minimal mais suffisant d'entités génériques qui portent en elles-mêmes suffisamment d'aspects virtuels pour générer la totalité des entités spécifiques du monde. Ces dernières n'étant pas le produit de relations d'appropriation spécifiques entre les mots mais le résultat de dérivations imposées par l'évolution de l'usage: dérivations dans lesquelles la grammaire ne jouerait qu'un rôle aléatoire, difficilement prédictible et en dernière analyse relativement indépendant de la construction du sens.

Ces deux mouvements n'opposent pas Lafaye et Boissière. Ils sont communs aux deux auteurs même si les continuateurs de Boissière ont surtout retenu le second. Boissière lui-même n'écrivait-il pas dans sa préface que l'enfant qui apprend sa langue maternelle "*ne pense plus en quelque sorte que par les mots*" ? Il n'imagine pas que les *idées* ou les *notions* une fois fixées dans l'esprit humain de telle sorte qu'elles deviennent une dynamique de production du sens soient autre chose que les mots de la langue avec leurs propriétés linguistiques. Mais contrairement à Lafaye il ne semble pas avoir d'idée claire sur la manière dont ces mots en produisent d'autres. Il est excusable. Aujourd'hui encore, plusieurs linguistes et non des moindres répètent à l'envi que les mots naissent du contexte, de la situation, du point de vue, et de toutes ces conditions pragmatiques, par ailleurs fort pertinentes, mais dont ils ne décrivent jamais par quelle alchimie elles se transmutent en formes particulières soumises à des contraintes grammaticales résolument étrangères à ce qui n'est pas dans la langue.

Les deux mouvements permettent d'opérer des regroupements d'unités lexicales à partir desquels ou entre lesquels on peut calculer des degrés de distance ou de proximité.

Si l'on se réfère à des travaux récents associant une procédure d'analyse automatique à chacun des deux mouvements on trouvera du côté du traitement synonymique classique⁷ les recherches commencées à l'initiative de l'INALF et de l'ELSAP de Caen en ayant recours à la notion de *clique*. Ces recherches menées principalement sous la direction de Sabine Ploux, Bernard Victorri et Jean-Luc Manguin, ont abouti à l'élaboration d'un dictionnaire électronique des synonymes consultable sur le site du laboratoire CRISCO de Caen. La clique est *un sous-graphe complet maximal* qui constitue une projection topologique des relations synonymes constatées entre un ensemble d'unités lexicales. Elle présente le triple avantage de

⁷ "La relation de synonymie pure est une relation d'équivalence: elle est réflexive, symétrique et transitive. (...) Il existe très peu (sinon pas du tout) de synonymes purs. (...) Aussi utiliserons-nous ici la notion de synonymie partielle que nous appellerons désormais synonymie tout court que l'on peut définir de la façon suivante: Deux unités lexicales sont en relation de synonymie si toute occurrence de l'une peut être remplacée par une occurrence de l'autre dans un certain nombre d'environnements sans modifier notablement le sens de l'énoncé dans lequel elle se trouve. Cette relation est toujours réflexive et symétrique, mais elle n'est plus transitive: une unité peut être paraphrasable par une première unité dans certains environnements, et par une deuxième unité dans d'autres, sans que ces deux unités soient elles-mêmes synonymes" (PLOUX, VICTORRI 1998: 163).

rendre compte de la *hiérarchisation* des structures lexicales et partant de degrés d'homogénéité et de pertinence dans l'extension de la synonymie, de la *continuité* entre les différents sous-ensembles manifestée dans les parcours immédiatement visibles au sein d'un graphe et entre les différents graphes qui constituent une toile ou un réseau qui se veut plus ou moins isomorphe à l'ensemble des entrées lexicales, mais aussi, et, de notre point de vue, surtout, des entrées qui, en langue, ne sont pas et ne doivent pas être reliées. Cette démarche permet enfin une *extraction automatique de prototypes lexicaux* par le biais d'un calcul automatique et aveugle effectué directement sur le graphe et plus exactement sur la distance qui résulte de l'analyse du nombre de sommets directement, indirectement ou pas reliés au sein d'un sous-graphe.

Du côté de la démarche analogique, le meilleur exemple nous est donné par le travail qui se développe à Toulouse à l'initiative de Karine Duvignau; Olivier Gasquet, Bruno Gaume, Marie-Dominique Gineste (2002).

Dans les deux cas on ne calcule pas, nous semble-t-il, la même chose. Ce qui est commun aux deux démarches c'est qu'elles n'ont jamais recours de manière déterminante à la construction du mot dans la langue, à sa grammaire.

La première explique cet évitement par des raisons pratiques. L'introduction de la grammaire rendrait le système trop lourd et à la limite ingérable. Elle fait aussi plus ou moins explicitement le pari que le système peut, grâce notamment à des données liées aux fréquences, contourner les contraintes grammaticales et aboutir à des résultats à peu près équivalents à ceux qu'on obtiendrait en tenant compte des contraintes grammaticales.

La deuxième ne passe pas seulement par dessus les contraintes grammaticales strictes mais également par dessus les sélections argumentales élémentaires des verbes qui ont une incidence sémantique directe et se targue d'établir des liens qui n'existent pas ou rarement dans le dictionnaire, par exemple entre *écorcer un fruit, dénuder un fil électrique et déshabiller une personne*. Or il est clair et c'est probablement la raison pour laquelle la relation analogique traitée est indiscutablement présente chez les enfants de 3-4 ans que le dénominateur commun élémentaire et vraisemblablement universel à un certain niveau de la représentation symbolique à savoir *ôter matériellement ce qui couvre* ne revêt absolument pas *le même sens* c'est-à-dire n'entre pas dans *le même réseau de significations ni même dans des réseaux de signification analogues en langue*. Autrement dit que ce qui est présent dans la langue à un moment donné parce qu'il est motivé par un besoin cognitif n'y reste pas et n'a pas d'effet structurant décisif que ce soit dans la formation et l'évolution de l'architecture du lexique ou l'évolution parallèle et indissociable des propriétés lexicales et des contraintes grammaticales. L'analogie fait indiscutablement partie des principes linguistiques producteurs et on peut même identifier au sein de différentes langues, probablement de toutes, des *analogons*. Une partie importante de l'activité ludique à laquelle se prête la langue, par exemple les jeux de mots, fait jouer l'analogie contre la

ressemblance ou l'identité pour produire les effets qu'elle produit. Mais l'analogie, si elle s'applique à la langue, n'est pas un principe de fonctionnement de la langue. Si elle l'était les langues se ressembleraient beaucoup plus que ce n'est le cas et leurs grammaires n'auraient pas les formes qu'elles ont. Et c'est d'ailleurs parce que la construction du sens implique des formes de différenciation liées à des contraintes arbitraires qu'elle ne doit pas être décrite par des outils qui ne sont pas spécifiquement linguistiques.

Voilà également la raison pour laquelle nous proposons d'effectuer le calcul de la distance sémantique à travers l'élaboration d'une matrice analytique définitive dont aucun élément n'est étranger à la langue et à son fonctionnement.

Prenons l'énoncé de Laharpe cité par Lafaye:

"Les femmes en sont venues à s'habiller sans se vêtir"

il pourrait avoir la forme matricielle intermédiaire suivante:

"Les personnes adultes de sexe féminin sur lesquelles les personnes adultes de sexe masculin portent généralement leur regard sont, au moment et à l'endroit de mon énoncé, arrivées dans leur évolution à ce point où la façon dont elles se mettent des habits fait que c'est comme si elles ne portaient aucun vêtement"

Indépendamment du système de définition, sur lequel nous reviendrons, la distance sémantique à calculer est celle qui oppose *s'habiller* à *se vêtir*. La trace de cette opposition dans la matrice achevée est double: il y a d'abord celle de l'opposition lexicale structurelle, c'est-à-dire propre à l'architecture interne du lexique, dans la relation d'appropriation du verbe support *mettre* au prédicat nominal *habit* source de *s'habiller* qu'il actualise avec celle du verbe support *porter* au prédicat nominal *vêtement* source de *se vêtir*. Il y a ensuite la séquence discontinue qui relie les deux entités lexicales *la façon dont...fait que c'est comme si*. Cette séquence porte la trace d'une propriété lexicale classificatoire *la façon dont* qui aurait pu s'exprimer tout aussi bien par *la manière dont* et qui est justifiée par le fait que toutes *les définitions extérieures ou normatives*, c'est-à-dire celles des dictionnaires comme toutes *les définitions construites*, c'est-à-dire celles qui sont élaborées par induction à partir des expressions et collocations de l'usage indiquent que *s'habiller* construit sa signification à partir de la *manière de s'habiller* plutôt qu'à partir du fait de *porter* ou non des habits. Elle porte également la trace de l'*opérateur* (au sens de Ibrahim 2000b) qui rend l'opposition signifiante *faire comme si*.

Ces *traces* sont également l'état final d'un **parcours de restructurations** qui relie l'énoncé observé à la matrice analytique définitive. Dans le cadre que nous proposons, la complexité sémantique se confond avec ces parcours de restructurations et la distance sémantique entre deux expressions, que ce soit au sein d'un même énoncé ou entre deux occurrences apparaissant dans deux énoncés différents, correspond au différentiel qui oppose les parcours respectifs qui les relient à une matrice analytique.

Un parcours de restructurations comporte les opérations qui sont nécessaires pour **rendre compte des lexies complexes par des lexies plus élémentaires indécomposables** ainsi que par l'obligation de **convertir, toutes les fois où cela est possible, les précidats non-nominaux en prédicats nominaux** afin d'extraire de l'énoncé observé tout *l'outillage d'actualisation discursive qui est aussi un instrument d'abrègement, de condensation et de construction du sens*.

Les parcours peuvent être très variés et ne sont pas soumis à un protocole unique même si les exigences de simplification pour parvenir à la formulation exacte de la matrice analytique sont toujours les mêmes. En effet les degrés de complexité des lexies de la langue varient énormément selon la place qu'elles occupent dans le lexique, c'est-à-dire selon le type de grammaire locale qui gouverne leurs relations aux autres entrées du lexique.

Dans le cas qui nous intéresse les actualisations différentes au sein de la matrice de *habit* et *vêtement*, l'association au premier et non au second nom de *façon* et l'insertion de l'opérateur *faire comme si* constituent en quelque sorte l'aboutissement d'une série de contraintes constatées au cours de divers types de restructurations y compris celles qui aboutissent à des suites agrammaticales. On y trouve:

(1) Des contraintes d'ordre: **La façon dont elles portent leurs vêtements fait que c'est comme si elles ne mettaient aucun habit.*

(2) Des contraintes sur la possibilité et l'interprétation des formes nominales source ou dérivées:

(a) *du verbe conjugué*: On est vêtu des *vêtements* qu'on porte d'ordinaire par besoin ou par commodité. // On est habillé selon la manière dont on est mis, dont on a mis ses *habits* // Vêtu, nos *vêtements* couvrent notre corps, habillé, nos *habits* s'ajustent à notre corps, sont mis d'une certaine façon, sont mis pour mettre en valeur, montrer ou dire quelque chose.

On ne pourra pas dire sans éprouver un fort sentiment d'incongruité sémantique **Il était habillé des habits qu'il portait d'ordinaire* ou qu' **Il portait ses habits ainsi par habitude* pas plus que **On est vêtu selon la manière dont on a mis ses vêtements*.

(b) *du verbe support* : *Leur mise fait que c'est comme si elles ne portaient aucun vêtement.* // *Leur mise correspond à une absence de vêtement* mais pas **Leur port correspond à la façon dont elles mettent leurs vêtements* ou **Leur port correspond à leur mise*.

(c) *du nom source du verbe*: *L'habillement du roi occupait une bonne douzaine de personnes pendant une bonne partie de la matinée* mais pas **Le vêtement du roi....*

(3) Des contraintes sur les dérivations associées aux privatifs ou aux antonymes: *Elle s'habillent de déshabillés* // *Elles s'habillent sans s'habiller* mais non **Elles se vêtent de dévêtus* // *Elles se vêtent sans se vêtir*. Ces contraintes ont des conséquences directes sur l'usage: on ne peut pas substituer *dévêtez-vous* à *déshabillez-vous* dans *Déshabillez-vous dans l'entrée. Il fait très chaud dans le salon*. De même on peut *déshabiller* mais non *dévêtir* quelqu'un *du regard*.

- (4) Des différences d'interprétation résultant de l'association de la forme verbale avec un adverbe de jugement : *Qu'est-ce que la reine d'Angleterre peut être mal habillée!* ou *Qu'est-ce qu'elle s'habille mal!* ou encore *Déjà à l'époque elle s'habillait comme on ne s'habillait plus!* - Des énoncés tout simplement inconcevables avec *vêtir* ou *se vêtir*.
- (5) Des différences radicales d'acceptabilité lorsque le résultatif dérivé se trouve sous la portée d'un intensificateur: *Il s'étendait sur le lit, le plus souvent tout habillé* (Vailland cité par le *Lexis*) mais non **tout vêtu* ou *Cette robe fait très habillé* et non **très vêtu*.
- (6) Et naturellement toutes les contraintes qui dérivent indirectement d'une *définition de la manière*. Ainsi il est impossible de substituer *vêtu* à *habillé* dans *Jeune avocat habillé à Londres* (Elsa Triolet citée par le *Lexis*), ou *Un enfant habillé en Indien* pas plus qu'il n'est possible de substituer *se vêtir* à *s'habiller* dans *S'habiller chez un tailleur réputé*.

Enfin cette complexité n'est pas simplement un phénomène lié à la combinatoire des constructions verbales et aux transformations ou dérivations qu'on peut ou non leur appliquer. Elle touche également des entrées nominales non-prédicatives pour lesquelles des associations de type analogique sont envisageables. Par exemple l'entrée *femmes* de notre énoncé est en fait une entrée très complexe dans la mesure où la définition que nous lui avons donnée résulte de l'analyse de l'évolution des définitions dictionnaires qui ont été construites par le classement de séries d'énoncés fabriqués par le lexicographe ou tirés de citations littéraires à travers lesquels se dégagent au moins cinq types d'oppositions paradigmatiques:

- (a) Une définition *organique* où *femme* s'oppose explicitement ou par défaut à *fille* et à *homme* et qui permet entre autres de comprendre pourquoi l'opposition *jeune femme* / *jeune fille* n'est pas symétrique.
- (b) Une définition *essentialiste* où *femme* est un *attribut* terminal, c'est le cas notamment dans des exemples du type: *Elle est femme, très femme*. [Rédac. P.R.] ou *Elle est femme dans toute l'acception du mot, par ses cheveux blonds, par sa taille fine (...) par le timbre argentin de sa voix* [Théophile Gautier] ou encore *Réponse bien féminine! Que vous êtes femme, mon Dieu, que vous êtes femme! Que vous êtes charmante!* [Claude Aveline]
- (c) Une définition *relationnelle* dans la construction *femme de N_{hum}* (= épouse de)
- (d) Une définition *catégorisante* dans les constructions *femme Prep N_{nr}* ou *N_{hum}femme* ou *femme-N_{nr}*
- (e) Une définition *générique* correspondant généralement à l'occurrence de *femme* en fonction *complément*, le plus souvent en construction directe mais également dans une construction *transitive indirecte* et parfois en fonction de sujet. Il s'agit généralement d'un statut définitoire construit par le regard de l'autre et qui, malgré sa généralité définit un sous-ensemble qui ne se confond pas avec *toutes les femmes* ni avec *les femmes en général*.

Enfin la matrice doit montrer sur quel mode s'actualisent les coordonnées spatio-temporelles de l'énonciation: *au moment de mon énoncé, au moment de mon énonciation, au moment et à l'endroit de mon énoncé, de mon énonciation*. Mais aussi des insertions comme *généralement* dans *portent généralement leur regard*.

Une matrice analytique est le développement définitoire maximal d'un énoncé de la langue naturelle au moyen d'un enchaînement grammatical très faiblement contraint d'unités linguistiques élémentaires indécomposables. La matrice ainsi constituée décrit simultanément deux parcours corrélés et qui ne peuvent pas être démêlés efficacement dans un énoncé terminal observé: le parcours grammatical qui téléscopie, synthétise et abrège les formes élémentaires primitives et le parcours interprétatif qui accompagne le parcours grammatical en assignant des valeurs définitoires aux opérations grammaticales au fur et à mesure qu'elles se réalisent. La matrice a donc, par le biais de sa fonction définitoire, en même temps qu'elle défait ou déplie la contrainte grammaticale, un certain pouvoir explicatif.

L'élaboration de ces matrices ne se justifierait pas si:

- /1/ La grammaire n'était pas un vaste processus d'abrègement et de condensation.
- /2/ Les mots et les suites de mots n'étaient pas en majorité des entités complexes décomposables en entités plus élémentaires de deux types:
 - (a) des entités élémentaires appariées en paires ou en triades plus ou moins fortement appropriées l'une à l'autre.
 - (b) des entités élémentaires indécomposables à combinatoire ouverte, libres de toute contrainte d'appropriation.
- /3/ Les contraintes syntaxiques et plus généralement toutes les contraintes grammaticales n'étaient pas pour l'essentiel liées à la nature des unités lexicales.
- /4/ Le métalangage de description qui formule ces contraintes n'était pas entièrement contenu dans la langue naturelle et ne se manifestait pas spontanément au cours de son usage.
- /5/ S'il n'existait pas des *classes d'équivalence plus ou moins imparfaites* qui permettent la conversion des éléments observés dans des séquences condensées en leurs composants élémentaires déployés dans des *séquences étendues plus ou moins fortement redondantes*.
- /6/ La construction du sens n'était pas un processus de *différenciation*, la construction d'une *valeur différentielle* au travers d'une série ordonnée de classes d'équivalences imparfaites qui garantissent une *continuité identitaire* ou une *permanence référentielle* c'est-à-dire ce processus complexe de reprise sélective, orientée et appropriative qu'est la *reformulation* (cf Martinot dans cette livraison).

Références

Boissière, Jean-Baptiste Prudence, 1862, *Dictionnaire analogique de la langue française, Répertoire complet des mots par les idées et des idées par les mots*, Paris: Auguste Boyer et C^{ie}, Libraires - Editeurs.

- Duvignau, Karine, Gaume, Bruno, Gasquet, Olivier, Gineste, Marie-Dominique, 2002, "Categorisation of Actions by Analogy: from the Analysis of Metaphoric Utterances to a Computational Model",
- Gross, Maurice, 1975, *Méthodes en syntaxe*, Paris: Hermann
- Harris, Zellig Sabetai, 1969, "The Two Systems of Grammar: Report and Paraphrase", *Papers in Structural and Transformational Linguistics* (1970), Dordrecht: D. Reidel.
- Ibrahim, Amr Helmy, 1999, "Constructions figées et constructions à supports", in MEJRI, Salah, CLAS, André, GROSS, Gaston, BACCOUCHE, Taïeb (Eds), *Le figement lexical*, Actes des 1ères Rencontres Linguistiques Méditerranéennes de Tunis, 17-19 septembre 1998, Tunis: Université de Tunis I / CERES, 373-387.
- 2000a, "Constantes et variables de la grammaire des supports dans quelques langues romanes", *Actes du XXIIe Congrès international de Linguistique et de Philologie romanes* (Bruxelles 23-29 juillet 1998), Vol. VI, *De la grammaire des formes à la grammaire du sens*, Tübingen: Max Niemeyer Verlag, 241-250.
- 2000b, "Une classification des verbes en 6 classes asymétriques hiérarchisées", *Syntaxe et Sémantique 2*, SÉMANTIQUE DU LEXIQUE VERBAL, Caen: Presses Universitaires de Caen, 81-98.
- 2001, "Dans la langue, par la langue mais toute entière", & "Argumentation interne et enchaînements dans les matrices définitives", *Langages* 142, *Les discours intérieurs au lexique* (Amr Helmy Ibrahim Ed.), 3-9 & 92-126.
- 2002, "Maurice Gross: une refondation de la linguistique au crible de l'analyse automatique" – *Actes de la 9^{ème} Conférence Annuelle sur le Traitement Automatique des Langues Naturelles* (TALN 2002), et de la 6^{ème} *Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues* (RECITAL 2002), Tome I, 5-30 – CD Rom JEP / TALN / Récital 2002, Nancy: CNRS & Université de Nancy 2.
- 2003a, "Les verbes supports en arabe", *Bulletin de la Société de Linguistique de Paris*, Tome XCVII, fasc.1, Louvain: Peeters, 315-352.
- 2003 (à paraître), "Le cadre du *Lexique-grammaire*", *LINX* 48
- Lafaye, Pierre-Benjamin, 1853, *Dictionnaire des synonymes de la langue française avec une introduction sur la théorie des synonymes* – Édition consultée: 1865, Paris: Librairie de L. Hachette et C^{ie}, Boulevard Saint-Germain, n° 77 (Deuxième édition, revue, corrigée et suivie d'un supplément).
- Martinot, Claire, 2003, "Acquisition des verbes et reformulations", *Revue d'Intelligence Artificielle*, Paris: Hermès.
- Niobey, Georges (Dir.) et alii, 1979, *Nouveau dictionnaire analogique*, Paris: Larousse.
- Ploux, Sabine & Victorri, Bernard, 1998, "Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes", *Traitement automatique des langues*, 39 (1): 161-182.