

Les sémèmes permettent de raisonner

Une des utilités de l'analyse sémique est de permettre la génération automatique de taxinomies dans lesquelles les lexèmes sont ordonnés en hyperonymes et hyponymes.

Par définition, un hyperonyme possède un sémème qui est un sous-ensemble du sémème de son ou de ses hyponymes. Sémantiquement, l'hyponyme est lié à son hyperonyme par une relation « sorte de... », relation transitive.

L'analyse sémique permet de construire des raisonnements simples et de répondre à des questions sur un texte ou une domaine de connaissance pour lequel aura été constitué un dictionnaire sémique ou base de connaissances sémantique.

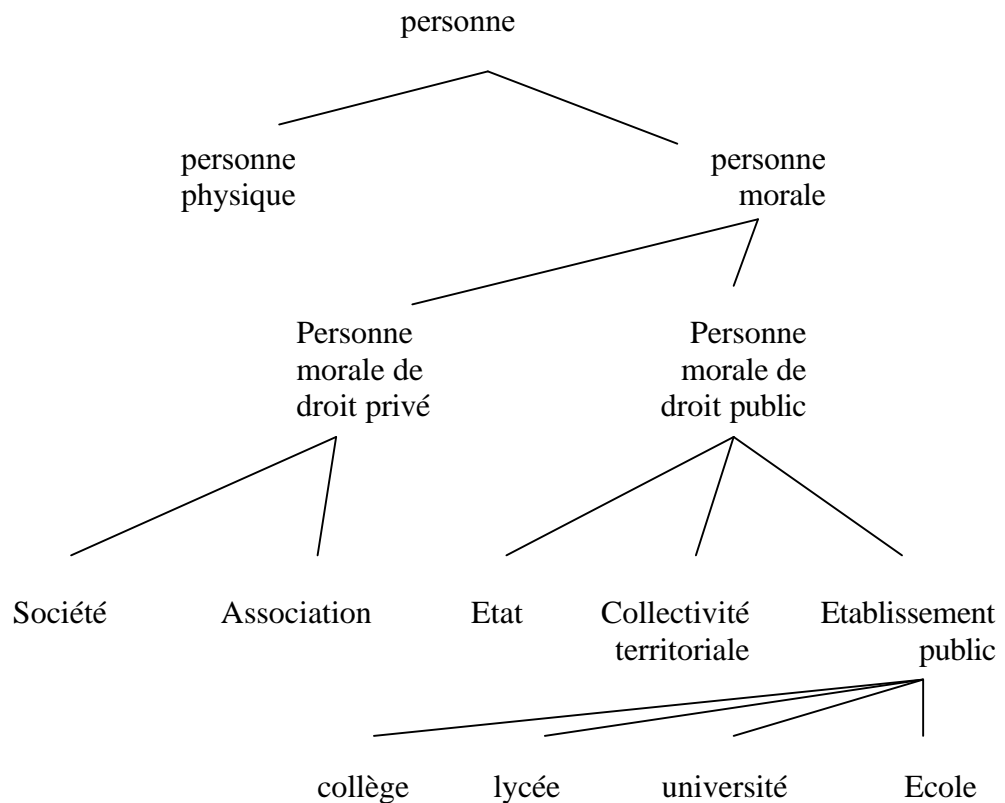
Si, par exemple, nous définissons une université comme étant un établissement public, nous pouvons en inférer qu'une université possède la personnalité morale du fait que :

- l'établissement public est une personne morale de droit public
- le fait d'être une personne morale entraîne que l'on possède la qualité de personne morale.

Ceci se déduit sans difficulté de l'analyse sémique.

La question étant « une université est-elle une personnalité morale? », la réponse à cette question est subordonnée à la réalisation de trois conditions :

- si la base de connaissance sémantique a été construite de façon à contenir non seulement le sémème d'établissement public, mais aussi son ensemble de définition, dont nous donnons un aperçu ci-après, cette base de connaissance comporte toutes les informations permettant de répondre à la question;
- il convient de disposer d'un algorithme de recherche adéquat
- le système doit être en mesure de comprendre le sens de la question? Cet aspect est d'une grande importance, mais, bien que lié à notre recherche, il n'en constitue pas l'élément principal.



S'agissant de l'algorithme de recherche, pour arriver à faire le lien entre personne morale et université, le système doit d'abord rechercher dans le texte si celui-ci contient la réponse. A cet égard, l'analyse de la question montre que nous sommes au regard de la théorie des voix (cf. p. **Erreur ! Signet non défini.**) en présence d'un *équatif* et donc nous pouvons poser sur la base de données un filtre qui ne sélectionne que les relations de type équatif. Ainsi, il est inutile de traiter une relation qui dirait que l'université est habilitée à délivrer des diplômes, qui est une relation de type *descriptif*. En cas d'échec, il convient d'interroger les taxinomies en partant soit de personne pour arriver à université, soit d'université pour arriver à personne morale.

Selon le premier cheminement, il faut d'abord naviguer dans la base de connaissance en parcourant tous les nœuds de l'arborescence de l'ensemble de définition contenant « personne morale » dans le sens descendant, jusqu'à trouver « université ».

L'autre cheminement possible est de partir de « université », et d'interroger la base de connaissance sémantique et de parcourir l'arbre dans le sens ascendant jusqu'à trouver "personne morale".

Il est plus facile d'interroger la taxinomie à partir de « personne morale » qu'à partir de « université » parce que « université » peut appartenir à plusieurs taxinomies. Par exemple, université peut appartenir aussi bien à la taxinomie qui détaille toutes les formes de personnalité juridique, qu'à la taxinomie qui classe les différents types d'établissements d'enseignement, non pas en raison de leur statut juridique, mais en raison de leur fonction pédagogique. Donc, il convient de conduire la recherche de préférence à partir de l'hyperonyme et non de l'hyponyme.

Donc, nous pensons que le fait d'identifier dans la question un certain type de relation est un élément de compréhension de la question essentiel pour conduire ensuite la recherche avec efficacité.

On notera toutefois que si la question est libellée « L'université a-t-elle la personnalité morale », il ne s'agit pas d'une relation équatative, mais d'une relation descriptive de la seconde forme, et qu'il n'est pas possible de rattacher directement cette formulation à une relation du texte qui dirait « l'université est un établissement public ». Il faudrait que le texte dise « l'université a la qualité d'un établissement public » pour que l'algorithme que nous venons de décrire soit opératoire. Il est donc nécessaire dans cet exemple d'opérer sur le libellé initial de la question une transformation dans la voix équatative. Et comme on ne sait jamais à l'avance la variante utilisée dans le texte, il sera toujours nécessaire, en cas d'échec de la première recherche, de reformuler la question

Cet exemple peut être analysé autrement. L'expression « avoir la qualité de » peut être interprétée comme caractéristique d'un équatif et non d'un descriptif. Auquel cas, la reformulation qui doit être appliquée consiste à rechercher les synonymies au sein des relations caractéristiques d'une voix donnée. Dans notre exemple où nous avons le module « SN être SN », la synonymie (relative) entre « être » et « avoir la qualité de » dans ce contexte est facile à établir. Quoi qu'il en soit, pour que la recherche soit une recherche intelligente, il est nécessaire de prévoir d'opérer sur le texte même de la question différentes transformations en cas d'échec de la recherche directe. Et il est évident que le nombre de transformations possibles s'accroît avec la longueur de la question de manière exponentielle. Les transformations pourront porter sur différentes visions au sein du module, mais aussi sur la synonymie et les variantes paraphrastiques éventuelles. Il s'agit

d'une combinatoire redoutable qui ne nous intéresse pas directement ici mais qui constitue une des difficultés majeures du traitement en langue naturelle ici envisagé.

Nous venons de voir un exemple dans lequel la recherche doit parcourir le taxème dans le sens ascendant ou dans le sens descendant et où la réponse dépend de la découverte sur ce parcours d'un terme existant dans le texte.

Mais nous pouvons imaginer que le terme de la question et le terme de la réponse se trouvent dans l'ensemble de définition, mais sur des branches différentes de cet ensemble. Imaginons que la question soit « l'université est-elle une association? ». L'association est une personne morale, mais une personne morale de droit privé. La réponse pourrait être simplement « non, l'université n'est pas une association », ce qui traduirait l'échec de la recherche. Mais si le programme constate que « établissement public » fait partie du même ensemble de définition, mais qu'ils diffèrent au moins par les sèmes « public » et « privé », la réponse pourrait être « non, l'université est un établissement public ».

Il convient de souligner ici un aspect très important découlant de l'analyse sémique et de la notion d'ensemble de définition ou taxème. L'appartenance au même ensemble de définition permet de répondre « non, l'université est un établissement public », mais interdirait de répondre « non, l'université est une institution de formation », parce que « institution de formation » ne fait pas partie du même ensemble de définition que « association ».

On notera toutefois que la structure sémantique que l'on peut extraire de l'analyse des sèmes fait apparaître des relations diverses qu'il convient d'identifier.

A la base, les éléments d'une même taxinomie sont unis par une relation « sorte de », « est un », « est une espèce de », etc., ce qui est le cas de « personne », qui en droit peut se décomposer en « personne physique » et « personne morale », « personne morale » pouvant se décomposer en « personne morale de droit public » et « personne morale de droit privé », etc. (voir schéma).

Toutefois, si l'on prend « personnalité », on ne pourra dire que « personnalité » appartient à la catégorie de « personne ». Selon des règles qui s'enseignent aujourd'hui à l'école primaire, « personnalité » s'obtient par dérivation de « personne » par l'ajout d'un suffixe porteur d'un sème « ité » qui veut dire « qui a la qualité de ».

Autrement dit, à partir d'un taxème donné, on devra construire un arbre comprenant non seulement les sous-catégories du lexème racine, mais aussi les différents lexèmes construits par dérivation ou par composition (« reconstruire qui s'obtient par composition à partir de « construire »).

Egalement, il convient d'inclure dans la même démarche les verbes ou les notions contraires, ceux que Gérard Sabah (opus cit. p.95) appelle les verbes converses qui forment des couples tels que : vendre-acheter, apprendre-enseigner, construire-démolir, etc. Ces cas dont l'analyse peut être systématisée relèvent tous des modèles sémantiques fondamentaux évoqués plus haut.

Allons plus loin. Si le taxème comporte dans son sème la notion d'un ensemble composé d'éléments, on devra en déduire de l'appartenance à cet ensemble un certain nombre de caractéristiques héritées de l'ensemble de niveau supérieur. Prenons par

exemple le taxème « nation », la « nationalité » qui est une qualité attachée à « nation » se reporte sur chacun des éléments composants cette nation. Ainsi, un citoyen de la nation française possède la « nationalité française ». Il s'agit d'une forme d'héritage distincte de l'héritage découlant de la relation "sorte de ...".

Il est bien évident que, compte tenu de ce qui vient d'être dit, les cas de synonymie, mais aussi de paraphrase, sont susceptibles d'être traités au travers de l'analyse sémique. Quand on lit « le service public de l'éducation contribue à l'égalité des chances », et si l'on pose la question « le service public favorise-t-il l'égalité des chances? », la forte ressemblance au niveau du sémème entre « contribuer à » et « favoriser », doit conduire à une réponse intelligente, c'est-à-dire positive.

La question évidemment délicate à laquelle nous souhaitons pouvoir répondre est de savoir à partir de quel moment ou jusqu'à quel point on peut considérer que deux lexèmes sont synonymes ou quasi synonymes, les sèmes qui les distinguent éventuellement pouvant être négligés.

Nous nous plaçons ici dans une perspective uniquement sémasiologique orientée vers l'interprétation des questions posées et non dans une perspective onomasiologique orientée vers la production d'énoncés en relations mutuelles de paraphrase. La production d'énoncés obéit à une combinatoire rapidement incontrôlable. Cependant, nous ne voyons pas d'autres moyens de générer des énoncés voulant dire à peu près la même chose que de partir d'ensembles de sèmes tels que si je veux signifier « qui mange des végétaux », j'ai le choix entre « qui mange des végétaux » et « herbivore ». On doit en effet partir de ce que l'on veut dire pour, à la suite de différents choix sémantiques puis syntaxiques, produire l'énoncé final. Il s'agit ici d'une perspective qui sort du champ immédiat de notre recherche, mais il nous apparaît que seul l'analyse sémique permet de progresser dans cette direction.

Les précédentes remarques suggèrent l'idée que l'analyse sémique est susceptible de jouer un rôle de reconnaissance du sens, par analogie avec la notion de reconnaissance des formes. Des syntagmes, dont on a vu qu'ils pouvaient par l'effet des transferts de catégorie recouvrir des ensembles complexes, sont ainsi reconnaissables à leur sémème. Il faut par exemple que deux énoncés syntaxiquement différents comme « j'entends hurler le vent » et « j'entends le vent hurler » puissent être interprétés comme des énoncés proches sinon identiques, ne serait-ce que parce que l'inversion entre « hurler » et « le vent » induit une différence de pondération sémantiquement significative mais que l'on peut vouloir négliger pour éviter de les considérer comme des énoncés radicalement différents.

Une hypothèse qui pourrait être étudiée et qui dépasse également le champ de cette étude, serait d'utiliser l'analyse sémique à des fins d'indexation.

Dans une recherche bibliographique, au lieu de procéder par mots clés, liés à toute une série de synonymes ou de mots apparentés, on devrait pouvoir produire le ou les sémèmes composant l'interrogation et rechercher dans le texte des sémèmes identiques ou voisins.

Cette démarche devrait être comparée aux recherches actuelles qui, voulant dépasser le stade quelque peu limité des mots clés, et dans le but de mieux cibler les recherches, partent de relations syntaxiques incluses dans la question posée et recherchent dans le texte indexé des relations similaires (cf. Annie Coret, Bruno Menon, Danièle Schibler, Christophe Terrasse, 1994; Anne-Marie Guimier-Sorbets, 1993).

On peut aussi se demander si les traits logiques qui caractérisent certaines relations, et à ce titre, ont a priori leur place parmi les sèmes, ne permettent pas de construire d'authentiques raisonnements.

Les exemples que nous avons utilisés jusqu'ici reposaient sur une relation logiquement transitive (à différencier de la transitivité au sens syntaxique) dans la mesure où si l'on pose que si l'université est un établissement public et que si l'établissement public est une personne morale de droit public, on peut en déduire que l'université est une personne morale de droit public.

Des opérations logiques sont également possibles par utilisation d'autres traits logiques des relations.

Prenons par exemple la relation « être marié à ». Cette relation est symétrique parce que si Pierre est marié à Jeanne, Jeanne est mariée à Pierre. Mais la relation « être l'épouse de » est antisymétrique, sans qu'il y ait besoin de le démontrer. Mais dans la mesure où la relation « être l'épouse de » implique la relation « être marié à » ou « être le conjoint de », de la même manière que « épouse » a pour hyperonyme « conjoint » ou « marié », mais comporte en outre le sème « féminin », on peut en inférer que si Jeanne est mariée à Pierre et que comme Jeanne est une femme, alors Jeanne est l'épouse de Pierre. Ceci dans un contexte où l'on a le droit de poser comme règle de gestion que la polygamie n'étant pas admise, Jeanne est bien l'épouse de Pierre et non *une* épouse de Pierre. On voit dans cet exemple intervenir des règles implicites dans l'énonciation, mais nécessairement explicites dans tout traitement, tandis que la déduction que Jeanne est l'épouse de Pierre se déduit directement des sémèmes.

Le fait que la relation « être marié à » soit symétrique rend impossible la double affirmation « Jean est marié à Jeanne » et « Jeanne est marié à Paul ».

Le fait que la relation « être marié à » soit intransitive interdirait a fortiori de déduire de ces deux propositions que Jean est marié à Paul. (Brian Bowen et Pavel Kocura, 1993)

Dans tous les cas qui précèdent, nous avons fait appel aux taxinomies associées soit aux substantifs en relations mutuelles d'équivalence, soit aux relations elles-mêmes.

Mais l'utilisation des taxinomies n'est pas toujours possible. Si la question est « les universités sont-elles habilitées à délivrer les diplômes ? », aucune taxinomie qui reposerait sur une relation de base « sorte de... », ne peut contenir la réponse. La seule ressource est d'interroger l'ensemble des relations contenues dans le texte ayant pour base « université ». Nous utiliserons néanmoins un filtre, la question étant fondée sur une relation descriptive. Seules les relations descriptives devront donc être examinées.

Nous pensons donc que l'analyse sémique donne les moyens d'élaborer des raisonnements et nous regrettons de n'être pas en mesure de pousser plus loin dans cette voie car nous sortirions du champ plus restreint de notre recherche.

Toutefois, par un dernier exemple, il peut être utile de montrer les limites du traitement logique lui-même.

Dans une étude très profonde, déjà évoquée (cf. **Erreur ! Signet non défini.**), Paul Amselek (1992) dans le but d'illustrer les limites de la logique déontique, donne l'exemple de la règle suivante : « il est interdit de marcher sur les pelouses ». Si l'on

considère que ce qui n'est pas interdit est autorisé, un traitement purement logique permettrait d'affirmer qu'il est autorisé de faire de la moto sur les pelouses.

Or, si les règlements dans les jardins publics ne posent que cette règle simple, c'est parce que précisément elle est simple, courte et directement compréhensible par nos concitoyens qui, devant cette interdiction, n'auraient pas l'idée, sauf par malveillance ou pour défier les règlements, de faire de la moto sur les pelouses.

En fait, il faut comprendre que cette réglementation a un but de protection de la pelouse qui est un objet fragile susceptible de détérioration. Alors que « marcher » sur un trottoir n'implique aucune idée de détérioration, « marcher sur une pelouse » peut impliquer une telle idée, sans que l'on puisse toujours l'affirmer, toute pelouse n'étant ni nécessairement fragile, ni nécessairement protégée. Mais si une pelouse fait ainsi l'objet d'une mesure de protection, toute activité plus néfaste que le simple fait de marcher sur la pelouse doit logiquement être également interdite sans qu'il soit nécessaire de l'écrire. Qui interdit le moins, interdit également le plus.

Autrement dit, si la logique déontique a tort, ce n'est pas parce que la logique est intrinsèquement incapable de résoudre un tel cas, mais parce que la règle logique que l'on voudrait y appliquer n'est pas la bonne. Dès lors, la question devient de savoir comment déterminer la bonne règle.

En fait, l'interdiction en droit est généralement associée à une idée de nuisance à autrui, de nuisance à un bien fragile et protégé ou de danger. Dans ces hypothèses, interdire le moins, implique que le plus est également interdit. La règle selon laquelle tout ce qui n'est pas interdit est autorisé demeure dans tous les cas où la règle précédente ne s'applique pas. La question devient de savoir comment déterminer le plus par rapport au moins. Comment faire, sémantiquement parlant, pour comprendre qu'il est au moins aussi nuisible pour la pelouse d'y faire de la moto que de la piétiner. Il est bien clair que l'analyse sémique en contexte nous donne la réponse, mais il faut convenir que cette analyse échappe à toute possibilité d'automatisation, à moins d'introduire dans le sémantème de moto le sème « détériore les sols fragiles », et dans le sémantème de « pelouse » le sème « sol fragile ». Dans ce cas, l'interpréteur devrait établir le lien et, appliquant la règle précédente, donner une réponse satisfaisante. Mais, il n'y a aucune raison de ne pas étendre cette démarche à tous les cas les plus farfelus qui pourraient se présenter. Ainsi, il ne devrait pas non plus être possible de faire du vélo sur les pelouses, de la patinette, de la voiture à pédales, voire d'y déposer des objets lourds et encombrants. Malgré la difficulté de prévoir toutes les situations possibles, notons toutefois que c'est précisément la difficulté à laquelle se trouve confrontée toute autorité investie d'un pouvoir de réglementation. Par ailleurs, le raisonnement que nous venons d'amorcer est précisément celui que tout juge est conduit à faire devant une situation de ce type : d'abord apprécier les faits (le caractère de nuisance ou non de l'acte ou de l'activité incriminée) et ensuite déterminer la règle applicable.

Justement, peut-on envisager de conférer à un programme intelligent une capacité d'appréciation des faits égale à celle d'un juge. On peut en douter.

Cette discussion nous paraissait utile bien qu'elle se situe aux marges de notre recherche aux objectifs beaucoup plus modestes.