

Caractéristiques sémantiques de quelques “ Noms scientifiques ” dans l’article de recherche en français

Cristelle Cavalla, Francis Grossmann

► **To cite this version:**

Cristelle Cavalla, Francis Grossmann. Caractéristiques sémantiques de quelques “ Noms scientifiques ” dans l’article de recherche en français. Torodd Kinn. Akademisk prosa, Skrifter fra KIAP Romansk institutt, Universitetet i Bergen, pp.47-59, 2005. hal-00394208

HAL Id: hal-00394208

<https://hal.archives-ouvertes.fr/hal-00394208>

Submitted on 10 Jun 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

(2005), Torodd Kin (Eds). *Akademisk prosa*. Bergen, Skrifter fra KIAP Romansk institutt, Universitetet i Bergen, 47-59.

Caractéristiques sémantiques de quelques « Noms scientifiques » dans l'article de recherche en français

Cristelle Cavalla, Francis Grossmann

Lidilem, Université Stendhal-Grenoble³

1. Introduction

Ce travail fait partie d'un projet *Emergence*¹ dans lequel il s'agit d'étudier la phraséologie 'académique' spécifique de l'article scientifique, dans différentes spécialités. L'un des objectifs à terme, est la constitution d'un outil d'aide à la rédaction d'articles scientifiques (anglais – français).

Le lexique 'académique' et/ou 'transdisciplinaire' – en d'autres termes, le lexique commun à différentes disciplines scientifiques² –, n'est pas toujours facile à définir : certains termes peuvent en effet, sembler proches du lexique général, tandis que d'autres peuvent parfois être spécifiques d'une ou de plusieurs disciplines, et il faut se résoudre au fait qu'il s'agit en définitive d'un ensemble flou. Malgré ces difficultés, une étude de ce lexique se révèle intéressante à plusieurs égards. D'abord, il est parfois pertinent, de contraster le fonctionnement dans différents champs disciplinaires, en pointant les similitudes et les dissemblances dans l'usage. Ensuite, on peut aussi faire réfléchir les chercheurs néophytes sur l'emploi de certains de ces termes, qui fonctionnent comme des emblèmes de scientificité : utiliser la collocation 'faire une hypothèse', c'est bien d'une certaine façon, revendiquer son appartenance au champ scientifique, même si ce que l'on appelle *hypothèse* peut varier sensiblement selon les traditions épistémologiques, les cultures nationales ainsi que les formes d'écritures scientifiques.

L'objectif de la présente étude se veut cependant plus limité : dans la lignée d'études antérieures portant sur la phénomène collocatif (Tutin et Grossmann, 2002, Grossmann et Tutin, 2003), nous étudierons les collocations verbales de quelques unes de ces lexies (*hypothèse, postulat, thèse et théorie*) dans l'usage français en vue d'une étude contrastive ultérieure anglais-français. Il s'agit d'intégrer les noms étudiés dans des classes sémantiques qui permettraient de prédire, sans trop

¹ Acquisition automatique de traductions d'expressions semi-figées pour l'élaboration d'outils d'aide à la rédaction scientifique et technique. Coordinateurs : Olivier Kraif, Agnès Tutin (LIDILEM, Université Stendhal). Projet financé par la Région Rhône-Alpes.

² Voir, pour le français, les travaux concernant le lexique scientifique menés par Phal (1971).

d'erreurs, les principaux verbes associables à ces dimensions. Ainsi, en vue de l'outil d'aide à la rédaction en anglais pour des francophones et en français pour des anglophones, il est important de modéliser les collocations visées afin d'avoir toutes les correspondances possibles dans l'une et l'autre langue.

Nos objectifs immédiats sont d'observer les co-occurents verbaux [V + N] des quatre noms sélectionnés afin de préciser leur usage dans l'article scientifique. Nous cherchons à vérifier si ces noms partagent certaines caractéristiques sémantiques et si leur usage diffère en fonction de traits sémantiques ou de dimensions sémantiques privilégiées.

2. Démarche adoptée et aspects méthodologiques

Notre démarche s'est effectuée selon deux phases : dans un premier temps nous avons travaillé à partir de définitions de dictionnaire et dans un deuxième temps sur un corpus d'articles scientifiques rédigés en français.

2.1. L'analyse lexicologique

La première partie du travail, l'analyse lexicologique, a consisté à dégager les traits sémantiques (Rastier, 1996) de chaque lexie à l'aide de sa définition prise dans le *Trésor de la Langue Française* (1977 ; désormais abrégé TLF).

L'étude du sens de chaque lexie, à l'aide des définitions, nous a permis d'effectuer une désambiguïsation sémantique afin de sélectionner les acceptions pertinentes pour notre corpus. De ce fait, nous observerons la multiplicité des utilisations d'une même lexie en contexte ; par exemple, le nom *thèse* peut aussi bien être considéré sous sa facette épistémologique ('la thèse X appartient à tel ou tel paradigme scientifique') que sous sa facette argumentative ('une thèse peut se discuter, être infirmée', etc.). L'analyse micro-sémantique des définitions aide à la mise en exergue des traits sémantiques des lexies ainsi qu'au regroupement de ces derniers au sein de dimensions sémantiques. Les exemples utilisés par le TLF ont servi également de corpus d'appoint pour la vérification des co-occurents privilégiés.

2.2. L'étude de corpus

L'étude portant sur le corpus d'articles a permis la recherche et le tri des co-occurents verbaux à partir des concordances. Le corpus utilisé a été constitué de différents corpus existants :

- a. Un corpus d'articles de linguistique³
- b. Un corpus d'articles de sciences humaines (Revue « Per Hermès »)
- c. Le corpus de KIAP (des articles de linguistique, d'économie et de médecine)

Le corpus entier représente environ 1 700 000 mots. Contrairement à nos attentes, le nombre d'occurrences des collocations [V + Nsc] s'est révélé encore insuffisant pour une étude quantitative ; il permet déjà, cependant, d'observer des tendances qui doivent être confirmées par l'examen d'un corpus plus vaste.

Par ailleurs, le choix des quatre noms s'est effectué en fonction d'un double critère : il fallait qu'ils soient suffisamment proches, au plan de leur fonctionnement sémantique et syntaxique, pour qu'une comparaison se révèle intéressante. Mais il fallait aussi qu'il puisse y avoir des différences suffisantes pour que l'on puisse observer d'éventuelles incompatibilités dans la sélection des verbes qu'ils admettaient comme arguments ; c'est ce qui justifie par exemple le choix de *théorie*, par opposition aux noms hypothèse et postulat ; *thèse* a été choisi parce que ce lexème semblait représenter un cas limite – mais intéressant – entre lexique 'méthodologique' et lexique général. Enfin, la structure collocationnelle retenue [V + Nsc] l'a été non seulement sur un critère de productivité, mais aussi parce qu'elle nous semblait particulièrement intéressante du point de vue sémantique : en effet, elle présente des collocations plus typées sémantiquement que la structure [Nsc + V]. Cependant, il va de soi qu'il serait intéressant de compléter l'étude par l'examen d'autres types de collocations (par exemple [Nsc + Adj.]).

2. 3. La construction des dimensions sémantiques

L'élaboration des dimensions sémantiques s'est avérée complexe au vu des nombreux sens et utilisations en contextes des lexies. Comme nous travaillions dans le domaine de la phraséologie, il n'était pas possible en outre, de nous attacher à des distinctions sémantiques trop fines : il nous fallait en effet adopter un système de catégorisation suffisamment large pour neutraliser les problèmes de figement dont la distinction sémantique n'est pas réellement significative dans l'usage ; par exemple pour 'émettre une hypothèse'/'faire une hypothèse', il est difficilement envisageable de classer les deux verbes dans deux dimensions différentes étant donné qu'en contexte la différence sémantique reste minime voire inexistante.

³ Constitué par F. Rinck, Lidilem, dans le cadre de son travail de thèse.

C'est d'abord uniquement à partir du corpus que des grandes classes sémantiques sont apparues pertinentes pour une taxinomie des verbes associés aux noms. Grâce à l'extraction des traits sémantiques de la définition, nous avons dégagé quatre dimensions qui apparaissent privilégiées dans le corpus. Ces dimensions contiennent les verbes associés le plus fréquemment aux noms dans le corpus. Certains verbes co-occurents, intuitivement pourtant associables aux noms, n'apparaissent que très rarement (une seule occurrence pour « énoncer une hypothèse »). Le regroupement des verbes dans des classes plus larges pallie en partie cet inconvénient :

- La dimension EMISSION rassemble les verbes qui permettent l'énonciation, la présentation de l'objet. Les verbes emblématiques seraient *émettre* et *énoncer*.
- La dimension ELABORATION contient les verbes qui décrivent la construction de l'objet, sa création. Les verbes rencontrés sont *faire*, *construire*, *effectuer*.
- La dimension VERIFICATION associe les verbes qui viennent attester de la validité de l'objet tels que *vérifier*, *valider* ou *infirmer*.
- La dimension ARGUMENTATION regroupe des verbes qui permettent de placer l'objet en position d'enjeu d'un débat. Les verbes *défendre*, *récuser* ou encore *appuyer* s'y côtoient.

Afin de ne pas surcharger cette présentation avec trop de définitions, nous avons choisi de ne présenter que les sens pertinents à l'étude. Les quatre noms choisis ont deux types de sens décrits par le TLF, un sens dit 'courant' et un sens 'scientifique'. Le plus souvent, le sens courant ne nous concerne pas, même s'il peut être utile d'en tenir compte pour l'extraction des traits sémantique ; en revanche le sens scientifique nous préoccupe davantage. Certains noms vont appartenir à une science spécifique qui n'entrera que rarement dans le discours d'autres sciences (c'est le cas pour *postulat*). En revanche les trois autres noms ont davantage de champs d'utilisation et vont apparaître dans les 'sciences' de façon générale (même si certains affichent un domaine privilégié : la philosophie pour *hypothèse*).

Nous avons remarqué que le terme *science* est utilisé de façon polysémique dans le TLF. Il apparaît autant pour désigner toutes les sciences (sans distinction de 'sciences exactes'/'sciences humaines') que pour n'en désigner que certaines non citées explicitement dans la définition. L'exemple de *hypothèse* est intéressant puisque le domaine 'philosophie' est associé à celui de 'sciences' ce qui interroge la catégorisation des domaines effectuée par le TLF.

3. Analyse des données recueillies

3.1. Etude du nom 'hypothèse'

Dans le TLF, le sens spécialisé de *hypothèse* est en philosophie et en sciences. Il se dégroupé en deux acceptions spécifiques :

- a. Proposition reçue, indépendamment de sa valeur de vérité, et à partir de laquelle on déduit un ensemble donné de propositions. Synonyme : *Principe*.
- b. Proposition (ou ensemble de propositions) avancée, provisoirement, comme explication de faits, de phénomènes naturels et qui doit être, ultérieurement, contrôlée par la déduction ou par l'expérience. Synonyme : *Conjecture*.

Il existe un sens courant dérivé du sens spécialisé en sciences : 'Supposition, conjecture'.

A partir de ces sens dans les domaines scientifiques, nous avons extrait les traits sémantiques qui nous ont permis à la fois d'établir et de valider les dimensions sémantiques. Voici la liste des traits sémantiques dégagés pour *hypothèse* :

/principe/, /sert de fondement/, /provisoire/,
/à contrôler/, /supposition/, /vraie ou fausse/,
/vers la réalisation d'un fait/, /conséquences/.

Le trait /à contrôler/ associé à l'exemple issu du TLF 'rejeter une hypothèse' permet d'insérer ce nom dans la dimension ARGUMENTATION dès l'instant où le type de verbe associé à cette dimension apparaît en co-occurrence avec *hypothèse*. En comparant ces relations sémantiques, nous constatons que les traits sémantiques de la définition s'associent en contexte avec les quatre dimensions décrites précédemment :

La dimension EMISSION : Avancer (17⁴), émettre (16), énoncer (1), formuler (9), introduire (2), poser (9), présenter (4), proposer (6)

La dimension ELABORATION : Effectuer (4), faire (113), former (1)

La dimension VERIFICATION : Confirmer (12), corroborer (3), examiner (3), infirmer (4), mettre à l'épreuve (3), tester (35), valider (6), vérifier (6)

La dimension ARGUMENTATION : Appuyer (1), conforter (2), défendre (9), discuter (3), étayer (3), justifier (4), légitimer (1), opposer (1), privilégier (2), récuser (1), renforcer (4), soutenir (2)

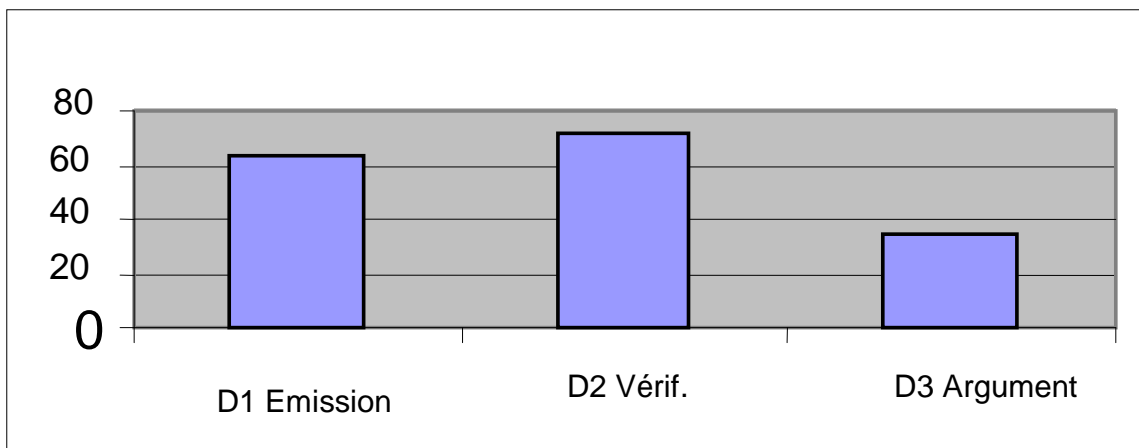
⁴ Les chiffres entre parenthèse représentent le nombre d'occurrences répertoriées dans tout le corpus.

La fréquence du verbe *faire* dans le corpus s'explique aisément. Les occurrences en contexte ont révélé la fusion de deux sens distincts dans la plupart des cas. Il s'avère que le verbe *faire* est associé à deux fonctions distinctes :

- a. *Faire* comme verbe support (cf. $oper_1$ dans le DEC ; Mel'Cuk, 1984) : *faire* est un outil grammatical que nous pourrions remplacer par le verbe *Hypothéser* si celui-ci existait en français (il existe en anglais *to hypothesize*).
- b. *Faire* comme verbe d'élaboration : le verbe signifie alors *construire et/ou fabriquer*.

Les exemples trouvés dans le corpus révèlent cette ambiguïté sémantique qu'il est quasiment impossible de lever : « Le premier programme scientifique en linguistique qui a fait l'hypothèse du langage naturel comme système formel. »

Il est donc difficile de prendre en compte ce verbe pour notre étude. Le nombre élevé d'occurrences fausse les résultats d'analyse qui deviennent caduques dès l'instant où nous ne pouvons trancher quant à la dimension dans laquelle classer les occurrences ; doit-on privilégier le sens de la dimension EMISSION ou celui de la dimension ELABORATION ? Serait-ce dès lors pertinent d'élaborer des classes sémantiques plus fines pour un tel cas ? Seule une analyse micro-sémantique en contexte nous permettrait, peut-être, de distinguer les deux sens apparemment très liés. Cependant, nous n'avons pas pour objectif de décrire tous les sens spécifiques de chaque lexie dans tous les contextes.



Graphique : Principales dimensions sémantiques pour hypothèse

3.2. Etude du nom 'postulat'

Le sens spécialisé dans le TLF est prioritairement celui de la géométrie puis celui des sciences. Ce sens se dégroupé en trois acceptions spécifiques :

- a. Proposition que l'on demande d'admettre comme principe d'une démonstration, bien qu'elle ne soit ni évidente ni démontrée.
- b. Proposition faisant partie de l'axiomatique formulée au départ d'un système hypothético-déductif. Synonyme : *Axiome*.
- c. Principe non démontré que l'on accepte et que l'on formule à la base d'une recherche ou d'une théorie.

Notons que le sens courant est dérivé du sens spécialisé : 'Représentation qui est admise de façon implicite et sur laquelle se fonde un système de pensée' ; il est utilisé en sciences humaines.

Les traits sémantiques de postulat extraits de la définition :

/proposition admise/,	/évidente/,	/non démontrée/,
/vrai par définition/,	/principe/,	/fondement/.

Le nom *postulat* est apparu comme très peu productif dans notre corpus. Seules treize occurrences apparaissent pour la structure étudiée [V + postulat]. En revanche, notons la concurrence non négligeable du verbe *postuler* qui est en dehors du champ de notre étude. *Postulat* apparaît d'une utilisation rare en sciences humaines (notre corpus est majoritairement constitué d'articles scientifiques en sciences humaines). De ce fait, les dimensions ne contiennent que peu de verbes mais ils sont surprenants. En effet, nous constatons l'émergence de la dimension VERIFICATION qui, selon la définition du TLF, ne devrait pas apparaître puisqu'un postulat est un fondement qui ne se discute pas. Nous retiendrons les verbes suivants pour cette dimension : Affiner un postulat (1), confirmer un postulat (1), décliner un postulat (1), infirmer un postulat (1), redoubler un postulat (1), relativiser un postulat (1), ruiner un postulat (1), valider un postulat (1)

L'apparition de cette dimension s'explique aisément dès lors que l'on se situe en discours : en effet, le fondement accepté un temps par une recherche peut être remis en cause lors d'une phase ultérieure, ou bien une étude peut contester les postulats adoptés par d'autres.

Nous trouvons également des verbes typiques de la dimension EMISSION : Ajouter un postulat (1), appliquer un postulat (1), avancer un postulat (1), implémenter un postulat (1), réactiver un postulat (1)

En sciences humaines le nom *postulat* semble souvent être utilisé avec un sens proche de *hypothèse*. Les deux lexies contiennent les traits sémantiques /principe/ et /fondement/ ce qui peut contribuer à leur rapprochement sémantique en contexte. Nous pourrions cependant nous

demander pourquoi un tel emploi alors que *hypothèse* suffirait. *Postulat* ferait-il ‘plus’ scientifique ? Une investigation dans un corpus complémentaire (la partie ‘traités’ de la base Frantext de 1900 à 2000) semble cependant montrer qu’il peut exister aussi une dimension argumentative (« Admettre le postulat » : 20 occurrences ; « rejeter le postulat » : 14 occ. ; « nier » : 1 occ. ; « confirmer » : 2 occ. ; « infirmer » : 2 occ. ; « démontrer » : 6 occ.). En Sciences Humaines tout particulièrement, *postulat* revêt alors souvent un sens négatif, les postulats étant conçus comme des implicites à démasquer : « Mais ce n’est là que <*postulat*> et oubli des vérités » (Foucault).

3.3. Etude du nom ‘thèse’

La définition de *thèse* révèle deux sens dont celui d’ouvrage ‘pour obtenir le grade de docteur’ qui ne convient pas à nos objectifs sémantiques. De ce fait, le seul sens qui soit pertinent à notre étude est non spécialisé mais utilisé en science : « Proposition ou théorie que l’on tient pour vraie et que l’on soutient par une argumentation pour la défendre contre d’éventuelles objections. »

Les traits sémantiques de *thèse* retenus pour notre corpus :

/proposition/,

/à défendre/,

/théorie/,

/vraie pour ses défenseurs/.

Deux dimensions principales émergent malgré les réserves d’usage concernant les faibles fréquences :

EMISSION : Avancer (1), énoncer (1), exposer (1), formuler (1)

ARGUMENTATION : Appuyer (3), contester (1), défendre (6), exclure (1), privilégier (1), réfuter (1), rejeter (1), soutenir (1)

La dimension ARGUMENTATION arrive en tête pour *thèse* (par rapport à EMISSION, ce qui présente le cas inverse de ce qui se passe pour *hypothèse*). Cette prégnance argumentative s’actualise en contexte comme le prévoyait la définition : une thèse doit être défendue.

3.4. Etude du nom ‘théorie’

Le TLF donne un sens spécialisé en sciences, énoncé comme suit : « Construction intellectuelle, hypothétique et synthétique, organisée en système et vérifiée par un protocole expérimental. Ensemble de lois formant un système cohérent et servant de base à une science, ou rendant compte de certains faits. »

Ses traits sémantiques sont les suivants :

/ensemble de notions/,	/intellectuel/,
/hypothétique/,	/organisation/,
/fondement abstrait/,	/cohérence/.

Contrairement à *hypothèse* et *postulat* qui présentent des sens différents selon le domaine scientifique dans lequel ils sont utilisés, *théorie* ne présente pas de polysémie caractérisée dans le domaine scientifique. Cependant, la polysémie est bien présente dès l'instant où nous utilisons la première ou la deuxième phrase de la définition. Dans la première phrase il est question 'de construction hypothétique' tandis que dans la deuxième nous avons à faire à 'un système servant de base' ce qui rappelle le trait sémantique /fondement/ déjà rencontré. De ce fait, l'utilisation du nom *théorie* avec des verbes de la dimension ARGUMENTATION conduit à envisager la *théorie* comme un objet discutable donc à caractère hypothétique et non immuable (qui est dans le trait /fondement/). Dans ce cas c'est le sens de la première phrase de la définition qui est pris en compte et non la deuxième.

La polysémie n'est donc pas marquée entre les disciplines scientifiques mais, au sein de chacune d'entre elles, le terme peut être utilisé avec l'un ou l'autre des deux sens de sa définition, voire parfois avec les deux facettes.

Dans le corpus la construction [V + théorie] est peu fréquente, mais toutes les dimensions sémantiques sont représentées par quelques verbes :

ELABORATION : Construire (3), déboucher sur (1), édifier (2), élaborer (2), esquisser (3), établir (1), fonder (2), jeter les bases (2), produire (1)

EMISSION : Décrire (1), exposer (3), introduire (1), présenter (2), proposer (5)

VERIFICATION : Confirmer (1), évaluer (2), falsifier (1), tester (1), vérifier (1)

ARGUMENTATION : Accepter (1), admettre (1), adopter (1), défendre (1), démonter (1), justifier (1), mettre en cause (1), reconsidérer (1), rejeter (1), tourner en dérision (1)

4. Regroupement des traits selon les dimensions sémantiques

Il nous a paru possible de regrouper certains des traits obtenus à partir de l'analyse lexicologique des lexèmes dans les dimensions préalablement construites. Nous fournissons (tableaux 1 et 2) deux exemples de ces regroupements. Il convient de noter que chaque lexème peut être analysé

sur plusieurs dimensions, les différences étant généralement seulement tendancielle : un lexème sera plus fortement marqué sur telle ou telle dimension.

Noms	Dimension	Traits sémantiques	Exemples du TLF	Exemples du corpus
Hypothèse	ARGUM	/à contrôler/	Rejeter une hypothèse	Appuyer une hypothèse, défendre une hypothèse
Théorie		/hypothétique/ (sens 1)	Admettre une théorie	Défendre une théorie, accepter une théorie
Thèse		/à défendre/	Défendre une thèse, rejeter la thèse de	Défendre une thèse, appuyer une thèse

Tableau 1 : Exemples de regroupement : la dimension Argumentation

On remarquera que la dimension argumentative, sémantiquement marquée pour *thèse*, est plus implicite pour *théorie* et pour *hypothèse*, et ne se manifeste qu'à travers des chaînes inférentielles du type : si on a contrôlé une hypothèse, on peut argumenter en sa faveur, la défendre, etc. Cette dimension argumentative fournit cependant des collocations identiques (par exemple « défendre une hypothèse, une théorie, une thèse »). Une part d'idiosyncrasie demeure toutefois : 'appuyer une théorie', bien que non totalement impossible, n'a pas été trouvé, ni dans les exemples du TLF, ni dans le corpus. La dimension VERIFICATION fournit un autre exemple de complexité, avec le statut différent accordé à la vérité pour les lexèmes considérés : ainsi, si *hypothèse* peut avoir deux valeurs de vérité (V ou F), *postulat* a par défaut la valeur V, tandis que *thèse* a une valeur relative V, suivant qui la défend (une thèse est vraie pour celui qui la défend, mais peut être fautive pour celui qui la discute. Le tableau 2 résume ces différentes possibilités.

Noms	Dimension	Traits sémantiques	Exemples du TLF	Exemples du corpus
Hypothèse	VERIF	/vraie ou fautive/	Admettre une hypothèse	Confirmer une hypothèse, vérifier une hypothèse
Postulat		/vrai par définition/	Admettre le postulat	Confirmer un postulat, infirmer un postulat
Thèse		/vraie pour ses défenseurs/	Accréditer, confirmer,	Pas d'occurrence

			invalider une thèse	
--	--	--	---------------------	--

Tableau 2 : Exemples de regroupement : la dimension VERIFICATION

5. Conclusion

Les quatre noms étudiés comportent de nombreux co-occurents communs (liés aux quatre dimensions). On note qu'on a plutôt affaire à des affinités qu'à des exclusions : « émettre ou énoncer une hypothèse » plutôt que « exposer une hypothèse », « exposer une théorie » plutôt que « énoncer une théorie ».

Notre approche multidimensionnelle permet au moins partiellement d'expliquer ce phénomène. En effet, les configurations des lexèmes sur chacune des dimensions rendent compte tendanciellement des collocations privilégiées avec certains noms. Leur repérage sur les dimensions sémantiques met bien en évidence le fait qu'on a des fonctionnements réguliers correspondant à tous les lexèmes étudiés (ex. 'défendre une thèse, une hypothèse, un postulat, une théorie'), tout en expliquant pourquoi « défendre une thèse » est plus prototypique. Cependant, certaines dimensions plus que d'autres favorisent l'idiosyncrasie : on a ainsi des sélections différentes des co-occurents verbaux en fonction des noms pour la dimension EMISSION, sans qu'il soit toujours possible de trouver des raisons sémantiques évidentes. On ne pourra donc pas échapper complètement à un traitement au cas par cas, même si notre approche permet au moins de suggérer la probabilité des associations lexicales en fonction du typage sémantique des noms.

En ce qui concerne les différences d'usage, les lexèmes étudiés sont, pour certains, utilisés indifféremment quel que soit le domaine scientifique : ex. *thèse*, et *théorie*. D'autres sont d'emploi plus spécifique : ex. *postulat*, très peu utilisé en sciences humaines, et dont le sens se rapproche alors de celui d'*hypothèse*. Une investigation complémentaire serait ici nécessaire, mais elle supposerait la mise au point d'un très gros corpus, équilibrant le poids de chaque discipline, ce qui se révèle pour l'instant difficile. A partir de notre investigation, il nous est apparu qu'il y avait effectivement des différences importantes selon cette variable disciplinaire. On peut prendre à titre d'exemple le statut respectif d'*hypothèse* en médecine et en linguistique : nos résultats indiquent une différence sur la dimension VERIFICATION, ce qui s'explique par des différences épistémologiques (cf. la place de la démarche expérimentale en médecine). L'approche linguistique aurait donc tout à gagner à être ainsi complétée par une meilleure prise en compte des variables sociolinguistiques : les différences de champs disciplinaires, mais aussi

les facteurs liés au degré d'expertise des scripteurs, ou encore les représentations de la science dans les différents champs disciplinaires.

Bibliographie

GROSSMANN Francis et TUTIN Agnès, 2003, "Quelques pistes pour le traitement des collocations", in *Les collocations : analyse et traitement*, Amsterdam: De Werelt.

IMBS Paul et QUÉMADA Bernard, 1977, *Trésor de la Langue Française - Dictionnaire de la langue du XIX^e et XX^e siècle (1789-1960)*, Paris: Centre National pour la Recherche Scientifique.

MEL'ČUK Igor, 1984, *Dictionnaire explicatif et combinatoire du français contemporain : recherches lexico-sémantiques*, Montréal: Presses de l'université.

PHAL André et BEIS Lucette, 1971, *Vocabulaire général d'orientation scientifique : part du lexique commun dans l'expression scientifique*, Paris: Centre de recherche et d'études pour la diffusion du français, CREDIF.

RASTIER François, 1996c, *Sémantique interprétative*, Paris: Presses Universitaires de France.

TUTIN Agnès et GROSSMANN Francis, 2002, "Collocations régulières et irrégulières : esquisse du phénomène collocatif", *Revue française de linguistique appliquée* n°VII-1, 7-25.