

## **Etude de la corrélation entre morphosyntaxe et sémantique dans une perspective d'étiquetage automatique de textes médicaux arabes**

Tatiana El-Khoury

Laboratoire TIMC-IMAG - Université Joseph Fourier-Grenoble 1,  
BP 53-38041 Grenoble  
tatiana.elkhoury@imag.fr

**Résumé** Cet article se propose d'étudier les relations sémantiques reliant base et expansion au sein des termes médicaux arabes de type « N+N », particulièrement ceux dont la base est un déverbal. En étudiant les relations sémantiques établies par une base déverbale, ce travail tente d'attirer l'attention sur l'interpénétration du sémantique et du morphosyntaxique ; il montre que, dans une large mesure, la structure morphosyntaxique de la base détermine l'éventail des possibilités relationnelles. La découverte de régularités dans le comportement de la base déverbale permet de prédire le type de relations que peut établir cette base avec son expansion pavant ainsi la voie à un traitement automatique et un travail d'étiquetage sémantique des textes médicaux arabes.

**Abstract** This paper examines the semantic relations existing in Arabic medical texts between the head and its extension in a two-noun compound, particularly when the head is a deverbal noun or a nominalization. By studying semantic relations encoded by nominalizations, this research work aims at underlining the correlation between morphosyntax and semantics notably the influence of the head noun structure on the set of semantic relations that can be established. The discovery of regularities in the functioning of the head noun allows thus to predict the type of relation that will be encoded. Such data are a pre-requisite for natural language processing and automatic part-to-speech tagging of medical Arabic texts.

**Mots-clés :** étiquetage automatique, terminologie médicale arabe, morphosyntaxe, sémantique.

**Keywords:** Part-of-speech tagging, Arabic medical terminology, morphosyntax, semantics.

La représentation des connaissances terminologiques a longtemps été envisagée à travers le prisme des liens taxonomiques et sémantiques reliant différents termes d'un même domaine. Des relations sémantiques ont ainsi été encodées entre des termes pour diverses raisons : classement documentaire, recherche d'informations (liens entre synonymes, variantes, hyperonymes et hyponymes...), organisation des connaissances dans un domaine spécialisé, etc. Cet encodage a également été largement utilisé dans le domaine de la terminographie et du traitement automatique des langues (TAL). Toutefois, la particularité du travail que nous présentons aujourd'hui consiste à encoder des relations sémantiques, non pas entre des termes donnés, mais au sein même de chaque terme, entre la base du syntagme terminologique et son expansion. Nous nous inscrivons, ce faisant, dans le prolongement des études menées depuis quelques années, notamment aux Etats-Unis<sup>1</sup>, et qui visent l'exploration des relations sémantiques, au sein de syntagmes terminologiques formés de deux lexèmes, à des fins d'étiquetage sémantique, d'amélioration des systèmes de recherche « question-réponse », de génération automatique de résumés, de construction d'ontologies et d'autres applications du TAL. Ainsi, en prélude à un étiquetage sémantique de textes médicaux arabes, nous nous proposons d'étudier, dans le présent travail, les relations sémantiques au sein des termes de type « N+N », soient les termes à expansion annective<sup>2</sup>. Parmi les différents types d'expansions - annective, qualificative, complétive ou modale (A. Roman, 1999) - qui participent à la construction de syntagmes terminologiques en arabe, nous analysons plus précisément l'expansion annective des syntagmes dont la base est formée d'un déverbal.

En étudiant les relations sémantiques établies par une base déverbale, nous désirons attirer l'attention sur l'interpénétration du sémantique et du morphosyntaxique. Nous estimons que, dans une large mesure, la structure morphosyntaxique de la base détermine l'éventail des possibilités relationnelles. Ainsi, si nous arrivons à déceler des régularités dans le comportement de la base, nous serons à même de prédire le type de relations que peut établir une base déverbale avec son expansion. C'est là un préalable à tout traitement automatique et à un travail d'étiquetage sémantique.

Une base déverbale est dérivée d'un verbe ; son expansion est donc interprétée comme un argument de ce verbe. Pour cette raison, la quasi-totalité des relations entretenues par une base déverbale et son expansion sont des relations actantielles (relations d'agent, de patient, d'instrument ou de bénéficiaire). Les études portant sur les relations actantielles au sein des termes à construction annective (N+N) tentent d'y repérer un comportement systématique et des régularités récursives afin de pouvoir modéliser ces relations. Partant de là, nous avançons l'hypothèse selon laquelle en arabe médical, il est possible de déterminer automatiquement la relation sémantique reliant la base déverbale à son expansion annective à condition de tenir compte de certains paramètres morphosyntaxiques.

Pour confirmer cette hypothèse, nous présentons dans une première partie les relations actantielles dans le discours médical arabe étudié puis détaillons dans une seconde partie les paramètres qui doivent être pris en compte dans l'identification automatique de ces relations. Nous nous aidons dans cette entreprise d'un corpus électronique de textes médicaux arabes d'environ 600 000 mots et étudions plus précisément un échantillon de 860 termes de structure annective prélevés sur ce même corpus et dont la base est un déverbal.

---

<sup>1</sup> Cf. Les travaux de Gildea et Jurafsky (2002), Rosario, Hearst et Fillmore (2002), Moldovan *et al.* (2004).

<sup>2</sup> L'expansion annective correspond au « complément de nom » de la tradition arabe qui établit avec son nom un rapport de subordination direct sans préposition.

## Les relations actantielles

### Les types de relations actantielles

L'examen de l'ensemble des expansions annectives étudiées fait ressortir un fait notable : toutes les relations sémantiques au sein des termes de type « N+N » dont N<sub>1</sub> est un déverbal de type *masdar* sont des relations actantielles d'agent ou de patient. L'*actance* est définie par Lazard (1994) comme « les faits relatifs aux relations grammaticales qui s'établissent entre le prédicat verbal et les termes nominaux qui en dépendent »<sup>3</sup>. Les actants sont les termes connexes au prédicat par opposition aux termes annexes que sont les circonstants. Par *actants*, on englobe donc les constituants syntaxiques entraînés par la valence de certaines classes lexicales (notamment le verbe) qui nécessitent d'être saturées. Ces constituants suivent ce qu'on appelle un schéma actantiel. Mais « le schéma actantiel, n'est pas une exclusivité du verbe, mais bien une caractéristique de tout processus impliquant un faire transformateur au sens le plus large du terme » comme le précise Laurendeau (2004). En fait, certains mots mobilisent l'actantialité de plain-pied sans être pour autant des verbes. Deux parties différentes du discours (par exemple « rejeter » et « rejet ») peuvent entraîner le même schéma actantiel. Pour décrire le phénomène linguistique liant ces deux parties du discours nous empruntons à Mel'čuk *et al.* (1995) le terme de *dérivation syntaxique*. Parmi les types de dérivations syntaxiques, la nominalisation du verbe nous importe tout particulièrement dans ce travail. En arabe, la nominalisation du verbe correspond au « *masdar* » appelé aussi « nom verbal » ou « nom de procès ». Le *masdar* est un déverbal qui comporte des traits nominaux (il admet l'article et les marques casuelles) mais aussi des traits verbaux. Ainsi, au niveau sémantique, le *masdar*, décrit un procès comme le font les verbes tandis qu'au niveau syntaxique, il peut régir un complément. C'est cette actantialité commandée par le statut de la base-*masdar* que nous allons tenter d'illustrer dans ce qui suit.

- Dans la **relation actantielle d'agent**, l'expansion annective renvoie à un actant à l'origine de l'action exprimée par la base ou l'actant responsable de l'existence de la base syntagmatique. L'expansion annective a donc le rôle de l'acteur ou de l'instigateur de l'action sous-tendue par la base nominale.

| Terme arabe                | Nom de base –<br>Nom d'expansion                     | Equivalent français                               |
|----------------------------|--|---|
| سريان الدم                 | <i>Circulation – sang</i>                            | La circulation du sang                            |
| تنامي اللثة                | <i>Développement –<br/>gencive</i>                   | L'hyperplasie gingivale                           |
| هجوم خلايا نخاع<br>المزروع | <i>Attaque – cellules [de<br/>la moelle osseuse]</i> | La réaction de greffe<br>contre hôte <sup>4</sup> |

<sup>3</sup> Tesnière (1959) fut le premier à introduire la notion d'actant. Il l'a définie comme suit « êtres ou choses qui, à un titre quelconque, et de quelque façon que ce soit, même au titre de simples figurants et de la façon la plus passive, participent au procès » (Tesnière, 1959 : 102).

<sup>4</sup> Ce terme est l'équivalent du terme anglais « Graft-versus-host disease » employé parfois sous sa graphie originale anglaise dans les textes français.

Dans les exemples susmentionnés, c'est le sang qui circule, la gencive qui se développe et les cellules de la moelle osseuse qui attaquent les tissus du receveur.

- Dans la **relation actantielle de patient**, l'expansion annective renvoie à un actant qui subit l'action de la base ou sur lequel l'action a lieu. Comme le laisse deviner son étymologie, « patient » du latin *patiens*, participe présent du verbe *patior* (souffrir, supporter), représente celui sur qui on agit, par opposition à agent, celui qui agit (du verbe latin *ago*, « agir »). La relation actantielle de patient s'établit donc entre une base qui agit ou qui traduit une action et une expansion qui subit.

| Terme arabe     | Nom de base –<br>Nom d'expansion | Equivalent français    |
|-----------------|----------------------------------|------------------------|
| استزراع الخلايا | <i>Culture – cellules</i>        | La culture de cellules |
| تصفية الدم      | <i>Filtration – sang</i>         | La dialyse             |
| زراعة الرئة     | <i>Greffe – poumon</i>           | La greffe du poumon    |

Dans les termes ci-dessus, l'expansion subit l'action dénotée par la base. Ainsi, les « cellules » sont « cultivées », le « sang » filtré et le « poumon » greffé.

## La modélisation des relations actantielles

Les récentes études portant sur les relations actantielles au sein des termes à construction annective tentent d'y repérer des régularités récursives de sorte qu'un sous-ensemble de faits puisse rendre compte de l'ensemble. La découverte d'une telle récursivité est une condition *sine qua non* préalable à tout traitement informatique. Mais la modélisation des relations actantielles au sein de termes à base déverbale ne fait pas l'unanimité. Si des auteurs (Lapata 2002) affirment qu'elles sont susceptibles de faire l'objet d'une caractérisation linguistique, d'autres (Hobbs *et al.* 1993) estiment qu'elles nécessitent pour leurs interprétations des paramètres pragmatiques et contextuels qui rechignent à toute modélisation. En arabe médical, nous estimons qu'il est possible de voir des régularités concernant le comportement des bases déverbaux car seuls des emplois atypiques semblent déroger à la « règle » et nécessitent par conséquent des éléments pragmatiques pour l'interprétation de la relation sémantique qui unit base et expansion. Quant aux régularités, elles peuvent être captées grâce à trois paramètres : la **valence** de la base déverbale, le degré d'**agentivité** du verbe dont elle est dérivée et l'implication de son **schème**. Nous développons ces trois paramètres dans ce qui suit.

## Les paramètres pour l'identification automatique

### La valence de la base déverbale

La valence du verbe est le nombre de compléments à lui donner pour construire un énoncé simple et complet. Elle est tributaire, en arabe, de son appartenance à l'un des trois groupes suivants : transitif direct, transitif indirect et intransitif. Le verbe peut être *transitif direct* et régir un complément d'objet sans intermédiaire, *transitif indirect* et admettre un complément d'objet précédé d'une préposition ou encore *intransitif* ; dans ce cas il n'admet pas de complément d'objet.

Cette différence de régime se répercute sur les *masdars* issus de chacun de ces types de verbes. Ainsi, la base-*masdar* d'un verbe intransitif ou transitif indirect, n'accepte, en théorie,

qu'une relation d'agent avec son expansion. A titre d'exemple, si le verbe de l'énoncé /*tadjallata ad-damm*<sup>5</sup> (glosé « a coagulé le sang ») est nominalisé, son sujet original /*ad-damm*/ transformé en complément de nom du *masdar* /*tadjallut*/ « coagulation » occupera le rôle d'agent dans /*tadjallut ad-damm*/ (la coagulation du sang). Par contre, dans le cas d'une base qui est *masdar* d'un verbe transitif direct, celle-ci peut commander soit une relation d'agent soit une relation de patient avec son expansion. Ainsi, lorsque le verbe de l'énoncé « *shakhhasa at-tabibu al-marada* » (glosé « a diagnostiqué le médecin la maladie ») est nominalisé, le *masdar* peut entretenir deux types de relation avec son expansion. On peut soit obtenir « le diagnostic du médecin », la relation base-expansion est alors une relation d'agent, soit « le diagnostic de la maladie », la relation base-expansion est, dans ce cas, une relation de patient.

Dans les termes médicaux à expansion annective, la valence des *masdars* formant les bases terminologiques semble avoir donc une incidence sur le type de relation sémantique reliant bases et expansions. En d'autres termes, l'intransitivité ainsi que la transitivité directe et indirecte commandent la relation sémantique et l'orientent dans un sens donné (agent ou patient). Mais lorsque la valence du *masdar* permet différentes relations comme dans le cas d'un *masdar* issu d'un verbe transitif direct, existe-il une relation prioritaire ? et si oui, laquelle ? Pour répondre à ces questions, nous confrontons les données recueillies et interprétons les résultats.

Sur les 860 syntagmes terminologiques de structure annective dont la base est un *masdar*, 603 bases sont dérivées de verbes transitifs directs. Quoique ces *masdars* admettent les deux relations d'agent et de patient, il semblerait que, dans le corpus étudié, les *masdars* de verbes transitifs directs n'acceptent qu'une relation de patient ; les relations actantielles d'agent étant confinées à quelques rares exemples de phraséologismes terminologiques ou d'emplois atypiques. Plus de 99% des bases-*masdars* dérivées de verbes transitifs directs établissent, en effet, avec leur expansion une relation de patient. Voici quelques exemples illustrant ce type de relation :

| Terme arabe   | Nom de base –<br>Nom d'expansion | Equivalent français |
|---------------|----------------------------------|---------------------|
| تشبيط المناعة | <i>Dépression –<br/>immunité</i> | L'immunodépression  |
| زرع القلب     | <i>Greffe – cœur</i>             | La greffe du cœur   |
| بيع الأعضاء   | <i>Vente – organes</i>           | Le trafic d'organes |

Dans ces exemples, l'« immunité » est « déprimée », le « cœur » greffé, et les « organes » vendus. Les expansions des syntagmes susmentionnés entretiennent donc avec leurs bases une relation actantielle de patient. C'est le cas de la quasi-totalité des bases-*masdars* dérivées d'un verbe transitif direct. La relation *masdar*-patient semble donc prioritaire par rapport à la relation *masdar*-agent dans la formation terminologique arabe. La présence de cette relation pourrait s'expliquer par la raison suivante : Lorsque la base est un *masdar* dérivé d'un verbe transitif direct, il semblerait que la charge spécifique véhiculée par les deux actants

<sup>5</sup> Les symboles de transcription utilisés dans cet article sont ceux de l'Encyclopédie de l'Islam. La transcription de certains caractères de l'arabe n'étant pas possible avec nos logiciels de traitement de texte, les phonèmes emphatiques sont mis en gras : **h, s, d, t, z et k**.

possibles, l'agent et le patient, ne soit pas dotée de la même importance; le patient aurait, en terminologie médicale arabe, une charge spécifique plus pertinente, plus révélatrice sur le plan sémantique que celle de l'agent. Il s'agit donc d'optimiser les traits conceptuels du syntagme terminologique en choisissant la charge spécifique la plus à même d'affiner la place du syntagme dans une sous-catégorie.

Seuls quelques rares cas de bases-*masdars* dérivées de verbes transitifs directs dérogent à cette tendance en établissant une relation d'agent avec leur expansion. Il s'agit essentiellement de phraséologismes terminologiques ainsi que d'usages atypiques ou ambigus susceptibles de se prêter aux deux relations d'agent et de patient.

Les 257 bases-*masdars* restantes sont dérivées de verbes intransitifs et transitifs indirects et entretiennent, en principe, avec leur expansion une relation actantielle d'agent comme dans les exemples suivants :

| Terme arabe       | Nom de base –<br>Nom d'expansion     | Equivalent français       |
|-------------------|--------------------------------------|---------------------------|
| انكماش حجم التوتة | <i>Résorption – taille du thymus</i> | L'involution du thymus    |
| تلازن الدم        | <i>Agglutination - sang</i>          | L'agglutination du sang   |
| استرخاء العضلات   | <i>relâchement – muscles</i>         | Le relâchement musculaire |

Dans les syntagmes terminologiques susmentionnés, c'est la « taille du thymus » qui se résorbe, les « cellules sanguines » qui s'agglutinent et les « muscles » qui se relâchent. Les référents des expansions sont donc les agents des actions dénotées par les dénominations des bases.

*A priori*, lorsque la base est un *masdar* dérivé d'un verbe intransitif ou transitif indirect, toute la charge spécifique doit être concentrée dans l'agent en l'absence de tout autre actant. Mais l'encodage des relations sémantiques actantielles n'obéit pas à cette systématisme simplificatrice car les rôles sémantiques (agent, patient...) ne se superposent pas forcément aux rôles grammaticaux (sujet, complément d'objet direct...). Pour cette raison, un *masdar* issu d'un verbe intransitif est susceptible d'accepter, dans certains cas, un patient quand bien même son verbe ne pourrait pas accepter un complément d'objet direct. Pour mieux comprendre ce point épineux, force est de prendre en compte le second paramètre, celui de l'agentivité du verbe dont est issue la base-*masdar*.

### L'agentivité du verbe de la base déverbale

L'agentivité est définie comme la participation de l'agent au procès dénoté par le verbe. Le critère de cette « participation » est le contrôle ou du moins le degré de contrôle de l'agent sur le procès. Selon Roman (1990), il y a en arabe trois degrés d'agentivité : dans *l'agentivité pleine*, l'agent contrôle effectivement le procès ; dans *l'agentivité partielle*, l'agent ne contrôle le procès qu'en partie ; dans *l'agentivité zéro*, l'agent ne contrôle pas le procès. Les verbes d'action sont ainsi des verbes pleinement agentifs, les verbes d'émotion et de réaction des verbes partiellement agentifs et les verbes d'état des verbes non-agentifs.

L'agentivité des verbes se répercute sur leur *masdar*. Ainsi, les verbes intransitifs ayant une agentivité pleine ou même partielle acceptent un agent comme actant mais pour les verbes

non-agentifs, les agents ou plutôt les « pseudo-agents » qui n'ont aucun contrôle sur le procès du verbe sont assimilés à des patients.

La question de « pseudo-agent » ou de « patient » pour des verbes intransitifs remet au centre des discussions la distinction traditionnelle faite en arabe entre « sujet réel » et « sujet grammatical ». Comme le fait remarquer Hamzé (1999), « les grammairiens arabes distinguent nettement dans leur théorisation sur le /fâ<sup>c</sup>il/ (littéralement : celui qui fait) entre les plans syntaxique et sémantique ». Sur le plan syntaxique, le /fâ<sup>c</sup>il/ est le sujet du verbe ; on l'appelle le /fâ<sup>c</sup>il nahwiyy/ (le sujet grammatical) alors que sur le plan sémantique, il est soit l'agent qui fait véritablement le procès (il est alors appelé /fâ<sup>c</sup>il haqîqiyy/) soit il ne l'est pas ; dans ce cas, il reste cantonné à son rôle de sujet grammatical /fâ<sup>c</sup>il nahwiyy/. Ce sujet, qui ne fait pas le procès, accompagne les verbes dits non-agentifs<sup>6</sup>. Ainsi, Al-Makhzûmî (1986 : 47) estime qu'il y a une différence sensible entre un sujet qui choisit de faire l'action du verbe comme /sâfara Khalid/ « Khaled est parti en voyage » et un autre qui subit cette action comme dans /'inkasara az-zudjâdju/ « le verre s'est cassé » (construit avec un verbe sur le schème /infa<sup>c</sup>ala/) qui se rapproche sémantiquement de /kusira az-zudjâdju/ « le verre a été cassé » (construit avec la forme simple du verbe /kasara/ à la voix passive).

Toutefois, pour pouvoir repérer automatiquement les relations actantielles établies par une base déverbale, il faut trouver un critère formel pour distinguer d'une part les verbes intransitifs à agentivité pleine et partielle qui acceptent un agent et d'autre part les verbes intransitifs à agentivité nulle qui ne peuvent accepter sémantiquement qu'un patient. Or, les schèmes produisant des verbes à agentivité nulle ou zéro tels que /fa<sup>c</sup>ila/ (ex. /marida/ « tomber malade ») ne peuvent pas constituer un critère fiable car ils produisent également des verbes à agentivité partielle (ex. /dahika/ « rire »). Pour les fins d'un traitement automatique, nous estimons que le critère le plus fiable est l'existence ou pas, pour un verbe intransitif, d'un participe actif<sup>7</sup>. Le sens des verbes à agentivité zéro, tels que les verbes d'état, exclut en fait le participe actif car les valeurs d'achèvement et de progressivité de ce dernier sont incompatibles avec la qualité ou l'état durable et constant caractérisant les verbes d'état. Dans ces verbes, le sujet n'a aucune prise sur le procès. La formation d'un participe actif (qui a le sens de « celui qui fait ») est donc impossible. Faute d'avoir un participe actif, la morphologie de la langue arabe leur associe ce qu'il y a de plus analogue : « l'adjectif analogue » /as-sifa(t) al-mušabbaha(t)/. Cet adjectif est une forme dérivée d'un verbe intransitif et dont le fonctionnement est « analogue » à celui du participe actif. Il exprime, cependant, des qualités durables contrairement au participe actif et ne peut être dérivé que d'un verbe trilitère intransitif. Dans ce qui suit, nous confrontons les données obtenues sur les verbes intransitifs, appliquons le critère participe actif *vs.* adjectif analogue et interprétons les résultats.

Avant d'exposer les données, un récapitulatif s'impose. Sur les 860 syntagmes terminologiques de structure annective formés avec des bases déverbales, 603 bases-*masdars* sont dérivées de verbes transitifs directs. Ces bases, nous l'avons vu, acceptent dans la quasi-totalité des cas (précisément 99,3%) un patient. Les quelques rares cas d'agent sont des usages largement atypiques ou phraséologiques. Dans ce qui suit, nous nous intéressons donc

---

<sup>6</sup> De nombreux auteurs abordent cette notion de façon relativement succincte, parmi lesquels nous citons Al-'Astarâbâdî (m. 686/1287) qui cite dans Šarh aš-Šâfiya une catégorie de verbes dont le sujet réel n'est pas mentionné et dont la majeure partie de verbes exprime des pathologies tels que /djunna/ (devenir fou), /zukima/ (attraper un rhume), /humma/ (avoir de la fièvre), etc. Al-'Astarâbâdî ajoute que s'il n'est pas fait mention du sujet réel, c'est parce qu'il est communément admis que c'est Dieu lui-même. Le sujet fut donc supprimé parce que redondant

<sup>7</sup> Nous n'évoquons pas le participe passif puisqu'il est spécifique aux verbes transitifs.

aux 257 bases-*masdars* restantes et dérivées de verbes intransitifs ou transitifs indirects. Même si *a priori* ces bases ne peuvent accepter qu'un agent, l'agentivité nulle de certains verbes intransitifs complique la tâche. Voici les données relatives aux 257 *masdars* issus de verbes intransitifs<sup>8</sup> :

*Premièrement*, sur les 257 *masdars*, moins d'une vingtaine sont issus de verbes non-agentifs soit 6,6%.

*Deuxièmement*, les *masdars* issus de verbes non-agentifs construisent tous un adjectif analogue et n'acceptent pas de participe actif.

*Troisièmement*, tous ces *masdars* entretiennent avec leurs expansions une relation actantielle de patient comme en témoignent les exemples suivants :

| Terme arabe | Nom de base –<br>Nom d'expansion | Equivalent français | Adjectif analogue du<br>verbe correspondant |
|-------------|----------------------------------|---------------------|---|
| فقر الدم    | Appauvrissement –<br>sang        | L'anémie            | فقير  |
| موت الدماغ  | Mort – cerveau                   | La mort cérébrale   | ميت   |

Dans ces exemples, le sang subit un appauvrissement en globules rouges et le cerveau est sujet à la mort<sup>9</sup>. Les actants impliqués dans ces relations sémantiques sont donc des patients. Or les verbes dont sont issues les bases-*masdars* sont intransitifs. Nous pouvons donc avancer qu'une base issue d'un verbe intransitif entretient une relation d'agent lorsque son verbe est pleinement ou partiellement agentif et une relation de patient lorsque son verbe a une agentivité nulle.

Les remarques que nous venons de formuler sur l'agentivité mettent en évidence un troisième paramètre, celui des schèmes des *masdars* dont les implications sur le type de relation actantielle sont nettes dans les résultats de cette étude.

## Les schèmes de la base déverbale

Le tableau suivant récapitule les résultats exposés plus haut et obtenus sur le corpus de l'étude. Les pourcentages très élevés de relations d'agent pour les bases-*masdars* issues de verbes intransitifs et de relations de patient pour les bases-*masdars* issues de verbes transitifs posent la question de savoir dans quelle mesure les schèmes participent à cette répartition et s'ils peuvent constituer un critère fiable capable d'être utilisé à des fins de traitement automatique.

<sup>8</sup> Par souci de commodité, nous incluons dans la suite de cet exposé sous l'appellation « intransitifs » les verbes qui ne sont pas transitifs directs soit les verbes véritablement intransitifs et les verbes transitifs indirects ; ces deux catégories ont en commun de ne pas accepter de complément d'objet direct /ma<sup>c</sup>ûl bihi/.

<sup>9</sup> Le fait de considérer le « cerveau » comme patient est intimement lié à la perception de la mort dans la culture arabe qui estime que l'on n'est pas actif dans sa propre mort mais que c'est au Tout-Puissant que revient la décision d'ôter la vie à l'humain. Un verbe synonyme de « mâ<sup>t</sup> » - qui est à l'origine de l'exemple susmentionné - celui de « tawaffâ » est plus explicite puisque la personne décédée occupe toujours la position de COD ; le sujet étant systématiquement Dieu.



|   | Nb. d'exemples | Relations d'agent | Relations de patient |
|---|----------------|-------------------|----------------------|
| Bases- <i>masdars</i> issues de verbes intransitifs | 257            | 93,4%             | 6,6%                 |
| Bases- <i>masdars</i> issues de verbes transitifs   | 603            | 0,7%              | 99,3%                |

Le schème est le « moule » syllabique qui sert à la formation des mots en arabe. Selon le schème combiné avec la racine du verbe, nous pouvons obtenir le nom verbal (ou *masdar*), le participe actif, le participe passif, le nom de lieu, le nom d'instrument, le nom d'abondance, le nom de métier, etc.

Certains schèmes, de par leur construction morphologique, se prêtent majoritairement (et parfois même exclusivement) soit à une relation d'agent, soit à une relation de patient car ils favorisent la construction de verbes transitifs ou intransitifs. Ainsi, une étude morphologique de la distribution des schèmes dans un discours donné peut tracer des tendances claires quant aux relations sémantiques *masdars*-expansions susceptibles d'être établies. Pour s'en convaincre, il suffit de regarder le tableau ci-dessous qui indique si les différents schèmes mentionnés contribuent, dans la terminologie médicale arabe, à construire des *masdars* dérivés de verbes intransitifs (donc avec une très forte probabilité d'entretenir une relation d'agent avec leur expansion) ou des *masdars* dérivés de verbes transitifs (donc établissant dans la quasi-totalité des cas une relation de patient avec leur expansion). Afin de conforter l'hypothèse évoquant le lien entre schème et relation actantielle, nous choisissons de présenter dans ce tableau des schèmes attestés par au moins 15 *masdars* dans le corpus étudié. Voici leur distribution :

| Schémes                                   | Transitifs | Intransitifs | Tendance     |
|---|------------|--------------|--------------|
| istif <sup>c</sup> âl                     | 92.5%      | 7.5%         | Transitifs   |
| inf <sup>i</sup> âl                       | 0%         | 100%         | Intransitifs |
| tafâ <sup>c</sup> ul                      | 5%         | 95%          | Intransitifs |
| tafa <sup>cc</sup> ul                     | 6.4%       | 93.6%        | Intransitifs |
| taf <sup>c</sup> ila(t)                   | 100%       | 0%           | Transitifs   |
| taf <sup>c</sup> îl                       | 98.5%      | 1.5%         | Transitifs   |
| fi <sup>c</sup> âla(t)                    | 94%        | 6%           | Transitifs   |
| fu <sup>c</sup> ûl/fu <sup>c</sup> ûla(t) | 95.6%      | 4.4%         | Transitifs   |

En conclusion, nous pouvons avancer que dans les termes arabes à expansion annective, il existe une corrélation entre les paramètres morphosyntaxiques (valence, agentivité et schèmes des bases-*masdars*) et les relations sémantiques base-expansion (relation d'agent ou de patient) pouvant être établies. Ainsi, les *masdars* issus de verbes intransitifs ne peuvent entretenir qu'une relation d'agent avec leurs expansions sauf si l'agentivité des verbes dont sont issues les bases-*masdars* est nulle. Les résultats de cette étude révèlent un autre phénomène surprenant. L'expansion annective d'un *masdar* issu d'un verbe transitif qui, *a priori*, peut établir avec sa base une relation d'agent ou de patient est, dans le corpus étudié, une relation presque exclusivement de patient ce qui laisse supposer

que cette relation est prioritaire dans la formation terminologique. Ainsi, au-delà des considérations purement syntactico-sémantiques, les résultats obtenus établissent une jonction entre grammaire et terminologie et pavent la voie à un travail d'étiquetage sémantique des textes médicaux arabes qui reste encore à faire.

## Références

- AL-MAKHZÛMÎ M. (1986). *Fî An-nahû al-<sup>c</sup>arabiyy. Naqdun wa tawjîh*. Beyrouth : Dâr Ar-Râ'id al-<sup>c</sup>arabiyy.
- COTTE P. (1999). *Langage et linéarité*. Villeneuve d'Ascq : Presses universitaires du Septentrion.
- FRANÇOIS J. (2004). *L'adjectif en français et à travers les langues*. Caen : Presses universitaires de Caen.
- GILDEA D., JURAFSKY D. (2002). Automatic Labeling of Semantic Roles. *Computational Linguistics* 28, Vol.3.
- HAMZE H. (1999). La position du sujet du verbe dans la pensée des grammairiens arabes. In (Cotte, 1999), 127-149.
- HOBBS J., STICKEL M., APPELT D., MARTIN P.M. (1993). Interpretation as Abduction. *Journal of Artificial Intelligence* 1-2, Vol.63.
- LAPATA M. (2002). The Disambiguation of Nominalizations. *Computational Linguistics* 28, Vol.3.
- LAURENDEAU M. (2004). L'actancialité de l'adjectif verbal d'origine gérondive dans deux vernaculaires du français. In (François, 2004), 53-70.
- LAZARD G. (1994). *L'actance*. Paris : Presses Universitaires de France.
- MEL'CUK I., POLGUERE A. CLAS A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- MOLDOVAN D., BADULESCU A., TATU M., ANTOHE D., GIRJU R., (2004). Models for the Semantic Classification of Noun Phrases. Proceedings of the *Computational Lexical Semantics Workshop at HLT-NAACL*, 60-67.
- ROMAN A. (1990). *Grammaire de l'Arabe*. Paris : Presses Universitaires de France.
- ROMAN A. (1999). *La créativité lexicale en arabe*, Lyon : Presses Universitaires de Lyon.
- ROSARIO B., HEARST M., FILLMORE C. (2000). The Descent of Hierarchy, and Selection in Relational Semantics. Proceedings of the *40<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL-02)*, 247-254.
- TESNIERE L. (1959). *Éléments de syntaxe structurale*. Paris : Klincksieck.