Représentation informatique de la connaissance lexicale

D. Kayser
LIPN - URA 1507 du CNRS
Institut Galilée - Université Paris-Nord

Introduction

Pour réaliser des machines effectuant un traitement "intelligent" de données textuelles, il est évident qu'il faut faire des progrès non seulement dans le domaine des analyseurs syntaxiques, qui ont absorbé une part importante de l'énergie des chercheurs au cours du dernier quart de siècle, mais aussi dans la représentation de la connaissance sémantique et pragmatique.

Même s'il est difficile de tracer entre eux une ligne de partage claire, on peut cependant définir au moins deux niveaux (fortement coordonnés) de connaissances sémantico-pragmatiques : un niveau lexical et un niveau grammatical. Celui-ci concerne l'interprétation des structures, indépendamment des occurrences de mots qui instancient ces structures, alors que celui-là concerne justement les éléments de signification liés au choix de ces instances.

Le problème est que les connaissances qui interviennent à ces deux niveaux sont loin d'être parfaitement déterminées, et qu'il est donc difficile de les représenter d'une façon utilisable par un ordinateur. Mais une méthode possible pour parvenir à mieux les définir est justement de commencer à les implanter en machine : l'observation des inadéquations du comportement obtenu permettra de diagnostiquer les lacunes des connaissances fournies.

Une façon d'initialiser ce processus est de tenter de "mettre en machine" les connaissances lexicales fournies par les dictionnaires. Cette idée a été à l'honneur au cours des dernières années (cf. notamment Véronis & al.1989) mais on se rend vite compte qu'elle ne peut pas nous mener bien loin. En effet, les dictionnaires sont faits pour des êtres humains qui disposent d'un grand nombre de connaissances. Un dictionnaire qui répond à leur attente n'est pas rempli de définitions, mais d'indices qui permettent d'accéder à l'information qu'ils recherchent. Comme le remarque, par exemple, Martin (1990), les dictionnaires se contentent rarement des définitions "minimales" et fournissent des éléments "stéréotypiques" d'une grande importance pour comprendre la variété des usages d'un mot (voir également la remarque de Putnam "*On a souvent stigmatisé comme un défaut des dictionnaires le fait qu'ils sont encombrés (...) de bribes éparses d'information factuelle (...) Ce n'est pas du tout un défaut (...)*" Putnam 1970).

En essayant de présenter ces éléments sous la forme de définitions, on obtient certes une apparente homogénéité, mais peu de véritables définitions et beaucoup de redondances et d'informations inutiles. Par exemple, la volonté compréhensible de limiter les renvois d'un article à un autre mène à de nombreuses redites. De plus, certaines "polysémies" obéissent à des règles simples ; par exemple, un mot désignant une action désigne généralement aussi le résultat de cette action ; or les dictionnaires répètent pour définir chacun de ces mots : "*action de xxx ; résultat de cette action*", ce qui les encombre inutilement (Kayser, 1992). Il est vrai cependant que peu de "polysémies" sont aussi régulières que celle que nous venons de mentionner ; il est fréquent que l'expression des règles étendant le sens d'un mot soit délicate, et nombreuses sont celles qui admettent des exceptions (Kayser & Abir, à paraître), au point qu'il est parfois moins économique de fournir la règle que d'indiquer directement son résultat pour chacun des mots auxquels elle s'applique.

Une autre caractéristique, reprise peut-être sans réflexion suffisante, des dictionnaires actuels est la présentation arborescente des différents "sens" d'un mot ; il s'agit probablement de la forme la plus adaptée à la nature de ces dictionnaires, mais les structures informatiques permettent une beaucoup plus grande souplesse (p.ex. un treillis permettant l'héritage multiple, ou plusieurs treillis, selon la propriété concernée).

Représentation logique

Ces remarques concernant la différence entre ce qu'on attend d'un dictionnaire destiné à des hommes et d'un dictionnaire utilisé par des machines ne nous donnent guère de piste pour concevoir ces derniers. Aller plus loin requiert d'ailleurs une connaissance de la tâche à accomplir : selon que l'on demande à la machine une simple vérification orthographique, une gestion de nomenclatures, une indexation automatique de documents, ou des réponses à des questions en utilisant le cas échéant un système déductif, il est évident que les connaissances que l'on doit trouver dans les dictionnaires ne seront pas les mêmes. Conformément aux objectifs que s'est toujours fixés l'Intelligence

Artificielle, nous nous placerons dans le cadre le plus ambitieux : celui de machines devant effectuer des **inférences** à partir de textes.

Les inférences admissibles obéissent à des régularités, et il serait agréable de les assimiler aux inférences valides définies en logique. Il suffirait alors de mettre dans les dictionnaires les expressions logiques correspondant à chaque mot, et d'exprimer la sémantique grammaticale par des lois de combinaison d'expressions logiques. C'est le programme que s'était notamment assigné R. Montague, et on le retrouve dans des ouvrages récents de sémantique (Chierchia & McConnell-Ginet 1990, Gamut 1991).

Sans vouloir développer ici tous les obstacles que rencontre ce programme (voir discussion p.ex. dans Kayser, 1990), je signalerai simplement qu'il présuppose la possibilité de décider pour chaque mot s'il correspond à une ou à plusieurs expressions du langage logique ; en d'autres termes, s'il est ou non ambigu. Or l'idée d'une correspondance "un mot $\rightarrow n$ sens" n'est qu'une approximation assez grossière. Ce n'est d'ailleurs pas un hasard si les dictionnaires divergent tant dans le nombre et l'articulation des différents "sens" prêtés à un même mot. À ma connaissance, aucun critère satisfaisant permettant de savoir à quelle condition deux emplois peuvent être rattachés au même "sens" n'a d'ailleurs jamais été énoncé.

Primitives Sémantiques

Supposons cependant, pour les besoins de l'exposé, qu'une liste de formules ait été définie pour chaque mot de la langue, liste présumée satisfaisante pour couvrir tous les usages de ce mot. Le problème des inférences admissibles sur ces formules reste posé. Les deux solutions extrêmes sont, comme Montague, d'affecter un prédicat différent à chaque mot, ou, comme Schank et Wilks, de réduire toute signification à une combinatoire sur un jeu restreint de primitives sémantiques.

Dans le premier cas, aucune inférence n'est possible tant que les prédicats ne sont pas reliés entre eux par des postulats de signification, mais aucune méthode n'est donnée pour trouver ces postulats ni, les ayant écrits, pour savoir s'ils suffisent à effectuer les inférences admissibles dont on cherche à rendre compte.

Dans le deuxième cas au contraire, le petit nombre de primitives sémantiques permet, en principe, d'exprimer et de justifier toutes les inférences acceptables. En d'autres termes, s'il était possible de trouver une "forme canonique" de la signification, c'est-à-dire une réduction des formules à un ensemble de prédicats primitifs, l'énoncé d'un système axiomatique adéquat sur ces prédicats primitifs suffirait à garantir une sorte de complétude du système déductif dans son ensemble.

Yorick Wilks (1977) a exprimé les conditions auxquelles un ensemble de prédicats pouvait être pris comme primitif (faible cardinalité, exhaustivité, indépendance, unicité de la décomposition des non-primitifs en primitifs, etc.). Lui-même, et d'autres chercheurs comme p.ex. Schank, ou, en France, Jean-Pierre Desclés, ont proposé des primitives satisfaisant plus ou moins ce cahier des charges.

Cependant, il semble que malgré l'intérêt de ces propositions, la réduction de la variété sémantique des mots de la langue à un petit nombre de prédicats impose à ceux-ci une multitude d'interprétations rendant la description de leur comportement inférentiel à peu près impossible. Cette constatation n'exclut pas une décomposition sur un ensemble de prédicats beaucoup plus vaste (des centaines, voire des milliers), mais l'espoir d'en donner une axiomatisation adéquate serait alors plus faible.

La relation de déductibilité

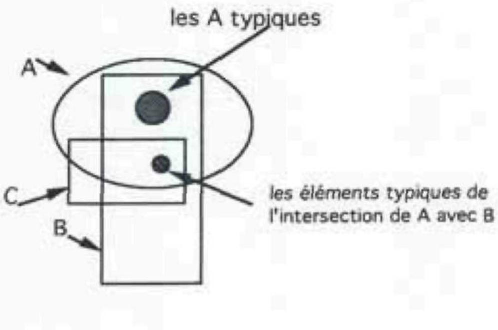
Venons-en maintenant à la nature de la relation de déductibilité nécessaire pour effectuer les inférences souhaitées. Beaucoup de définitions lexicales sont à caractère hypéronymique, et de la définition "A : sorte-de B", on est fondé à adopter le schéma d'inférence $A(x) \dashv B(x)$ où x s'instancie sur un terme du langage. Mais la relation notée $\dashv$ n'est pas la relation de déductibilité de la logique classique. En particulier, elle n'est pas transitive.

Un exemple significatif de non-transitivité est fourni par exemple par A.Nazarenko (Perrin 1989). Analysant les définitions présentes dans un petit dictionnaire, elle a obtenu successivement "*tombe*" sorte de "*fosse*" ; "*fosse*" sorte de "*trou*" ; "*trou*" sorte d'"*ouverture*" ; "*ouverture*" sorte de "*passage*" ; Chacune de ces définitions a sa raison d'être, mais la déduction obtenue par transitivité n'est pas valide.

Que s'est-il passé ? Chaque définition a été exprimée par rapport à un "sens typique" du mot, et s'il est vrai qu'une fosse typique est une sorte de trou, elle n'est pas un trou typique. Cette situation est très voisine de ce qu'on observe dans les modèles des logiques non-monotones. En effet, si on note $A \vdash B$ le fait que la proposition B est une conséquence non-monotone de A, de nombreux travaux montrent la correspondance de ce fait avec "B est vrai dans tous les modèles **préférés** de A". La non-transitivité de l'opérateur $\vdash$ vient de ce que la vérité de B ne suffit pas à établir que l'on est dans un modèle préféré de B. La ressemblance avec notre problème est grande ; c'est pourquoi nous allons considérer ce qui se passe si nous assimilons "B est vrai dans tous les modèles préférés de A" à "les A typiques sont des B".

Nous nous baserons pour cela sur la relation de déductibilité étudiée récemment par Kraus & al. (1990). Ces auteurs supposent cette relation fermée par 5 principes : la réflexivité, l'affaiblissement à droite, l'équivalence à gauche, la "coupure", et la "monotonie prudente".

- la *réflexivité* ($A \vdash A$) signifie "un A typique est un A", ce qui ne soulève guère de problème ;
- l'*affaiblissement à droite* (si $A \vdash B$ et $B \supset C$, alors $A \mid\!\sim C$) signifie "si les A typiques sont des B et si tous les B sont des C, alors les A typiques sont des C", ce qui est également conforme à l'intuition ;
- l'*équivalence à gauche* (si $A \equiv B$ et $A \vdash C$, alors $B \vdash C$) est un peu plus problématique ; en effet, ceci se traduit par "si les A typiques sont des C et si l'ensemble des A coïncide avec l'ensemble des B, alors les B typiques sont des C". En effet, on peut envisager que la notion de *voiture* coïncide avec celle de *bagnole*, mais que la voiture typique ait des propriétés que la bagnole typique n'a pas. Malgré tout, ce genre de situation ne semble pas très fréquent ;
- la "*coupure*" (si $A \vdash B$ et $A \wedge B \vdash C$, alors $A \vdash C$) correspond à l'assertion "si les A typiques sont des B et si les éléments typiques à l'intersection des A et des B sont des C, alors les A typiques sont des C". Une interprétation ensembliste aidera à concevoir une situation mettant cette assertion en défaut :



La situation qu'illustre ce diagramme serait celle dans laquelle les A typiques seraient entièrement distincts des éléments typiques de $A \cap B$ (ceci n'a rien d'impossible : on peut considérer que les *aliments* typiques sont solides (le pain, la viande,...) et qu'il y a des éléments typiques à l'intersection des *aliments* et des *liquides* (le lait)) mais il faudrait ajouter que les A typiques ont cependant la propriété B, et cette situation ne semble pas se produire dans le langage. Plus précisément, il ne semble pas possible que, si les A typiques sont des B, les éléments typiques de $A \cap B$ ne soient pas des éléments typiques de A[1] ;

- la dernière propriété, appelée "*monotonie prudente*" (si $A \vdash B$ et $A \vdash C$, alors $A \wedge B \vdash C$), est en quelque sorte la réciproque de la précédente, et l'hypothèse précédente (à savoir que si les A typiques sont des B, les $A \wedge B$ typiques sont des A typiques) suffit à la vérifier.

Un premier résultat de Kraus & al. (1990) est de montrer qu'un "modèle" permet de définir une relation de déductibilité vérifiant ces principes. Sommairement, ce qu'on appelle ici "modèle" est une sorte de généralisation des structures de Kripke. Au lieu d'avoir des "mondes" et une relation de visibilité entre les mondes, on se donne des états, une relation de précédence sur les états, et chaque état est "étiqueté" par un ensemble de mondes. On dit qu'un état **satisfait** une formule φ ssi tous les mondes étiquetant cet état vérifient φ. On note par $Sat(\varphi)$ l'ensemble des états qui satisfont une formule φ donnée. La relation de précédence n'est pas nécessairement un ordre, mais par abus de langage, on appellera "minimal" un état qui n'est précédé par aucun autre. On se donne pour contrainte que tout état soit ou bien minimal, ou bien immédiatement précédé par un état minimal.

---

[1] Pour essayer de mettre cette propriété en défaut, on peut concevoir la situation suivante :

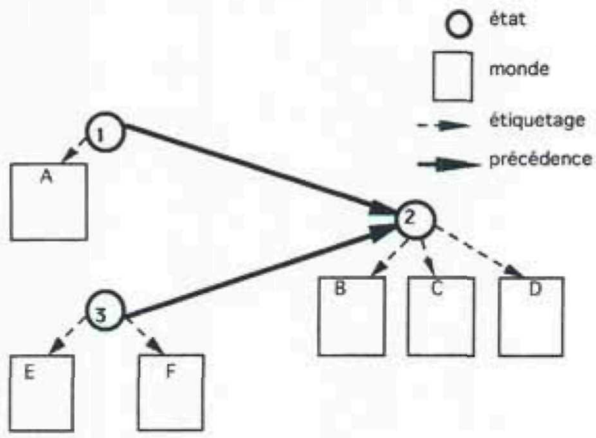
(α) en France, les informaticiens typiques parlent anglais

(car la plupart des documents qu'ils doivent lire sont en anglais), mais on ne peut pas dire que :

(β) un informaticien-parlant-anglais typique est un informaticien typique

En effet, l'image d'un informaticien-parlant-anglais typique est celle de quelqu'un qui a une bonne maîtrise de la langue anglaise, alors que l'informaticien typique ne fait que se débrouiller dans cette langue. Mais il me semble que cet exemple n'a de valeur qu'au prix d'un glissement de sens sur "parler". Dans (α), ce verbe doit être compris comme "être capable de prononcer et de comprendre la plupart des phrases", alors qu'en (β), il y a un intensif implicite : "parler" y signifierait "bien parler". L'exemple ne réfute donc pas la propriété énoncée, puisque les A typiques sont des B_1, mais ce sont les $A \wedge B_2$ typiques qui ne sont pas des A typiques.

Le diagramme suivant permet de mieux comprendre la notion de "modèle".



Ce diagramme décrit un "modèle" à 3 états, ceux-ci étant respectivement "étiquetés" par 1, 3, et 2 mondes. Supposons par exemple que le langage se limite à 3 propositions, p, q, et r, et, afin d'illustrer le mécanisme, choisissons arbitrairement des valeurs de vérité pour ces propositions dans chacun des mondes. Ainsi, admettons que p est vrai dans tous les mondes sauf D et F, q n'est vrai qu'en A et B, r est vrai sauf en E. Dans ce cas, **1** satisfait p, q, et r (car ces formules sont vraies en A, le seul monde étiquetant **1**); **2** satisfait r (car r est vrai en B, C, et D, les mondes étiquetant **2**) ; **3** satisfait $\neg$ q. On peut également calculer Sat(p) = Sat(q) = {**1**}, Sat(r) = {**1**, **2**}, Sat(p $\vee$ r) = {**1**,**2**,**3**}.

La relation de déductibilité définie par un tel "modèle" est la suivante : A $\vdash$ B ssi tous les états minimaux de Sat(A) — c'est à dire les états de Sat(A) qui ne sont précédés par aucun autre état de Sat(A) — sont dans Sat(B).

Sur notre exemple, on a p $\vdash$ q car le seul état de Sat(p), qui est donc minimal, est également dans Sat(q). On a aussi r $\vdash$ q, car parmi les deux états de Sat(r), un seul est minimal et celui-ci est dans Sat(q). Cependant, on n'a pas p $\vee$ r $\nvdash$ q car sur les trois états de Sat(p $\vee$ r), il y en a deux minimaux et ils ne sont pas tous les deux dans Sat(q).

Le résultat majeur de Kraus & al. (1990) est qu'il est possible de "synthétiser" une relation de préférence à partir de toute relation de déductibilité vérifiant les 5 propriétés. Cette préférence n'est pas nécessairement un ordre. Techniquement, on définit des classes d'équivalence de formules par A ~ B si et seulement si A ⊢ B et B ⊢ A.

C'est effectivement une relation d'équivalence (la réflexivité de ~ vient de la réflexivité de ⊢ ; la symétrie est donnée par la définition ; quant à la transitivité, supposons A ~ B et B ~ C, c'est-à-dire :

1. A ⊢ B
2. B ⊢ A
3. B ⊢ C
4. C ⊢ B

on peut montrer :

5. B ∧ A ⊢ C (Monotonie prudente 2,3)
6. A ∧ B ⊢ C (Équivalence à gauche 5)
7. A ⊢ C (Coupure 1,6)

et :

8. B ∧ C ⊢ A (Monotonie prudente 3,2)
9. C ∧ B ⊢ A (Équivalence à gauche 8)
10. C ⊢ A (Coupure 4,9)

donc, par définition, A ~ C)

Notons $\underline{A}$ la classe d'équivalence de la formule A. On associe un "état" à chaque classe d'équivalence, et on dit que l'on "préfère" l'état $\underline{A}$ à l'état $\underline{B}$ si ces états sont distincts et s'il existe une formule A de $\underline{A}$ telle que B ⊢ A.

Il s'agit bien d'une relation entre états, car s'il existe une formule A_0 de $\underline{A}$ telle que B ⊢ A_0, toute formule de $\underline{B}$ ⊢ A_0. En effet, soit B' une formule de $\underline{B}$. On a B' ~ B donc :

B' ⊢ B
B ⊢ B'
B ⊢ A_0 (hypothèse) ; ces 3 lignes correspondent, à un renommage près, aux lignes 1 à 3 ci-dessus, lignes qui nous ont suffi pour prouver 7 ; qui s'écrit ici B' ⊢ A_0.

On voit que cette relation de préférence est indépendante de la proposition considérée à l'intérieur d'une même classe d'équivalence. En assimilant "préféré" à "typique", ceci revient à dire que le fait qu'un état soit plus typique qu'un autre[2] ne dépend que de la classe de la formule dont il est typique, et non de la formule elle-même. Ceci explique d'une part la nécessité que les éléments typiques de deux prédicats logiquement équivalents coïncident, et d'autre part l'impossibilité des situations dessinées sur notre graphique : si les A typiques sont des B (si tous les états minimaux de Sat(A) satisfont B), un état typique de A ∩ B (i.e. un état minimal *e* de Sat(A ∧ B)) est obligatoirement typique de A.

e est dans Sat(A), et s'il n'était pas minimal, il y en aurait un minimal *f* qui lui serait préféré ; par hypothèse, *f* satisferait aussi B, donc serait dans Sat(A ∧ B) et *e* ne serait pas minimal dans cet ensemble : contradiction.

[2]Il convient de rappeler que ce vocabulaire, volontairement simplificateur, peut prêter à confusion, notamment parce que la préférence n'est pas une relation d'ordre.

En d'autres termes, comme la relation de préférence entre états est fixe (elle ne dépend pas de la relation de typicité que l'on cherche à établir), notre tentative d'assimilation entre "préféré" et "typique" mène à déclarer qu'un élément est typique avant de savoir précisément de quoi il est typique, ce qui est évidemment inadéquat. Cependant, il n'est pas évident que cette inadéquation ait des conséquences pratiques défavorables.

Un exemple

Revenons au paradoxe mis en évidence dans Perrin (1989), et regardons comment quelques dictionnaires gèrent les mots qu'ils utilisent.

Littré définit *tombe* par :

1. Grande table de marbre, de pierre, de cuivre etc. dont on couvre la fosse qui contient un mort ;

2. (par extension) tombeau (lui-même défini par 1. monument élevé à la mémoire d'un mort au lieu même où il est enterré ; 2. (par extension) lieu où l'on périt ; 3 à 9 : sens figurés ou particuliers) ;

3. (figuré et poétiquement) la Mort.

La définition hypéronymique *tombe* $\vdash$ *table* mène très rapidement au paradoxe (Littré répertorie 34 sens de *table*, les premiers étant "planche", "planche ou réunion de planches portée sur un ou plusieurs pieds" ; seul le 16ème sens ayant pour illustration les tables de la Loi peut éventuellement convenir) ; pour obtenir le passage souhaité tombe $\vdash$ fosse, il n'y a pas d'autre solution que de recourir à la métonymie. Le Petit Robert intègre cette métonymie (si c'en est une) en donnant pour sens premier de *tombe* "lieu où l'on ensevelit un mort, fosse recouverte d'une dalle", puis "pierre tombale, monument funéraire" et enfin, métaphoriquement "la mort".

Si on explore maintenant les définitions de *fosse*, on trouve dans Littré 14 sens, dont seuls deux sont généraux :

1. Creux fait dans la terre par la nature ou par la main de l'homme.
2. Trou creusé en terre et dans lequel on met les morts.

À noter que *creux* renvoie à cavité, partie concave, vide ; que *cavité* se définit comme "vide dans un corps solide", *trou* par "ouverture en creux faite dans un corps", *ouverture* par "fente" (où *fente* se définit par "petite ouverture en long"), "trou", "espace vide",... et par "action d'ouvrir".

On peut également remarquer que le Trésor de la Langue Française découpe différemment les sens de *fosse* :

I. Cavité d'origine artificielle
  - A. Cavité large et profonde creusée dans la terre, généralement destinée à recevoir qqch ;
  - B. Trou creusé en terre et destiné à l'inhumation des morts
  - C. 9 sens particuliers structurés en 3 catégories

II. Cavité d'origine naturelle
  - A. Cavité large et profonde
  - B. Plusieurs sens particuliers.

Le Petit Robert suit cette distinction, en modifiant la hiérarchie (1. Cavité creusée par l'homme, 2. Trou pour l'inhumation, 3. Cavité naturelle).

Comment représenter ces informations lexicales ? Sauf si nous suivons Littré, la règle *tombe* $\vdash$ *fosse* ne posera pas de grave problème. Mais les divers découpages proposés par les dictionnaires pour *fosse* conduisent à s'interroger : faut-il distinguer $fosse_1$ (avec *tombe* $\vdash fosse_1$ et $fosse_1 \vdash$ *trou*) et $fosse_2$ (avec $fosse_2 \vdash$ *cavité*) ? Ce serait prendre précisément la direction que l'on souhaite éviter, celle d'un dénombrement des significations.

Une autre possibilité serait l'introduction d'une disjonction : *fosse* $\vdash$ *trou* $\vee$ *cavité*, mais il est connu que la disjonction modifie considérablement le comportement déductif. En particulier, il faut décider si ayant $A + C$ et $B \mid\!\sim C$, on accepte $A \vee B + C$. Il convient de remarquer que ceci n'est pas une conséquence des 5 propriétés : nous avons vu dans la section précédente un cas où l'on avait $p + q$ et $r + q$ sans avoir $p \vee r \mid\!\sim q$, donc un bon contre-exemple. Pour éliminer de tels contre-exemples, c'est-à-dire, dans notre interprétation, pour que les éléments typiques de $A \cup B$ soient (dans) l'union des éléments typiques de A et des éléments typiques de B, Kraus & al. (1990) ont montré qu'il fallait n'étiqueter les états que par un seul monde.

Mais plutôt que de recourir à un dénombrement des sens où à la disjonction, il semble plus adéquat de fournir pour chaque mot, dans la mesure du possible, <u>une</u> acception privilégiée (p.ex. la fosse typique est une cavité destinée à recevoir quelque chose (grossièrement : fosse(x) $\vdash$ cavité(x) $\wedge \exists$y but (x, recevoir, y)) ; il suffit alors d'exprimer le fait qu'une tombe est une fosse dans laquelle ce quelque chose est un cadavre (tombe(x) $\vdash$ fosse(x) $\wedge \exists$y (cadavre(y) $\wedge$ but (x, recevoir, y))).

Cependant, pour éviter les paradoxes mentionnés, il faudra ajouter des prémisses négatives : si l'on accepte qu'une ouverture typique soit un passage (ouverture(x) $\vdash$ passage(x)), et qu'un trou typique soit une ouverture (trou(x) $\vdash$ ouverture(x)), il est utile de préciser que lorsqu'un trou est creusé dans une substance, il n'est pas typique qu'il traverse cette substance ; on peut aller jusqu'à exprimer la propriété contraire : trou(x) $\wedge$ dans(x, y) $\vdash$ $\neg$traverse(x, y) et se donner comme définition d'un passage le fait que celui-ci traverse quelque chose : passage(x) $\supset$ $\exists$y traverse(x, y). En traduisant le fait, signalé par de nombreux dictionnaires, qu'une fosse est un trou dans la terre, on arrive à la conclusion qu'une fosse n'est pas un trou typique, et donc pas nécessairement un passage[3].

Élaboration de la relation de déductibilité

Plus récemment, Lehman & Magidor (1992) ont étudié 3 nouveaux comportements déductifs, chacun strictement plus "fort" que le précédent, et le plus "faible" strictement plus fort que ceux qu'ils avaient envisagés en 1990. Il n'est pas sans intérêt de regarder l'adéquation de ces comportements au phénomène de description lexicale qui nous occupe ici. On suppose une relation de déductibilité obéissant aux 5 principes déjà énoncés et à :

1. la "*négation rationnelle*" : si ni de A$\wedge$B, ni de A$\wedge\neg$B on ne peut déduire C, c'est que de A seul, on ne peut pas déduire C. Dans notre interprétation, ceci se traduit par : "si les éléments typiques de A $\cap$ B ne sont pas tous des C, et que les éléments typiques de A $\cap$ $\neg$B non plus, alors les A typiques ne sont pas tous des C."

Il semble que cette hypothèse soit un peu forte : on peut vouloir exprimer que les êtres humains sont typiquement "égoïstes/altruistes", sans dire ni que l'homme typique, ni que la femme typique l'est (A = humain, B = féminin, C = "égoïstes/altruistes"). Mais on admettra également que de tels cas sont rares et que l'hypothèse n'est pas inacceptable ;

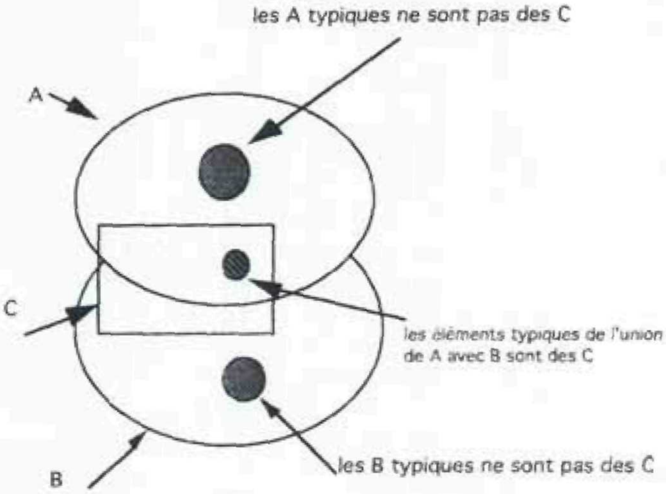
2. la "*disjonction rationnelle*" : si ni de A ni de B on ne peut déduire C, c'est qu'on ne peut pas le déduire de A $\vee$ B. Le même cas de figure (avec la même remarque sur son aspect un peu sollicité) sert ici aussi de contre-exemple, ce qui n'a rien d'étonnant puisque les auteurs ont montré que cette hypothèse impliquait la précédente.

La démonstration en est très simple : plaçons-nous dans les hypothèses de *négation rationnelle* : ni de A $\wedge$ B, ni de A $\wedge$ $\neg$B on ne peut déduire C. Par *disjonction ration-*

[3] Les formules données ici ne permettent toutefois pas de conclure qu'une fosse n'est pas un passage : il faudrait se donner les moyens de montrer non seulement qu'un trou dans la terre ne traverse pas la terre (!), mais qu'il ne traverse aucune autre substance !

nelle on ne peut pas le déduire non plus de (A ∧ B) ∨ (A ∧ ¬B) qui est logiquement équivalent à A seul. La propriété d'*équivalence à gauche* étant supposée respectée par tous ces nouveaux systèmes, C n'est donc pas déductible de A seul.

Comme nous l'avons dit, l'implication entre *disjonction rationnelle* et *négation rationnelle* n'est pas une équivalence. On pourrait donc trouver de nouveaux contre-exemples, dont le schéma serait le suivant :



Pour arriver à construire un contre-exemple sur ce modèle, il faudrait que l'union de A avec B ait des éléments typiques, donc vraisemblablement qu'elle soit lexicalisée, disons par U. Il ne semble pas possible dans la langue que B ne se définisse pas par "les U qui ne sont pas A", et dans ce cas, la disjonction se réduit à la *négation rationnelle* : si ni de U ∧ A (≡ A), ni de U ∧ ¬A (≡B) on ne peut déduire C, c'est que de U seul (≡ A ∨ B) on ne peut déduire C.

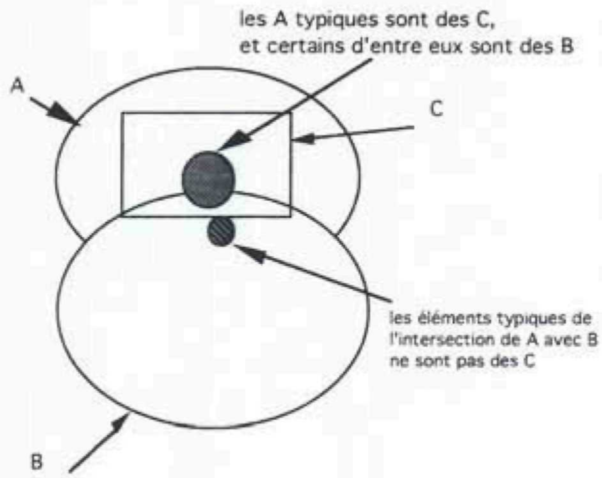
3. "*Monotonie rationnelle*" : si de A ∧ Bon ne peut pas déduire C, mais que de A seul on ne peut pas déduire ¬B, c'est que de A seul, on ne peut pas déduire C. Là aussi, il est facile de voir que les mêmes contre-exemples s'appliquent car, comme le montrent les auteurs, cette propriété est plus forte que la précédente.

Plaçons-nous dans les hypothèses de *disjonction rationnelle* : ni de A ni de B on ne peut déduire C. A est logiquement équivalent à (A ∨ B) ∧ A, et *équivalence à gauche* est toujours considérée comme vérifiée. Deux cas se présentent :

- si de A ∨ B, on ne peut pas déduire ¬A, alors par *monotonie rationnelle*, de A ∨ B, on ne peut pas déduire C ;

- si de $A \vee B$, on peut déduire $\neg A$, donc, par *affaiblissement à droite*, $\neg A \vee B$; supposons que de $A \vee B$ on puisse déduire C et montrons que cela conduit à une contradiction : par *monotonie prudente*, on conclurait que de $(A \vee B) \wedge (\neg A \vee B)$, on peut déduire C ; or $(A \vee B) \wedge (\neg A \vee B) \equiv B$: par *équivalence à gauche*, on conclurait donc que de B, on peut déduire C, ce qui est contraire aux hypothèses.
Donc dans les deux cas, de $A \vee B$, on ne peut pas déduire C.

Un contre-exemple à cette propriété aurait l'aspect suivant :



Un exemple qui peut venir à l'idée pour illustrer le schéma serait le suivant : A = compositeur ; si l'on demande de citer les "compositeurs typiques", il est probable que l'on recueille les noms de Mozart, Beethoven, Chopin, Bach, etc. et que l'on accepte que les compositeurs typiques aient consigné leurs œuvres par écrit (C) ; mais on accepte aussi qu'il y ait des compositeurs non-européens (B), et, même si on ignore leur nom, qu'il y en existe de "typiques" dont on ne s'attend pas à ce qu'ils aient écrit de partitions ; ceci donne une situation où de $A \wedge B$ (compositeur non-européen) on ne veut pas déduire C (œuvres écrites), de A (compositeur), on ne veut pas déduire $\neg B$ (européen), et pourtant de A (compositeur), on accepte de déduire C (œuvres écrites).

Discussion et Conclusion

Les travaux sur la relation de déductibilité que nous venons de passer en revue semblent assez bien répondre à notre souci d'éliminer des paradoxes induits par une représentation en logique classique des définitions lexicales, même si quelques points contestables mais relativement marginaux subsistent.

Nous avons vu cependant qu'ils correspondaient à une vision "absolue" de la préférence, ce qui nous a semblé un choix difficile à justifier sur le plan des principes, mais dont les conséquences pratiques n'apparaissent pas dans les exemples que nous avons envisagés jusqu'ici. Considérons maintenant un autre aspect de la connaissance lexicale : l'ambiguïté, qui permet souvent de multiples "lectures" d'un même énoncé.

La relation de déductibilité modélisée par Kraus & al. (1990) puis par Lehmann & Magidor (1992) considère implicitement qu'à partir d'un ensemble fixé de prémisses, il y a unicité de l'ensemble des conséquences déductibles. Or cette contrainte est incompatible avec la possibilité d'interpréter un texte de différentes manières. Si l'on veut rétablir cette possibilité, on peut envisager de représenter différents "points de vue" en indiquant la relation de déductibilité. Chaque point de vue permettrait de synthétiser une préférence différente, ce qui revient à relativiser la notion de préférence. Toutefois, associer à chaque relation de déductibilité (du genre ouverture(x) $\vdash$ passage(x)) un indice d'une façon cohérente, établir le cas échéant des hiérarchies sur les indices (p.ex. tel point de vue est une restriction de tel autre), et gérer convenablement cet ensemble de relations et d'indices conduit à de redoutables difficultés théoriques et pratiques.

Or il se trouve que la pluralité que nous recherchons s'intègre sans recourir à des mécanismes spécifiques, dans plusieurs autres approches de la non-monotonie, les plus élaborées étant la logique des défauts de Reiter (1980) et la logique auto-épistémique de Moore (1985) entre lesquelles des correspondances profondes existent (Konolige 1988).

En effet, l'expression de la conséquence typique s'écrit, par exemple chez Reiter dans le cas considéré : \F(ouverture(x) : **M**passage(x) ; passage(x)) où **M** est un opérateur modal dont la signification intuitive est "en l'absence d'une preuve du contraire". Pour éviter des paradoxes dans l'interprétation de cet opérateur, Reiter donne une construction technique utilisant la notion de point fixe, qui permet de définir ce qu'est l'**extension** d'un ensemble de formules. Cette notion correspond à celle d'ensemble de conséquences déductibles à partir de prémisses données. Il se trouve que dans les cas intéressants, il y a plusieurs points fixes, donc plusieurs extensions, et ceci correspond à ce que nous avions annoncé, à savoir une pluralité des "lectures" possibles d'un même ensemble de départ.

Nous avons tenté d'exprimer dans ce formalisme, légèrement modifié, des connaissances lexicales permettant d'analyser des phénomènes de glissements de sens (Kayser 1992, Kayser & Abir, à paraître). Il se trouve que la principale difficulté du système que nous proposons dans ces articles réside dans la nécessité d'avoir des techniques drastiques d'élimination des extensions, si on ne veut pas voir "exploser" le nombre de "lectures" d'une phrase.

Dans ces conditions, il pouvait apparaître opportun d'examiner de plus près, du point de vue de la connaissance lexicale, le récent travail de Kraus, Lehmann et Magidor sur la relation de déductibilité -et tel est le but de la présente étude-.

Au terme de cet examen, il nous semble possible d'affirmer son intérêt pour représenter d'une façon simple certains phénomènes comme la non-transitivité de la relation d'hypéronymie. Cependant, cette simplicité a pour contrepartie le refus d'envisager la pluralité des interprétations possibles, or ce phénomène est inhérent à la langue et il doit être pris en compte dès le niveau lexical.

C'est pourquoi, malgré ses difficultés propres, nous considérons qu'un système basé sur une logique multi-extensionnelle comme celle de Reiter est plus adéquat à la représentation de la connaissance lexicale.

Références bibliographiques

CHIERCHIA, G. & MCCONNELL-GINET S. (1990), *Meaning and Grammar. An Introduction to Semantics*, M.I.T. Press.

GAMUT (1991)[4] *Logic, Language and Meaning (vol.1 Introduction to Logic ; vol.2 Intensional Logic and Logical Grammar)*, The University of Chicago Press.

KAYSER, D. (1990), "Adéquation et inadéquation de la logique au traitement sémantique des langues", *Modèles Linguistiques* 12, 119-136.

KAYSER, D. (1992), "Sur les glissements de sens lexicaux" in TYVAERT, J.E. (ed.) (1992) *Lexique et Inférence(s)*, Recherches Linguistiques XVIII, 231-246.

[4] L.T.F.GAMUT, pseudonyme pour J.F.A.K.Van Benthem, J.A.G.Groenendijk, D.H.J.de Jong, M.J.B.Stokhof, H.J.Verkuyl.

KAYSER, D. & ABIR, H. (à paraître), "A Non-monotonic Approach to Lexical Semantics", in *Computational Lexical Semantics*, Cambridge University Press.

KONOLIGE, K. (1988), "On the relationship between Default Logic and Autoepistemic Logic", *Artificial Intelligence* 35, 343-382.

KRAUS, S., LEHMANN D. & MAGIDOR M. (1990), "Non-monotonic Reasoning, Preferential Models, and Cumulative Logics", *Artificial Intelligence* 44, 167-207.

LEHMANN, D. & M. MAGIDOR (1992), "What does a conditional knowledge base entail ?", *A. I. Journal* 55, 1-60.

PUTNAM, H. (1970), "Is Semantics Possible", in KIEFER, H.E. & MUNITZ M.K. (eds.) (1970), *Language, Belief, and Metaphysics*, State University of New York Press.[5]

MARTIN, R. (1990), "La définition «naturelle»", in *La Définition*, Larousse, Langue et Langage, 86-95.

MOORE, R.C. (1985), "Semantical Considerations on Non-Monotonic Logic", *A.I. Journal* 25, 75-94.

PERRIN, A. (1989), *Rapport de DEA d'Intelligence Artificielle* , Université Paris-Nord.

REITER, R. (1980), "A Logic for Default Reasoning", *Artificial Intelligence* 13, 81-132.

VÉRONIS, J., IDE N.M. & WURBEL N. (1989), "Extraction d'informations sémantiques dans les dictionnaires courants", *Actes du 7ème RFIA*, 1381-1395.

WILKS, Y. (1977), *Good and bad arguments about semantic primitives*, Technical Report 42, Dept. of Artificial Intelligence, University of Edinburgh.

[5] Trad. fr. de Jean-Marie MARANDIN : "La sémantique est-elle possible", in *La Définition*, Larousse, Langue et Langage, 1990, 292-304.