

Les noms

Il ne fait pas de doute que les noms sont la catégorie grammaticale la plus représentée – et la plus étudiée – en terminologie. Malgré les récentes ouvertures théoriques à d'autres catégories grammaticales, comme les adjectifs, les verbes et les adverbes, pour de nombreux terminologues l'implication *terme* => *nom* tient toujours.

Après l'analyse de quelques verbes des deux corpus à la lumière de la méthode du lexique-grammaire, dans ce chapitre nous nous attacherons à analyser ce que les méthodes liées au lexique-grammaire peuvent apporter à l'étude des noms en terminologie.

4.1. Catégories sémantiques ressorties lors de l'analyse de quelques constructions verbales

L'analyse d'un petit groupe de verbes dans le chapitre précédent a fait ressortir quelques catégories sémantiques de termes nominaux que l'on peut retrouver dans le discours de la médecine thermale. Ces catégories sémantiques, qui ont été retrouvées dans les deux corpus d'étude, sont :

- les **entités humaines** (*Nhum*) : il s'agit des noms qui désignent les dispensateurs et les bénéficiaires des soins thermaux, les premiers étant la plupart des fois des noms de profession, mais pouvant aussi être des noms propres de personnes, d'organismes officiels ou de stations thermales, qui désignent par métonymie les personnes qui y travaillent ;

- les **moyens** (*Nconc*) : dans cette catégorie entrent les noms relatifs à des substances, comme les médicaments et les types d'eaux, par exemple, mais aussi certains types d'appareils ;

- les **techniques de soin** et les **activités thermales** (*Nabs*) : c'est la catégorie qui compte le plus grand nombre d'éléments, apparaissant dans une quantité élevée de phrases ;

- les **noms de maladies** (*Nmal*) : dans cette catégorie, outre des hyponymes de *maladie*, nous trouvons aussi des co-hyponymes de *maladie*, tels qu'*affection* et *pathologie*, suivis d'un modifieur ;

- les **noms de partie du corps** (*Npc*) : outre des noms de membres et d'organes, on trouve dans cette catégorie aussi des formes composées typiques du style médical comme *segments corporels* ou *région lombaire*.

Ces noms peuvent être simples ou composés, sachant que, pour des raisons d'économie linguistique, les formes composées dépassent de loin, en nombre d'entrées, les formes simples.

4.2. Fréquence des verbes et de leurs V-n

Les catégories présentées à la section précédente ne couvrent qu'une partie des termes de nature nominale des deux corpus. Il y a d'autres noms qui n'entrent dans aucune des catégories citées plus haut. Parmi ces noms, il y a de nombreux noms dérivés de

verbes¹⁷⁹. Cela parce que dans les discours spécialisés les nominalisations ont une place privilégiée dans la construction du discours.

Si nous comparons le nombre d'occurrences de quelques verbes traités au chapitre précédent et des noms qui leur sont morphologiquement reliés (*V-n* dans les notations du lexique-grammaire), nous verrons que les premières sont moins nombreuses que les deuxièmes.

Corpus français

V	nombre d'occurrences	V-n	nombre d'occurrences
agir	22	action	229
contre-indiquer	15	contre-indication	134
indiquer	73	indication	500
prescrire	71	prescription	87
traiter	138	traitement	693

Tableau 1 : comparaison entre le nombre d'occurrences de quelques verbes et de leurs *V-n* dans le corpus français.

Corpus italien

V	nombre d'occurrences	V-n	nombre d'occurrences
agire	79	azione	399
controindicare	8	controindicazione	31
indicare	31	indicazione	105
migliorare ¹⁸⁰	71	miglioramento	178
somministrare	15	somministrazione	32
trattare	141	trattamento	441

Tableau 2 : comparaison entre le nombre d'occurrences de quelques verbes et de leurs *V-n* dans le corpus italien.

4.3. Les noms prédicatifs

En II.1.4., nous avons présenté une entrée – l'entrée *commande* – du *DAFA*¹⁸¹ (2000), ouvrage que nous avons qualifié de « description terminologique de qualité ». À cette occasion, nous avons vu que la description du nom était faite sur la base de sa distribution. En particulier, l'encadré « + verbe : qui fait quoi ? » avait retenu notre attention : divers schémas de phrases étaient fournis pour montrer les différents emplois du nom. Cette démarche rappelle le traitement des noms prédicatifs dans le lexique-

¹⁷⁹ La catégorie des *Nabs* atteste également bon nombre de déverbaux.

¹⁸⁰ Le verbe *migliorare* ne fait pas partie de l'ensemble de verbes traités au chapitre précédent. Toutefois, comme *miglioramento* est un nom très fréquent, nous avons voulu insérer la comparaison entre *migliorare* et *miglioramento* dans le tableau.

¹⁸¹ *Dictionnaire d'Apprentissage du Français des Affaires*.

grammaire¹⁸². D'ailleurs, les verbes listés dans l'article *commande* peuvent tous être qualifiés de verbes supports ou extensions de verbes supports. Outre les *Npréd* qui sont morphologiquement reliés à un verbe (*V-n*), il existe aussi des *Npréd* isolés. Leur identification se fait par le rétablissement d'une phrase à verbe support. Le rétablissement de la phrase à verbe support – aussi bien pour les *V-n* que pour les *Npréd* isolés – est une opération importante aussi pour l'identification des prépositions qui accompagnent les *Npréd*.

Les méthodes de description mobilisées dans le lexique-grammaire soulignent que les différents niveaux de description d'un nom (niveau purement lexical, niveau syntaxique) ne peuvent pas être traités totalement indépendamment. En effet, l'interprétation d'un nom peut dépendre de sa position dans la phrase.

Bien que cet aspect présente un intérêt indéniable pour la terminologie, une étude plus poussée aurait dépassé le cadre de cette thèse, en raison de la variété des constructions à verbe support et des critères qui permettent de les délimiter, et de la profusion de leurs propriétés syntaxiques. En revanche, nous avons pu explorer d'autres pistes, comme celles proposées dans M. GROSS et TREMBLAY (1985) sur la variabilité interne des noms composés et la notion de *figement*.

4.4. *Les faits de la syntaxe pour expliquer la continuité entre les unités de la langue générale et les unités spécialisées*

Comme nous l'avons déjà vu lors de l'introduction au lexique-grammaire (§II.3.), M. GROSS refuse l'image selon laquelle les termes sont « des mots [...] de contenu fixe » (1985 : 9). La soi-disant « exceptionnalité » des termes par rapport aux autres unités du vocabulaire de la langue n'est pas justifiée, au moins sous l'angle de la linguistique. À l'appui de cette affirmation, il porte de nombreux exemples qui montrent que les noms techniques obéissent aux mêmes contraintes syntaxiques que les noms non techniques. En conséquence de cela, la séparation entre vocabulaire général et vocabulaire technique n'a pas de raison d'être.

4.5. *Degré de figement des unités terminologiques nominales complexes*

En premier lieu, la continuité entre vocabulaire général et vocabulaire technique s'explique en recourant à la notion de *degré de figement*. À ce propos, DE GIOIA affirme que :

« Le figement n'est [...] pas une valeur absolue, et dans ce sens il est en rapport avec l'idée de continuum. Il existe des degrés de figement.

¹⁸² Un nom prédicatif désigne généralement un procès et occupe généralement la position N_I dans une phrase à verbe support :

Le patient (accuse+connait) une amélioration

= *L'amélioration du patient (est significative).*

LABELLE (1986) considère que les noms de maladie (*Nmal*) occupent la position N_I dans le schéma de phrase : N_0 *Vsup* (*avoir+souffrir de*) N_I .

Cette dernière notion de figement, formulée d'abord dans le cadre du lexique-grammaire, est applicable à l'analyse de séquences figées soumises à des contraintes non absolues et d'ordre syntaxique (mécanismes syntaxiques moins fermés), sémantique (opacité moins nette), lexical (paradigmes limités). » (2004 : 143-144)

Tout comme les unités du vocabulaire courant, les unités du vocabulaire technique peuvent présenter des degrés de figement variés. Il existe des noms composés terminologiques qui ne permettent aucune manipulation syntaxique – telles que l'insertion de déterminants ou modificateurs, la substitution d'un des composants par un synonyme ou une autre unité de la même catégorie grammaticale – mais il existe également des noms composés terminologiques qui en revanche sont moins contraints du point de vue de la combinatoire syntaxique. Ces derniers posent souvent des problèmes pour le découpage des termes. Avant d'aborder cet aspect, nous voudrions d'abord analyser quelques noms composés figés des deux corpus d'étude.

Commençons par un exemple de nom composé figé du vocabulaire courant¹⁸³. Soit la séquence *N de N*

bain de foule.

Le figement de la séquence, outre par son opacité sémantique – un *bain de foule* n'est pas un hyponyme de *bain* – est confirmé par le fait qu'elle ne permet pas de manipulations syntaxiques telles que :

- la relativation : **un bain qui est (de+en) foule* ;
- l'insertion d'un déterminant (= *la, des*) ou d'un modifieur dans la séquence : **un bain de la foule, *un bain des foules, *bain (long+exceptionnel) de foule* ;
- la substitution d'un de ses composants par un synonyme¹⁸⁴ ou un mot sémantiquement proche : **une baignade de foule, *un bain de masse.*

Toutes ces manipulations donnent lieu à des séquences inacceptables car la séquence *bain de foule* correspond à une seule unité de sens. Autrement dit, le sens de *bain de foule* ne résulte pas de la somme des sens de chacun de ses composants.

On comprendra que, dans une optique contrastive, l'attention accordée au phénomène du figement devient nécessaire dans l'assurance de qualité de la traduction. Ceci est aussi valable pour le vocabulaire courant que pour le vocabulaire technique. La traduction d'un nom composé figé sémantiquement opaque dans une autre langue peut donner comme résultat :

- 1) un terme composé figé parfaitement équivalent au terme composé de la langue source, aussi bien au niveau de la structure syntaxique qu'au niveau lexical ;
- 2) un terme composé figé ayant une structure syntaxique différente ;
- 3) un terme simple.

Voyons cela à l'aide de quelques exemples tirés de nos corpus.

Pour le premier cas, nous choisissons un terme composé du corpus français ayant une structure syntaxique *N de N*

col de cygne.

¹⁸³ Nous partons d'un nom composé de la langue courante pour introduire aux tests syntaxiques et à l'opacité sémantique.

¹⁸⁴ Pour le choix des synonymes nous avons consulté le *Dictionnaire des synonymes* d'Henri Bertaud du CHAZAUD (1989).

Les tests de manipulations syntaxiques illustrés plus haut pour *bain de foule* produisent des séquences inacceptables : **un col qui est d'un cygne*, **un col du cygne*, **un col long de cygne*.

Le terme désigne un type de robinet, donc n'a aucune relation ni avec un col, ni avec un cygne. La traduction en italien ne pose aucun problème : l'équivalent de *col de cygne* est le terme *collo di cigno*. L'équivalence entre les deux termes composés est totale, tant au niveau syntaxique qu'au niveau sémantique.

Un cas comme celui que nous venons d'analyser s'explique par le recours à des figures rhétoriques comme la métaphore ou la métonymie pour créer de nouvelles unités lexicales, aussi fréquent dans le vocabulaire courant que dans les vocabulaires techniques.

Pour l'analyse de la catégorie 2), nous partons d'un terme composé du corpus italien ayant une structure syntaxique *NA*,

*camminamento vascolare*¹⁸⁵.

La séquence ne permet pas de manipulations syntaxiques telles que :

- la relativation : **il camminamento che è dei vasi sanguigni* ;
- l'insertion d'un modifieur comme un adverbe d'intensité (= *molto, poco*) : **un camminamento (molto+poco) vascolare* ;
- la substitution d'un des deux éléments par un synonyme : **cammino vascolare*, **camminamento dei vasi*.

La traduction de ce terme composé en français donne comme résultat le terme composé *parcours de marche*¹⁸⁶.

Pour la catégorie 3), nous prenons l'exemple du terme composé français de structure *N de N*

sortie de bain.

De même que pour les termes composés des catégories 1) et 2), des tentatives de modifications de la séquence figée aboutissent à des séquences inacceptables. Ainsi, il n'est pas possible de :

- changer un des composants : **entrée de bain*, **sortie de douche* ;
- insérer un déterminant (= *le, un*) dans la séquence : **sortie du bain*, **sortie d'un bain*.

Il est indispensable de bien identifier la séquence, d'autant plus que la traduction en italien a pour résultat un terme simple, *accappatoio*.

Les cas que nous avons analysés jusqu'ici sont des cas de séquences figées sémantiquement opaques. Le sens du nom composé n'est pas compositionnel et des manipulations syntaxiques – comme celles que nous avons observées plus haut – donnent lieu à des séquences inacceptables. Le repérage de ces noms composés dans un corpus est donc relativement aisé.

Toutefois, dans les discours spécialisés il existe de nombreuses séquences composées dans lesquelles un des éléments spécifie en quelque sorte l'autre, à la différence des noms composés analysés plus haut. Dans ces séquences, la variation syntaxique est

¹⁸⁵ Le terme désigne une pratique thermale.

¹⁸⁶ Le corpus italien contient aussi le terme *percorso vascolare*, qui toutefois ne désigne pas une pratique thermale mais un dispositif, qui correspondrait plutôt au terme *couloir de marche*, repéré dans le corpus français.

plus importante. De même, du point de vue sémantique, ces séquences affichent un moindre degré d'opacité. La plupart des fois, il est possible d'expliquer le rapport de ces séquences figées à un de ses composants par le biais d'une relation d'hyponymie ou d'hyperonymie : un *bain bouillonnant* est un type de bain, une *eau soufrée* est un type d'eau, une *affection dermatologique* est une affection de la peau. La séquence composée peut même présenter des variantes. Tel est le cas de certains noms composés de maladie, comme

phénomène de Raynaud

dont les variantes suivantes sont attestées dans le corpus :

- *syndrome de Raynaud*,
- *maladie de Raynaud*,
- *Raynaud*.

Cependant, lorsqu'un nom composé contient un nom propre, il est fort probable de le retrouver sous sa forme canonique dans le corpus. Dans bon nombre de cas, en revanche, il n'en est pas ainsi et cela peut compliquer la tâche de découpage des termes, comme nous le verrons dans ce qui suit.

4.6. Problèmes dans le découpage des termes nominaux composés affichant un moindre degré de figement

La continuité entre vocabulaire courant et vocabulaire technique est confirmée par le fait que de nombreux termes composés non entièrement figés subissent les mêmes contraintes syntaxiques que les syntagmes nominaux non terminologiques. Ainsi, il peut arriver qu'un terme composé : a) n'apparaisse pas dans l'ordre séquentiel, mais que ses composants soient séparés par un nombre X de mots, b) soit réduit en raison d'une pronominalisation partielle (M. GROSS et TREMBLAY 1985 : 18). Dans ces cas-là, il faut procéder au rétablissement des termes, si l'on vise un recensement des termes qui soit le plus complet possible.

4.6.1. Découpage des termes composés dont les composants ne sont pas contigus

Pour les cas de type a), une bonne stratégie de repérage est l'observation des contextes des conjonctions de coordination. Il arrive souvent dans les textes spécialisés – surtout dans les listes et les énoncés de types définitoires – que les conjonctions de coordination (ainsi que des virgules) soient utilisées pour éviter de répéter un nom. Nous procédons à un rétablissement manuel de termes composés à partir de quelques exemples des deux corpus. Après chaque exemple nous listons les termes extraits manuellement.

Commençons par quelques exemples du corpus français :

1) « Les pratiques thermales associent bains en baignoire, en piscine **ou** en couloir de marche, bains locaux, massages sous l'eau et soins complémentaires (électrothérapie, ultrasons, infrarouges et cataplasmes de paraffine). »

- *bains en baignoire*
- *bains en piscine*

- *bains en couloir de marche*

2) « L'étuve thermique est un nouveau soin proposé aux patients en cure thermique rhumatologique **ou** rhumato-phlébologique à la station thermique de La Léchère. »

- *cure thermique rhumatologique*

- *cure thermique rhumato-phlébologique*

3) « L'évolution du périmètre de marche **et** de la consommation médicamenteuse est représentée sur le tableau IV. »

- *évolution du périmètre de marche*

- *évolution de la consommation médicamenteuse*

Passons à quelques exemples du corpus italien :

4) « È da tempo noto l'effetto terapeutico che le acque termali, sulfuree, salso-solfato-alcaline, salso-bromoiodiche **e** sulfureo-salco-bromo-iodiche, esercitano sulle affezioni flogistiche, acute **e** croniche, delle mucose respiratorie. »

- *acque termali*

- *acque sulfuree*

- *acque salso-solfato-alcaline*

- *acque salso-bromoiodiche*

- *acque sulfureo-salco-bromo-iodiche*

- *affezioni flogistiche acute delle mucose respiratorie*

- *affezioni flogistiche croniche delle mucose respiratorie*

5) « In particolare, per gli arti si utilizzano anche appositi contenitori dove immergere le gambe (fanghi a cassetta **o** in mastelli). »

- *fanghi a cassetta*

- *fanghi in mastelli*

Le rétablissement des termes composés à partir des exemples que nous venons d'examiner est relativement aisé. Toutefois, il n'en est pas toujours ainsi. Considérons par exemple :

6) « La rencontre avec les familles souvent accompagnantes est également très profitable (reconnaissance **et** information sur le trouble, motivation à une plus grande implication de leur part, etc.) »

Si dans les exemples 1) à 5) les noms composés étaient formés sur un même nom, dans l'exemple 6) nous nous trouvons face à deux noms composés qui n'ont pas la même base nominale, ce qui complique le rétablissement des termes. Si nous nous limitons à appliquer la même méthode, nous obtiendrions deux termes composés dont l'un est inacceptable :

- **reconnaissance sur le trouble*

- *information sur le trouble*

Car le nom *reconnaissance*, à la différence du nom *information*, ne prend pas la préposition *sur*, mais la préposition *de*. Cela devient plus clair si on rétablit les phrases à verbes support dans lesquelles entrent les noms *reconnaissance* et *information* en position prédicative. Ainsi, nous aurons :

(1) *On procède à la reconnaissance du trouble*

Et

(2) *On donne une information sur le trouble.*

Par effacement du verbe support, les termes rétablis seront alors :

- *reconnaissance du trouble*

- *information sur le trouble.*

Dans le corpus français nous avons également retrouvé bon nombre de termes composés dont les composants sont séparés par la suite de conjonctions *avec ou sans* : tel est le cas de *bains avec ou sans radon*, suite à partir de laquelle on peut rétablir les termes :

- *bains avec radon*

- *bains sans radon.*

Ce procédé stylistique est en revanche peu attesté dans le corpus italien (seules 2 occurrences de la suite de conjonctions *con o senza* ont été retrouvées).

4.6.2. Rétablissement de termes réduits en raison d'une pronominalisation partielle

Pour des raisons stylistiques – notamment pour éviter d'alourdir des phrases qui sont déjà très denses – les auteurs de textes spécialisés recourent au procédé de la pronominalisation. M. GROSS et TREMBLAY expliquent le phénomène de la façon suivante :

« Les discours où la pronominalisation opère ont des formes variées. Il arrive qu'une partie d'un terme soit réduite lorsque ce terme est rattaché par une conjonction à un terme qui lui est sémantiquement lié et qui paraît dans sa forme complète. La réduction évite alors de répéter l'élément qui est commun aux deux termes. » (1985 : 24)

La partie du terme que l'on veut éviter de répéter est d'habitude « substituée par un pronom ou par un adjectif possessif coréférent à une partie explicite ». (*Ibid.*)

Ces cas reflètent une référence lexicale que le terminologue doit retrouver pour l'extraction de terminologie. Les deux corpus sont riches en exemples de ce type.

Comme nous l'avons fait plus haut, nous listons les termes extraits manuellement à partir de chaque exemple.

Quelques exemples tirés du corpus français :

1) « Ces contrôles volontaires s'ajoutent aux contrôles obligatoires faits par les services de santé sur les boues, **leur** maturation, **leur** préparation, **leur** utilisation et sur les eaux thermales de **leur** point d'émergence à **leur** utilisation. »

- *maturation des boues*

- *préparation des boues*

- *utilisation des boues*

- *point d'émergence des eaux thermales*

- *utilisation des eaux thermales*

2) « Il y a tout d'abord la prise en charge de la douleur (douches, application de boues thermales) puis **celle** de la spasticité (bains thermaux froids et mobilisations spécifiques suivant les schémas d'inhibition de la spasticité). »

- *prise en charge de la douleur*
- *prise en charge de la spasticité*

3) « Les migraines sont la troisième indication classique en hépato-gastro-entérologie ; mais pas n'importe quelles migraines ; les migraines sans aura et surtout **celles** à expression digestive, c'est à dire celles qui s'accompagnent de nausées, de vomissements et de troubles intestinaux. »

- *migraines sans aura*
- *migraines à expression digestive*

4) « Le bain est une manière d'appliquer à l'individu chaleur et pression. Il peut être sédatif ou stimulant, selon **sa** température, **sa** durée, **sa** concentration. »

- *température du bain*
- *durée du bain*
- *concentration du bain*

Pour le corpus italien, nous avons choisi ces exemples :

5) « In particolare l'acqua sulfurea è utile per trattare le malattie delle vie respiratorie, quella bicarbonata cura il "diabete della maturità". »

- *acqua sulfurea*
- *acqua bicarbonata*

6) « **Quello** naturale è il fango così come si presenta alla sorgente, di formazione spontanea (quindi comprende il fango sorgivo e **quello** di salsa). »

- *fango naturale*
- *fango sorgivo*
- *fango di salsa*

7) « Tale timpanogramma viene registrato nelle Otitis Medie Secretive dove l'aria, nella cassa timpanica, è sostituita del tutto o in parte da un essudato, e la membrana timpanica perde la **sua** mobilità e il variare della pressione durante l'esame non dà luogo a modificazioni di impedenza, tale curva è tipica anche nella timpanosclerosi. »

- *membrana timpanica*
- *mobilità della membrana timpanica*

8) « Esistono strette correlazioni tra il sistema linfatico e **quello** venoso. »

- *sistema linfatico*
- *sistema venoso*

4.6.3. Conséquences pour le traitement automatique

Si pour un être humain avec une connaissance même inconsciente des règles de la grammaire le rétablissement de ces termes est une tâche relativement aisée, il n'en est pas de même pour un logiciel. Ce dernier pourra reconnaître sans problèmes un nom composé apparaissant sous sa forme canonique, mais le rétablissement de termes pour les cas cités plus haut sera bien plus ardu, voire impossible. En conséquence de cela, bien des termes composés n'apparaissant jamais sous sa forme canonique risqueraient de passer sous silence. En fait, la forme canonique attribuée à un terme composé par les terminologues

n'est pas nécessairement la plus fréquente : bien souvent, il s'agit de la forme qui simplifie le plus la description des termes. Toutefois, si le logiciel exploite des ressources linguistiques, il est possible de les récupérer – au moins pour certains des cas cités plus haut – par le biais de grammaires locales. C'est ce que nous allons montrer à la section suivante.

4.6.4. Retour à Unitex : définition de variables dans les grammaires locales pour le rétablissement de termes composés

Dans III.1.7. (et sous-sections relatives) nous avons présenté les graphes d'Unitex : leur création et les possibilités d'exploitation qu'ils offrent pour l'analyse de textes (spécialisés ou non). Nous y revenons à présent pour illustrer la possibilité de rétablir la forme canonique des termes apparaissant sous d'autres formes et procéder ainsi à l'extraction d'information. Pour que cela soit possible, il faut assigner des variables aux boîtes des graphes. Dans la fenêtre d'édition des graphes, il est possible d'assigner une variable à une boîte (les crochets rouges définissent une variable d'entrée, les crochets bleus une variable de sortie). Une fois que l'on a identifié les composants des termes à reconstituer à l'aide de variables dans un graphe, on peut produire des sorties affichant le terme sous sa forme canonique, même s'il n'apparaît jamais sous cette forme dans le texte.

Reprenons quelques cas cités plus haut en guise d'exemples pour illustrer cette méthode.

Dans les deux corpus, nous avons des exemples de termes composés non entièrement figés dont les composants peuvent être séparés par des virgules ou des conjonctions de coordination. Telle est la situation de séquences Nom Adjectif du type :

- *affections neurologiques ou rhumatismales*
- *acque termali, sulfuree, salso-solfato-alcaline.*

Pour pouvoir rétablir des séquences non contiguës, nous avons élaboré des graphes comme celui de la figure ci-dessous :

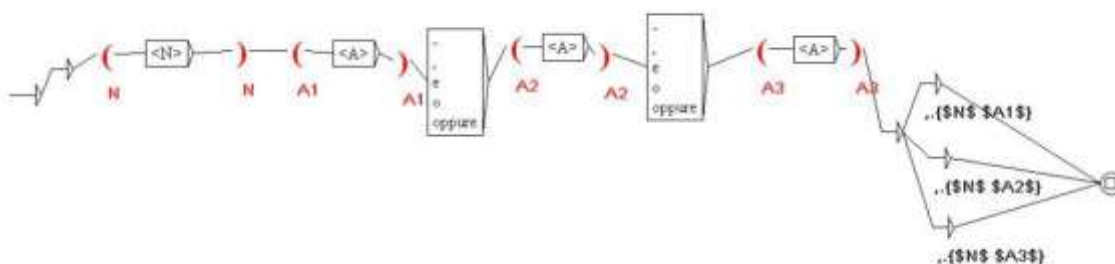


Figure 1 : graphe élaboré pour le rétablissement des séquences N A non contiguës du corpus italien.

Pour appliquer ce graphe et procéder à l'extraction de terminologie, il suffit de choisir dans le menu Locate Pattern (I.1.5.) l'option « Merge with input text ». Les résultats de l'extraction seront visibles dans l'affichage de la concordance :

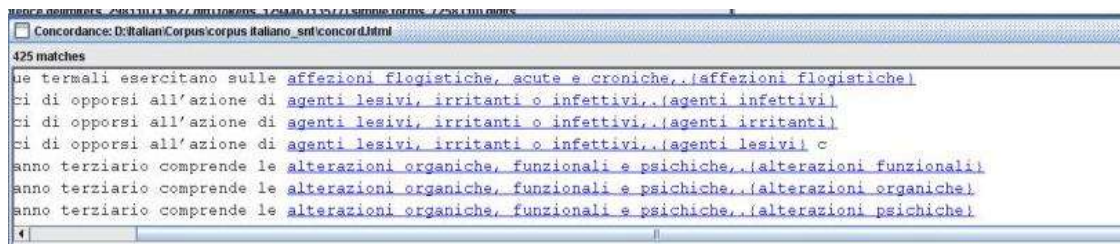


Figure 2 : affichage d'un extrait de la concordance produite par l'application du graphe de la figure 1.

Voici le graphe réalisé pour l'extraction des séquences séparées par la suite de conjonctions *avec ou sans* :

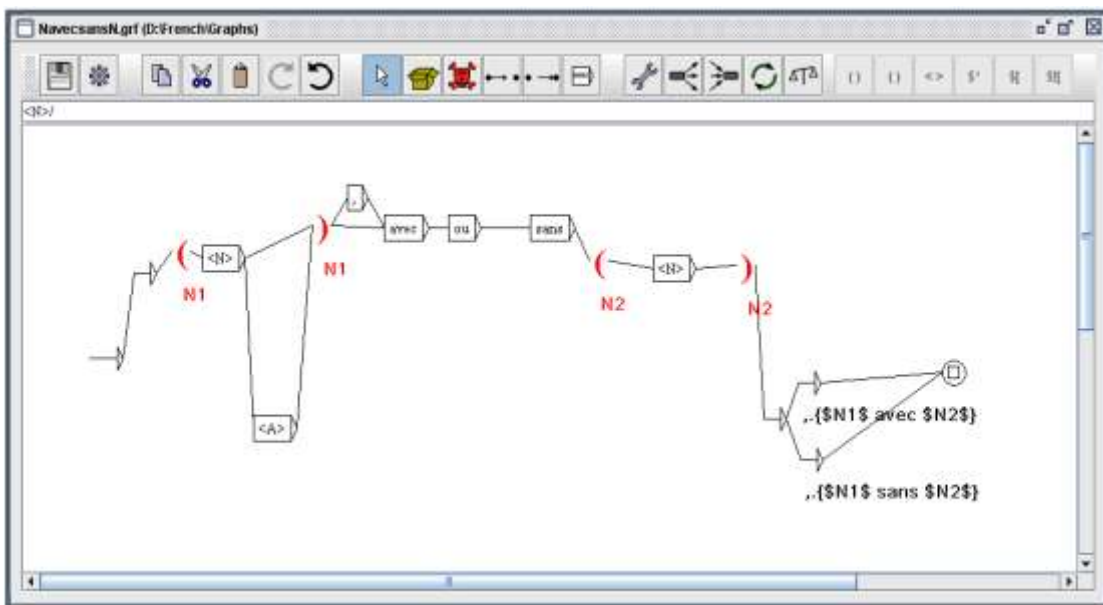


Figure 3 : graphe élaboré pour le rétablissement de séquences N (et N A) Prép (*avec/sans*) N du corpus français.

Et voici ci-dessous un extrait de la concordance produite par l'application de ce graphe :

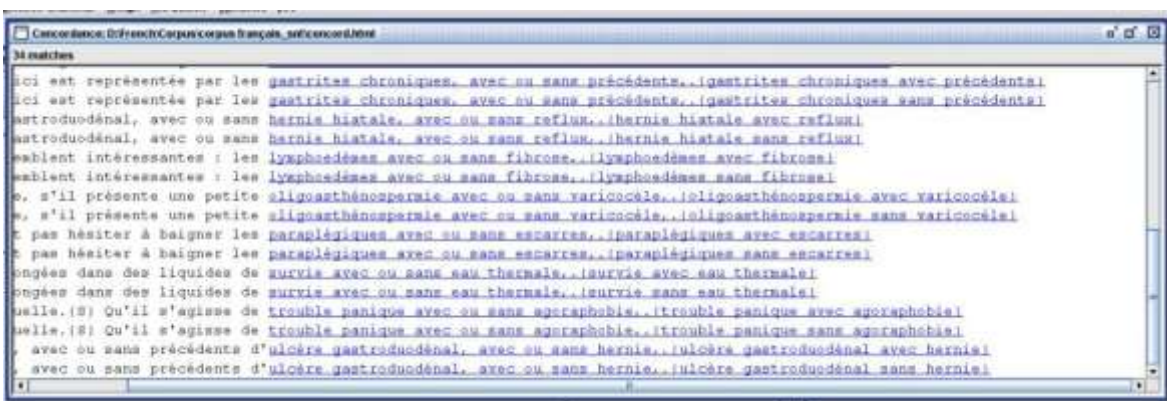


Figure 4 : affichage d'un extrait de la concordance produite par l'application du graphe de la figure 3.

Les exemples que nous venons d'illustrer nous sont utiles pour montrer une fois de plus que la réflexion sur la distribution des termes en discours peut améliorer nettement la description terminologique. Lorsqu'un logiciel exploite des méthodes symboliques, il est possible d'employer cette réflexion pour automatiser partiellement la tâche d'extraction terminologique.

Pour résumer

Dans ce dernier chapitre, nous avons suivi quelques pistes de réflexion offertes par les travaux du lexique-grammaire dans l'analyse d'unités terminologiques à base nominale.

Avant cela, nous avons rappelé les catégories sémantiques ressorties de l'analyse de quelques constructions verbales au chapitre précédent (§4.1.).

Ensuite, nous avons fourni quelques données à propos de la fréquence de quelques verbes des deux corpus et des noms qui leur sont morphologiquement reliés (§4.2.). Nous avons rapidement fait référence au traitement des noms prédicatifs dans le lexique-grammaire (§4.3.).

Suivant M. GROSS et TREMBLAY 1985, nous avons montré dans quelle mesure les faits de la syntaxe peuvent être utilisés pour montrer la continuité entre vocabulaire courant et vocabulaire spécialisé (§4.4.). Pour cela, nous avons d'abord mobilisé les notions de *figement* et de *degré de figement* (§4.5.). Nous avons ensuite montré quelques cas problématiques de découpage des unités terminologiques complexes affichant un moindre degré de figement (§4.6.). En premier lieu, nous avons analysé les noms composés qui ne se présentent pas dans le texte sous leur forme canonique (§4.6.1.). En deuxième lieu, nous avons passé en revue les cas de termes réduits en raison d'une pronominalisation partielle (§4.6.2.). Ces cas sont problématiques en vue d'une analyse automatique de textes (§4.6.3.). Un retour au logiciel Unitex a été nécessaire pour montrer qu'il est possible de semi-automatiser l'extraction terminologique des cas cités plus haut (§4.6.4.).