

Adjectifs dérivés sémantiques (ADS) dans la structuration des terminologies

Marie-Claude L'Homme

Observatoire de linguistique Sens-Texte (OLST)
Département de linguistique et de traduction
Université de Montréal
C.P. 6128, succ. Centre-ville, Montréal (Québec), H3C 3J7, Canada
lhommem@ling.umontreal.ca

Résumé

Le présent article présente les résultats préliminaires d'une étude plus globale visant à mieux comprendre le rôle des adjectifs dans la structuration des terminologies. Il décrit les résultats d'une analyse portant sur les adjectifs dérivés sémantiques (ADS), à savoir les adjectifs dénominaux et déverbaux, observés dans trois corpus spécialisés. Les corpus utilisés appartiennent au domaine de l'informatique, du droit et de la médecine. L'étude essentiellement descriptive vise à estimer l'importance des ADS dans des corpus spécialisés de nature différente et à décrire leur contribution à la structuration des terminologies. Cette contribution est étudiée en fonction du lien sémantique partagé par l'ADS avec le nom ou le verbe formant sa base.

Mots clés : adjectif, déverbal, dénominal, dérivé sémantique, relation lexico-sémantique, structure terminologique, corpus spécialisé

1. Introduction¹

Les conséquences des analyses terminologiques fondées sur corpus sur les modèles fondateurs de la discipline ont été examinées sous une variété d'angles (Bourigault et Slodzian 1999; Meyer et Macintosh 1996, entre autres) et ne sont probablement pas encore toutes mises au jour.

En ce qui concerne la structuration des terminologies, divers travaux montrent l'importance de tenir compte de relations de nature lexico-sémantique et de ne pas focaliser exclusivement sur les relations conceptuelles (voir un numéro spécial de TAL (2002) sur la structuration des terminologies et un numéro spécial de *Terminology*, à paraître en 2004). Les relations conceptuelles, qui résultent d'une organisation consensuelle des connaissances, ne révèlent qu'une partie de la structure terminologique d'un domaine.

Parmi les relations lexico-sémantiques étudiées récemment, citons la variation terminologique (Daille 2001; Jacquemin 2001), la dérivation

morphologique (Zweigenbaum et Grabar 2001), les liens syntagmatiques, notamment les collocations (Heid 1994).

En outre, la contribution de catégories grammaticales autres que nominales à la structuration des terminologies est généralement admise et mieux décrite. Le rôle du verbe, notamment comme prédicat sémantique, a été étudié dans Sager (1987). Une application au domaine de l'informatique a été réalisée par L'Homme (1998; 2003). De même, des classes d'adjectifs commencent à retenir l'attention (Assadi et Bourigault 1995; Daille 2001; Maniez 2001; Normand et Bourigault 2001) (les études les plus proches de la nôtre sont présentées à la section 2).

Le présent travail s'inscrit dans cette problématique. Nous nous penchons plus précisément sur les adjectifs dénominaux (ex. *algorithmique*, *fibreuse*) et déverbaux (ex. *compilable*, *injustifié*) que nous étiquetons *adjectifs dérivés sémantiques* (ADS). Nous tentons de décrire leur contribution à la structuration des terminologies par l'analyse des liens qu'ils partagent avec les bases nominales et verbales.

¹ Ce travail a été réalisé avec l'aide de Mériam Haddara, Chantal Lemay et Elizabeth Marshman que je tiens à remercier. Je remercie également Patrick Drouin qui a mis au point le concordancier utilisé pour mener à bien cette étude ainsi que Sylvie Vandaele et les relecteurs anonymes pour leurs suggestions et commentaires sur une version préliminaire du présent article.

2. Travaux antérieurs

L'adjectif ayant retenu la plus grande attention dans les travaux récents de terminologie est l'adjectif dit « de relation ». Contrairement aux autres formes d'adjectifs, l'adjectif de relation exprime un actant du nom qu'il modifie et est souvent distingué des autres en raison d'un comportement syntaxique différent (par exemple, il ne peut être modifié par un adverbe et admet difficilement la position attributive). Cette distinction, fondée sur le comportement syntaxique, est cependant remise en cause (Mélis-Puchulu 1991).

Maniez (2002) propose une stratégie de distinction automatique des termes complexes composés d'un nom et d'un adjectif des collocations (groupes semi-figés) en s'appuyant sur les propriétés de l'adjectif de relation (notamment, le fait qu'il ne puisse être modifié par un adverbe). D'après le chercheur, les groupes comprenant un adjectif relationnel sont plus susceptibles de former un terme complexe, alors que ceux qui comprennent un adjectif qualificatif se comportent comme des collocations.

Daille (2001) s'intéresse également aux adjectifs relationnels, mais dans une optique différente. Elle identifie automatiquement les groupes formés d'un nom et d'un adjectif relationnel en tenant pour acquis que le terme contenant la forme nominale correspondante sera sémantiquement équivalente (ex. *épanchement sanguin* = *épanchement du sang*). Cette stratégie, qui s'appuie sur la parenté morphologique de l'adjectif de relation et du nom ainsi que sur la présence du nom dans le corpus, permet de réunir des variantes terminologiques.

L'accent placé sur l'adjectif de relation ne permet pas de se faire une idée générale du rôle joué par les autres formes d'adjectifs. Quelques cas de figure sont examinés ci-dessous :

- a) Un nom peut partager toute une série de relations sémantiques avec différentes formes d'adjectifs : par exemple, *allergie* partage des liens sémantiques avec *allergisant*, *allergène*, *allergénique* (ex. « qui peut provoquer une allergie »); *allergique* (ex. « présentant une allergie »), *anallergique* (ex. qui ne provoque pas d'allergie), etc. *Loi* est lié à *légal*, *législatif*, *légitime*, *licite*.
- b) De même, un verbe peut être lié à différents adjectifs : par exemple, *compilable* (x *compile* y; y est *compilable*), *compilé* (x *compile* y; y est *compilé*); *évolutif* (x *évoque*; y est *évolutif*).

- c) Le même adjectif peut avoir des sens différents. Soit, le sens n'est pas le même (*populaire* : *éducation populaire*, *mesure populaire*), soit la relation qui lie l'adjectif à la base n'est pas la même (*symptomatique* ; *traitement symptomatique* (qui s'attaque aux symptômes), *amélioration symptomatique* (des symptômes)).
- d) Un syntagme nominal de type nom + adjectif peut apparaître beaucoup plus fréquemment dans un corpus spécialisé que son correspondant nom + prép. + nom. Par exemple, *traitement des symptômes* n'apparaît qu'une seule fois dans un corpus médical d'environ 500 000 mots, alors que *traitement symptomatique* y figure 30 fois.
- e) Les adjectifs dénominaux n'ont pas toujours une correspondance nominale dans les corpus. Par exemple, *traitement symptomatique* et *amélioration symptomatique* ainsi que *traitement* et *amélioration des symptômes* peuvent être observés dans un corpus médical. Toutefois, *épilepsie symptomatique* (qui se manifeste par des symptômes), *patient symptomatique*, *enfant symptomatique* (qui présente des symptômes) et *médicament symptomatique* (qui traite les symptômes), même si le lien avec *symptôme* est perceptible, ne donnent pas lieu à des paraphrases nominales dans le même corpus médical. Une seule occurrence de *épilepsie, accompagnée par des symptômes*, a été relevée.

En ne prenant pas en compte les autres adjectifs, et *a fortiori* les adjectifs dénominaux et déverbaux, les analyses terminologiques risquent de faire l'impasse sur certaines formes de variantes terminologiques et sur de nombreux contextes où des sens spécialisés sont exprimés.

3. Objectifs et méthode

Dans ce travail, nous nous pencherons sur un aspect de la contribution des adjectifs à la structuration des terminologies telle qu'elle peut s'observer dans les corpus spécialisés.

Les adjectifs qui nous intéressent sont ceux que nous étiquetons *dérivés sémantiques* (ADS). Ils présentent les propriétés suivantes :

- ils sont construits par conversion ou par affixation sur une base nominale (ex. *constitutionnel*, à partir de *constitution*, *informatique* adjectif, à partir de *informatique* nom) ou une base verbale (ex. *programmable*, à

partir de *programmer*) (ces adjectifs sont appelés respectivement *adjectifs dénominaux* et *adjectifs déverbaux*) ;

- ils ont un sens prédictible à partir de celui du nom ou du verbe de base ; les sens est dit *prédictible* lorsqu'on peut le paraphraser à partir du nom ou du verbe de base (par exemple, *imprimante configurable* : *l'actant 2 (imprimante) peut être configuré(e)*).

Nous tenons pour acquis que la contribution de ces adjectifs à la structuration des terminologies se définit en fonction des liens sémantiques qu'ils partagent avec leur base et ils pourront faire l'objet d'une représentation qui rend compte de cette relation sémantique.

Nous mettons donc de côté les adjectifs suivants.

- les adjectifs dits « qualificatifs » (ex. *grand, large, élevé*) ;
- les adjectifs dits « numéraux » ou « indéfinis » étiquetés comme les autres par l'étiqueteur (ex. *quadruple, certain, premier*) ;
- les adjectifs régissant des syntagmes prépositionnels (ex. *possible, facile*).

Nous avons classé dans un groupe à part certains adjectifs qui ne sont ni des adjectifs qualificatifs proprement dits, ni des dérivés dont on peut prédire le sens à partir du sens d'une base. Il s'agit d'adjectifs comme *clinique* ou *électrique*, dont la contribution à la structuration lexicale d'un domaine ne fait pas de doute, mais dont le sens ne peut plus se décrire à partir de celui d'une base verbale ou nominale. Souvent, ces adjectifs ont une parenté morphologique avec un nom, mais le lien sémantique est perdu (ex. *clinique* dans *d'autres sujets présentent des signes cliniques*).

Nous nous intéresserons à la place qu'occupent les ADS dans trois corpus spécialisés différents, à savoir un corpus d'informatique (CI), de droit (CD) et de médecine (CM)² ainsi qu'aux relations sémantiques les plus fréquemment observées dans les trois corpus. Le présent article présente les résultats préliminaires d'une étude plus globale visant à mieux comprendre le rôle des adjectifs dans la structuration des terminologies.

² Les corpus sont ceux du groupe de terminologie de l'Observatoire de linguistique Sens-Texte (OLST) (<http://www.olst.umontreal.ca>) de l'Université de Montréal. Chacun des corpus comprend approximativement 500 000 mots.

Nous nous sommes penchées (dans chacun des trois corpus spécialisés) sur 1000 concordances contenant un adjectif. Les textes avaient été préalablement traités au moyen de l'étiqueteur morpho-syntaxique (*WinBrill*, Lecompte 1998). Les contextes contenant un mot étiqueté ADJ³ ont été extraits au moyen d'un programme maison et placés dans une base de données. Des exemples de contextes extraits sont proposés en (1).

- (1) *Aux États-Unis, l'établissement des sites d'interconnexion concrétise cette organisation hiérarchique/ADJ de la prestation d'accès : des fournisseurs d'accès sont ... retard de maturation osseuse. Les anomalies squelettiques/ADJ sont le plus souvent latentes. Les signes radiologiques ... elle dispense d'obtenir sous serment un mandat d'une autorité décidant d'une façon judiciaire/ADJ. D'une manière générale ...*

Les 1000 concordances ont été prélevées à partir des contextes générés par le concordancier. Elles ont été corrigées afin d'éliminer les erreurs d'étiquetage.

4. Quelques résultats

4.1 Généralités

Tout d'abord, l'étude des adjectifs dans les trois échantillons révèle qu'ils ne distribuent pas de la même manière d'un corpus à l'autre⁴. Les adjectifs dérivés sémantiques sont nettement plus fréquents dans le CD et le CM (représentant respectivement 37,7 % et 40,8 % des adjectifs étudiés et près de la moitié des adjectifs en combinaisons libres) que dans le CI (20,8% des adjectifs extraits). Voir l'Annexe 1 pour le détail de la distribution.

Les trois corpus se distinguent également lorsqu'on examine les proportions d'adjectifs dénominaux et déverbaux. Parmi les ADS relevés dans le CM, 85,3 % sont des adjectifs dénominaux. La proportion de dénominaux dans le CD est légèrement inférieure (78,2 %). Le CI, quant à lui, accorde une plus grande place aux adjectifs déverbaux, puisqu'ils représentent 39,4 % des ADS analysés. L'Annexe 2 présente de détail de la répartition dans les trois corpus.

³ L'étiqueteur distingue les adjectifs des participes passés. Nous n'avons pas extrait les concordances contenant un participe passé en raison du bruit qu'elles engendrent. Ce choix, toutefois, risque sans doute de faire perdre quelques participes lexicalisés.

⁴ Il aurait été intéressant de connaître la proportion exacte d'adjectifs sur le total des mots dans chacun des corpus, mais les erreurs d'étiquetage n'ont été corrigées que sur l'échantillon et fausseraient les résultats.

4.2 Relations sémantiques entre ADS et bases nominales et verbales

Les relations sémantiques entre les adjectifs dérivés sémantiques avec une base nominale ou verbale sont valables pour la majorité des ADS analysés. De plus, la plupart d'entre elles peuvent être observées dans les trois corpus.

Le tableau 1 présente les relations sémantiques régulières entre adjectifs et bases verbales; le tableau 2, les relations observées entre l'adjectif et la base nominale. Des exemples ainsi que la paraphrase utilisée pour valider chaque relation sémantique sont également reproduits.

Certains adjectifs sont polysémiques et les interprétations reproduites ci-dessous sont valables pour les contextes cités. Par ailleurs, le même adjectif peut parfois se paraphraser à partir d'une base verbale ou nominale. Ce cas, rarement observé dans les corpus étudiés, touche pourtant l'adjectif *législatif* : 1. *texte législatif* : texte de loi; 2. *pouvoir législatif* : pouvoir de légiférer.

Tableau 1 : Relations sémantiques régulières verbe – adjectif

Verbe – adjectif	
1. l'actant 1 verbe	
CI	<i>évoluer – évolutif (ex. spécification évolutive)</i>
CM	<i>activer – activateur (ex. pouvoir activateur) prévenir – préventif (ex. traitement préventif)</i>
CD	<i>déroger ; dérogatoire (ex. clause dérogatoire) discriminer ; discriminatoire (ex. acte discriminatoire)</i>
2. l'actant 2 peut être verbe	
CI	<i>paramétrer – paramétrable (ex. carte paramétrable)</i>
CM	<i>injecter – injectable (ex. préparation injectable)</i>
CD	<i>dissocier – indissociable (ex. disposition indissociable)</i>
3. l'actant 2 est verbe	
CI	<i>formater – formaté (ex. document formaté)</i>
CD	<i>justifier – injustifié (ex. intrusion injustifiée)</i>

Tableau 2 : Relations sémantiques régulières nom – adjectif

Nom – adjectif	
1. l'actant 1 possède, a, avec + nom	
CI	<i>défaut – défectueux (ex. secteur défectueux)</i>
CM	<i>symptôme – symptomatique (ex. patient symptomatique), cellulaire (ex. zone « richement » cellulaire)</i>
CD	<i>énergie – énergétique (mesure énergétique)</i>
Les adjectifs pouvant être paraphrasés au moyen d'une préposition + nom de base (de, à, par, en), c'est-à-dire expriment un actant du nom modifié (agent, patient, instrument, etc.)	
CI	<i>bureautique – bureautique (ex. tâche bureautique), public – public (ex. service public)</i>
CM	<i>plasmatique – plasma (ex. concentration plasmatique); osseuse – os (ex. atteinte osseuse, fragment osseux)</i>
CD	<i>état – étatique (intervention étatique), gouvernemental – gouvernement (ex. intervention gouvernementale)</i>

Certaines relations sémantiques sont exprimées par un petit groupe d'adjectifs ou sont valables pour un seul corpus. Quantitativement, elles sont moins importantes que les relations sémantiques régulières. Toutefois, elles prennent une variété de formes.

La plupart de ces relations sémantiques spécifiques sont exprimées par les adjectifs préfixés (ex. *tridimensionnel, unilingue, vasodilatateur, sous-périosté*).

Toutefois, certaines relations sémantiques spécifiques sont également exprimées par des adjectifs formés par suffixation. Ci-dessous, quelques exemples de relations sémantiques apparaissant dans un seul corpus :

- Qui cause, qui engendre + nom : *infectieux* (ex. *agent infectieux*); *pathogène* (ex. *agent pathogène*) (CM);
- intervient sur + nom : *symptomatique* (ex. *médicament, symptomatique*); *cardiaque* (ex. *chirurgie cardiaque*) (CM);
- qui ressemble à + nom : *lipomateux* (ex. *masse lipomateuse*).
- conforme à + nom : *légal*; *conforme à la loi* (CD).

5 Conclusion

Les données examinées dans cette étude permettent de se faire une idée de la place occupée par les adjectifs dérivés sémantiques dans les corpus spécialisés. Ces derniers se distinguent sur un certain nombre de plans. La proportion d'ADS n'est pas la même d'un corpus à l'autre. Il est donc permis de croire que l'adjectif, et plus précisément l'adjectif dérivé sémantique, joue un rôle important dans la structuration de certaines terminologies. En outre, les adjectifs dénominatifs et déverbaux ne se distribuent pas de la même manière en informatique, en médecine et en droit.

La plupart des relations sémantiques observées entre les ADS et leur base apparaissant dans les trois corpus retenus pour cette étude. On peut donc penser que le relevé de ces classes d'adjectifs permettrait de couvrir la majorité des ADS dans les corpus spécialisés. Certaines relations, toutefois, sont spécifiques à un corpus. Même si quantitativement, elles sont moins importantes que les relations régulières, elles prennent une variété de formes et doivent être décrites isolément.

Cette étude permet de poser les bases d'une modélisation des liens entre ADS et bases verbales

et nominales. Nous comptons développer ce travail dans trois directions :

1. Extraire un plus grand nombre d'adjectifs des corpus informatique, médical et juridique et dresser la liste des relations sémantiques «non régulières» observées dans ces corpus ;
2. Étudier les répartitions respectives des bases et des adjectifs dans les corpus (voir si les bases sont toujours attestées et, si elles le sont, dans quelle proportion);
3. Voir s'il est possible d'associer des adjectifs associés à une relation sémantique donnée à des classes particulières de noms (le nom qu'ils modifient).

Cette dernière étude pourrait servir à résoudre un problème qui été qu'évoqué rapidement plus haut, à savoir celui des adjectifs polysémiques.

Par exemple, l'adjectif *symptomatique*, cité plus haut, modifie des nom appartenant à des classes différentes : 1. *patient* (<animé humain>) *symptomatique* : patient qui présente des symptômes; 2. *médicament symptomatique* (<substance>): médicament qui s'attaque aux symptômes; 3. *amélioration* (<activité>) *symptomatique* : amélioration des symptômes; 4. *épilepsie* (<maladie>) *symptomatique* : épilepsie qui se manifeste par des symptômes. Toutefois, il faudra voir si cette analyse reste valable pour un grand nombre d'ADS.

D'autres problèmes de polysémie sont entraînés par l'ellipse d'un élément dans le syntagme nominal. Par exemple, *cardiaque* : 1. *muscle cardiaque* ; muscle formant le cœur; 2. *fréquence cardiaque* : fréquence des contractions du cœur; 3. *débit cardiaque* : débit du sang éjecté par le cœur; 4. *volume cardiaque* : volume de sang dans le cœur.

6 Bibliographie

Assadi, S. et D. Bourigault (1995). « Classification d'adjectifs extraits d'un corpus pour l'aide à la modélisation de connaissances », In *Troisièmes journées internationales d'analyse des données textuelles*, pp. 313-320.

Bourigault, D. et M. Slozdzian (1999). « Pour une terminologie textuelle », *Terminologies nouvelles* 19, pp. 29-32.

Daille, B. (2001). "Qualitative terminology extraction: Identifying relational adjectives." In Bourigault, D., C. Jacquemin and M.C. L'Homme (eds.). *Recent Advances in Computational Terminology*, Amsterdam / Philadelphia: John Benjamins, pp. 149-166.

Heid, U. (1994). "On the Way Words Work Together – Topics in Lexical Combinatorics." In Martin, W. et al.

(Ed.), *Euralex '94 Proceedings*, Amsterdam, pp. 226-257.

Jacquemin, C. (2001). *Spotting and Discovering Terms through Natural Language Processing*, Cambridge: MIT Press.

Langages (1997). Numéro spécial sur le traitement automatique de l'adjectif, 126.

L'Homme, M.C. (1998). « Le statut du verbe en langue de spécialité et sa description lexicographique ». *Cahiers de lexicologie* 73(2): 61-84.

L'Homme, M.C. (2002). « Fonctions lexicales pour représenter les relations entre termes », *Traitement automatique des langues (TAL)* 43(1), pp. 19-41.

L'Homme, M.C. (2003). "Capturing the Lexical Structure in Special Subject Fields with Verbs and Verbal Derivatives. A Model for Specialized Lexicography", *International Journal of Lexicography* 16(4).

Lecomte, J. (1998). Le catégoriseur Brill14-JL5 / WinBrill-0.3, INaLF / CNRS.

http://www.inalf.fr/winbrill/BRILL14-JL5_WinBrill.doc

Maniez, F. (2002). « Distinguer les termes des collocations : étude sur corpus du patron <Adjectif – Nom> en anglais médical. », In *Actes du colloque TALN de Nancy*, 24-27 juin 2002, Tome 1, pp. 345-350.

Mélis-Puchulu, A. (1991). « Les adjectifs dénominaux : des adjectifs de "relation" », *Lexique* 10, pp. 33-60.

Meyer, I. and K. Macintosh (1996). "The Corpus from a Terminographer's Viewpoint." *International Journal of Corpus Linguistics* 1(2), pp. 339-348.

Monceaux A. (1997). « Adjectifs de relation, complémentation et sous classification », *Langages* 126, pp. 39- 59.

Normand S. and N. Bourigault (2001). "Analysing adjectives used in a histopathology corpus with NLP tools". *Terminology* 7(2), pp. 155, 164.

Sager, N. (1987). "Computer processing of narrative information." In Sager, N. et al. (eds.). *Medical Language Processing. Computer management of narrative data*, Reading, MA: Addison-Wesley, pp. 3-21.

Terminology (2004). Special issue on *Recent Trends in Computational Terminology* 10(1).

Traitement automatique des langues (TAL). (2002). Numéro spécial sur *La structuration des terminologies* 43(1).

Zweigenbaum, P. and N. Grabar, (2000). « Liens morphologiques et structuration de terminologie », In *IC 2000. Ingénierie des connaissances*, pp. 325-334.

ANNEXE 1 : Répartition des adjectifs dans trois corpus spécialisés

	CI	CM	CD
Dérivés sémantiques	208	407	377
Autres adjectifs ⁵	570	407	429
Adjectifs de type «clinique»	71	46	52
Expressions figées ou semi-figées, titres	136	103	125
Cas problèmes	15	40	17

Exemples d'expressions figées relevées dans le corpus : *personne morale ; petit mal ; convention collective.*

Exemples d'expressions semi-figées, c'est-à-dire des collocations : ex. *grande entreprise, caractères gras*

Exemples de titres (fréquents, notamment dans le corpus juridique) ; *Charte québécoise de la langue française ; Cour Suprême.*

Des cas d'ellipse ont été relevés dans les trois corpus, mais le corpus médical en recèle le plus grand nombre (ex. patient coronarien, logiciel libre).

ANNEXE 2 : Répartition des ADS dans trois corpus spécialisés

	CI		CM		CD	
Adjectifs dénominaux	126	60,6%	347	85,3%	295	78,2%
Adjectifs déverbaux	82	39,4%	60	14,7%	82	21,8%

⁵ La catégorie « Autres adjectifs » comprend les adjectifs qualificatifs, les numéraux et les adjectifs régissant un syntagme prépositionnel.