

Identification des adjectifs relationnels en corpus

Béatrice Daille

IRIN, 2, rue de la Houssinière,
BP 92208, 44322 NANTES cedex 3, France
<http://www.sciences.univ-nantes.fr/irin>
daille@irin.univ-nantes.fr

Résumé

Cet article présente l'identification en corpus des adjectifs relationnels considérés par les linguistes comme hautement dénominatifs. Notre approche utilise un programme d'extraction terminologique qui s'applique sur un corpus préalablement étiqueté et lemmatisé. Après avoir rappelé quelques propriétés linguistiques des adjectifs relationnels, nous présenterons le programme d'extraction de terminologie et les modifications apportées à celui-ci pour effectuer cette identification. Nous évaluerons le caractère dénominatif de ces adjectifs et des termes nominaux où ils apparaissent en les comparant à un thesaurus. Nous concluons sur l'intérêt de ces adjectifs à la fois pour l'extraction de terminologie mais aussi pour d'autres problématiques comme l'extraction de connaissances à partir de corpus ou la mise à jour d'un thesaurus.

1. Introduction

L'identification des adjectifs relationnels peut se révéler intéressante dans divers domaines du TALN comme par exemple l'acquisition de terminologie, la mise à jour de thesaurus, la détection des thèmes, du fait qu'ils possèdent une fonction dénominative forte communément admise chez les linguistes : (Bartning, 1976), (Mélis-Puchulu, 1991), (Lerat, 1995), (Levi, 1978), (Rey, 1979), etc.. Leur utilisation est donc privilégiée dans les domaines scientifiques (Monceaux, 1993).

Paradoxalement, les systèmes d'acquisition de terminologie TERMINO (David & Plante, 1990), LEXTER (Bourigault, 1994), TERMS (Justeson & Katz, 1995), pour ne citer que ces derniers, ne se sont pas intéressés aux adjectifs relationnels.

Dans le cadre de la modélisation de connaissances, (Assadi & Bourigault, 1995) proposent un module de classification des adjectifs extraits par LEXTER (Bourigault, 1994) pour aider à la validation terminologique. Le but est donc de regrouper les adjectifs apparaissant dans un même contexte, et plus précisément ceux qui partagent les mêmes noms de tête comme par exemple la classe composée des adjectifs *annuel*, *correspondant*, *total*, *réel* qui apparaît dans le contexte du nom *coût*. Notre approche, qui tend elle aussi à faciliter la tâche d'un expert, terminologue ou linguiste, est inverse et plus sélective puisque nous cherchons à regrouper les expansions par rapport au nom de tête, ainsi *produit laitier* sera associé à *produit au lait*, *produit de lait*, *produit*

issus du lait, etc.

Après un rappel de quelques propriétés linguistiques des adjectifs relationnels, nous décrirons le programme d'extraction de terminologie utilisé et les modifications effectuées de manière à identifier les adjectifs relationnels en corpus. Nous présenterons les résultats obtenus à partir d'un corpus technique de l'agriculture [AGRO] et comparerons les adjectifs relationnels identifiés ainsi que les candidats termes dans lesquels ils apparaissent avec un thésaurus du même domaine. Nous concluons sur l'intérêt qualitatif de la reconnaissance des adjectifs relationnels pour l'extraction de terminologie et d'autres problématiques du TAL.

2. Propriétés linguistiques de l'adjectif relationnel

Les adjectifs relationnels possèdent les propriétés linguistiques suivantes :

- relation morphologique avec un nom. Ces adjectifs dénominaux sont dérivés d'un nom au moyen d'un suffixe, citons les suffixes *-aire*, *-al*, *-el*, *-ien*, *-ique* pour les plus productifs. Ils peuvent être construits sur des bases non autonomes de classes de noms comme *card-* pour *cœur*, *mascul-* pour *homme*.
- la possibilité d'équivalence avec un complément prépositionnel d'un nom de tête au sein d'un syntagme nominal :
acidité sanguine, acidité du sang
conquête spatiale, conquête de l'espace
débit horaire, débit par heure
expérimentations animales, expérimentations sur les animaux
- et un certain nombre d'autres propriétés comme le fait d'être inusités comme attribut, l'incompatibilité avec le degré, la postposition immédiate après le nom dans une séquence d'adjectifs postposés, etc.

L'étude de (Monceaux, 1997) sur les adjectifs relationnels et leurs différentes propriétés remet en cause en partie certaines des propriétés citées ci-dessus :

1. Il existe une autre catégorie d'adjectifs qui se comportent comme des adjectifs relationnels : les adjectifs ayant un emploi substantival. Ces derniers partagent en particulier la propriété de la concurrence avec un groupe prépositionnel introduit par *de* : *un problème mathématique/un problème de mathématique*.
2. les propriétés syntaxiques de l'adjectif relationnel sont à nuancer. En particulier, il est démontré que l'adjectif relationnel, comme l'adjectif non dérivé à emploi substantival, peut accepter la construction attributive : *une pollution agricole/cette pollution est agricole, un problème mathématique/ce problème est mathématique*, le détachement, la modification et l'énumération : *les réformes politiques et économiques, une production surtout céréalière*, etc.

3. Identification des adjectifs relationnels

3.1. Programme d'extraction de terminologie

ACABIT, (Daille, 1996), le programme d'extraction de terminologie utilisé pour cette expérimentation a pour objectif de préparer la tâche du terminologue en lui proposant une liste

ordonnée de “candidats-termes” pour un corpus donné, c’est-à-dire des termes nominaux les plus représentatifs du domaine à ceux qui le sont moins. Ce programme utilise un score statistique pour classer les candidats. Les candidats-termes extraits du corpus correspondent à un type particulier de cooccurrences caractérisé par les propriétés suivantes :

- la cooccurrence est orientée et suit donc l’ordre linéaire du texte,
- elle est composée de deux unités lexicales pleines ;
- elle épouse l’une des structures morphosyntaxiques de ce que nous appelons “termes de base” ou l’une des variations recensées de ceux-ci.

Les termes de base s’appartiennent à l’une des structures morphosyntaxiques suivantes :

Nom1 Adj *emballage biodégradable*

Nom1 (Prep (Det)) Nom2 *ions calcium, protéine de poissons, chimioprophylaxie au rifampine*

Nom1 à Vinf *viandes à griller*

Ces structures de bases ne sont pas figées et admettent de nombreuses variations. Les variations prises en compte sont les suivantes :

1. Variations flexionnelles et variations syntaxiques faibles à l’intérieur de la structure de base :
 - les variations graphiques et orthographiques regroupent les variantes flexionnelles prévisibles : *conservation de produit, conservations de produit*, ou non : *conservation de produits* et les différences de casse.
 - la variation de préposition : *chromatographie en colonne, chromatographie sur colonne* ;
 - l’effacement ou non de la préposition et l’insertion ou non d’un déterminant : *fixation azote, fixation d’azote, fixation de l’azote*.
2. Variations de modification interne : l’insertion ou non d’un modifieur à l’intérieur de certaines structures de base comme par exemple l’insertion d’un adjectif au sein d’une structure **Nom1 Prep (Det) Nom2** : *lait de brebis, lait cru de brebis* ;
3. Variations de coordination : la coordination de termes de base : *alimentation humaine, alimentation animale et humaine* ;
4. Variations de l’adjectif en position attribut : *pectine méthylée, ces pectines sont méthylées*.

Le programme parcourt le corpus et relève les cooccurrences ainsi définies de termes ou de leurs variations en utilisant des grammaires locales à base d’expressions régulières. Les différentes occurrences sont regroupées sous l’entrée d’un couple.

La Figure 1 présente un exemple de candidat-terme : la première ligne indique quelle est la structure de base du candidat, un identificateur numérique du couple (13375), les deux lemmes qui le composent (*produit, surgeler*), sa fréquence (frq=4), la valeur du score statistique (stat=29.16)

| | | | | | | |
|--------------|-------|------------------|-------|------------|----------|--|
| na | 13375 | produit surgeler | frq=4 | stat=29.16 | VAR=1110 | ``produits surgelés`` |
| Flexion | | b007 | nbr=1 | | | text1=produits surgelés patt1=SBC ADJ2PAR lem1=produit surgeler nbo1=1 Pos=0372510-T-2/ |
| Modification | | i001 | nbr=2 | | | text1=produits alimentaires surgelés patt1=SBC ADJ ADJ2PAR lem1=produit alimentaire surgeler nbo1=1 Pos=0317818-T-5/ text2=produits halieutiques surgelés patt2=SBC ADJ ADJ2PAR lem2=produit halieutique surgeler nbo2=1 Pos=0421040-I-1/ |
| Coordination | | c001 | nbr=1 | | | text1=produits congelés ou surgelés patt1=SBC ADJ2PAR COO ADJ2PAR lem1=produit congeler ou surgeler nbo1=1 Pos=0421040-T-1/ |

FIG. 1 – Exemple de couple candidat-terme représentant une structure de base

et un résumé des variations rencontrées pour ce couple (VAR=1110). Le résumé des variations est une chaîne de quatre caractères, chaque caractère prenant la valeur 1 si le type de variation a été rencontrée, 0 sinon. Le premier caractère concerne les variations flexionnelles et morphosyntaxiques internes, le deuxième les variations de modification interne, le troisième les variations de coordination et le dernier, les variations attributives. Ensuite, pour chaque type de variation, les intitulés des règles utilisées (par exemple b007), le nombre d’occurrences ayant été reconnues par ces règles, les texte rencontrés, les patrons morphosyntaxiques et les formes lemmatisées correspondants, et leurs localisations dans le corpus (numéro de fichier, localisation dans le fichier : I pour titre, T pour texte, numéro de phrase du titre ou du texte). Le candidat-terme proposé à l’expert est par défaut la forme de base la plus fréquente apparaissant parmi les différentes occurrences recensées, *produits surgelés* pour l’exemple de la Figure 1.

3.2. Modification apportée au programme d’extraction de terminologie

Pour identifier les adjectifs relationnels, nous allons utiliser leur propriété d’équivalence avec un groupe prépositionnel. Nous allons donc regrouper les termes de base de structure Nom1 Prep (Det) Nom2 avec les termes de base Nom1 Adj où l’adjectif est soit dérivé de Nom2 comme par exemple *production de céréales*, *production céréalière*, soit possède un emploi nominal comme par exemple *muscle bovin*, *muscles de bovin*.

La reconnaissance des adjectifs relationnels s’exécute après l’extraction des différentes occurrences des termes et de leur variations syntaxiques. Chaque candidat étant associé à une structure de base et accompagné d’un résumé des variations rencontrées, il aurait été possible de vérifier la non modification, la non position attributive de l’adjectif si (Monceaux, 1997) n’avait pas dénoncé la généralité de ces propriétés.

Chaque candidat de structure Nom Adj va être examiné en fonction de la terminaison de son adjectif, les adjectifs relationnels possédant des terminaisons prédictibles. Les suffixes les plus courants du français entrant en jeu dans la dérivation des adjectifs relationnels ont été inventoriés dans (Guyon, 1993) ; il s’agit des suffixes favorables : *ain*, *aire*, *al*, *el*, *estre*, *ien*, *ier*, *il*, *in*, *ique* et des suffixes défavorables : *esque*, *eux*, *é*, *if*, *oire*.

Parmi les adjectifs relationnels, il existe deux sous-classes qui ne seront pas considérées : les adjectifs dérivés de noms propres et les adjectifs ethniques ; ceci principalement à cause du manque de noms propres dans les corpus techniques et aussi de la difficile prédictibilité de la forme adjectivale dérivée.

L’algorithme ci-dessous isole les candidats Nom Adj où l’adjectif se termine par *-ique* et décide

```

3076  BASE=0101 nbcand=16 stat=32.36
      npn 1693 fibre aliment frq=1 VAR= 1000 ``fibres dans les aliments``
      Flexion
          b014  nbr= 1
              text1= fibres dans les aliments patt1= SBC PREP DTN SBC
              lem1= fibre dans le aliment nbo1= 1 Pos=0066527-T-1/
      nar 1826 fibre alimentaire frq= 15 VAR= 1000 ``fibres alimentaires``
      Flexion
          b007  nbr= 15
              text1= fibres alimentaires patt1= SBC ADJ
              lem1= fibre alimentaire nbo1= 14
              Pos=0573344-I-1/0573344-T-1/0599904-T-1/0197642-I-1/etc.
              text2= Fibres alimentaires patt2= SBC ADJ
              lem2= fibre alimentaire nbo2= 1 Pos=0249374-T-1/
    
```

FIG. 2 – Exemple de couple candidat-terme acceptant une variante dérivationnelle

si Adj est relationnel ou non. Un algorithme similaire pour chaque suffixe a été écrit.

1. Examen de chaque candidat de structure Nom Adj ;
2. Vérification que le lemme de l’adjectif se termine par *-ique* ;
3. Application d’une règle de désuffixation/recodage de manière à générer toutes les formes du nom dérivé possibles :

| Règles | Exemples |
|-----------------------|--|
| <i>-élique/+el</i> | <i>pantagruélique/pantagruel</i> |
| <i>-énique/+ène</i> | <i>hygiénique /hygiène</i> |
| <i>-érique/+er(e)</i> | <i>éthérique / ether - atmosphérique /atmosphère</i> |
| <i>-étique/+et</i> | <i>alphabétique/alphabet</i> |
| <i>-cique/+x</i> | <i>thoracique/thorax</i> |
| <i>-ique/+(e)</i> | <i>ionique/ion - bronchique/bronche</i> |
| <i>-ique/+ie</i> | <i>chromatographique/chromatographie</i> |
| <i>-ique/+isme</i> | <i>catabolique/catabolisme</i> |
| <i>-ique/+icité</i> | <i>élastique/élasticité</i> |

Nous avons utilisé la méthode de surgénération décrite dans (Jacquemin & Tzoukermann, 1997). Comme l’existence des formes nominales générées à partir des règles est vérifiée dans le corpus, la production de monstres est grandement évitée : ainsi, avec l’adjectif *ionique*, les formes *ionie* et *ion* sont générées, mais seule la forme *ion* est attestée dans le corpus. Ces règles ont été établies manuellement à l’aide des exemples d’adjectifs relationnels et de leurs noms associés dans (Guyon, 1993), (Monceaux, 1993), et d’autres lectures. Nous ne prétendons pas couvrir toutes les formes nominales dérivationnelles possibles mais seulement les plus fréquentes.

4. Recherche dans l’ensemble des candidats de structure d’un Nom1 Prep (Det) Nom2 du candidat comportant le Nom1 (identique dans N1 Adj-*ique*) et le Nom2 généré à l’étape précédente.
5. Si l’étape 4 réussit, réunion des différentes structures de base. Extraction du candidat possédant un adjectif relationnel de l’ensemble des candidats de structure Nom Adj et affectation d’une structure Nom AdjR. Mise à jour des scores statistiques.

Pour les adjectifs non éthiques ayant un emploi substantival comme par exemple *problème technique* et *problème de technique*, il suffit de permettre qu’un couple candidat puisse posséder

deux structures de base : l'une de type Nom1 Prep (Det) Nom2, et l'autre de type Nom1 Adj. Il n'y aura aucun calcul sur la forme lemmatisée du Nom2, celle-ci étant identique à celle de l'adjectif.

Un exemple de regroupement comportant un adjectif relationnel est présenté dans la Figure 2. La première ligne identifie le candidat à l'aide d'un identificateur numérique unique (3076), d'un résumé des différentes structures de base rencontrées (BASE=0101), de sa fréquence (nb-cand=16) et de son score statistique (stat=32.36). Le résumé des structures de base (BASE=0101) est une chaîne de quatre caractères, chaque caractère prenant la valeur 1 si le type de base a été rencontrée, 0 sinon. Le premier caractère concerne la structure Nom Adj, le deuxième Nom1 (Prep (Det)) Nom2, la troisième Nom Prep Vinf et le dernier la structure Nom AdjR. Les autres lignes reprennent les informations fournies précédemment (cf. Fig 1 et section 3.1).

4. Résultats et Évaluation sur [AGR]

Notre corpus est constitués de 7272 résumés de texte français (2,5 Mbytes) extraits de la base PASCAL¹ sur le domaine de l'agriculture (corpus désigné par [AGRO]). Nous avons utilisé l'étiqueteur de Brill (Brill, 1992) entraîné pour le français à l'INaLF (Lecomte & Paroubek, 1996) et le lemmatiseur développée par F. Namer dans le cadre du projet ILIAD (Toussaint *et al.*, 1998). Cette chaîne de traitements satisfait nos besoins sauf pour le traitement des accords au sein d'une phrase ou même au sein des différents constituants de la phrase. Il est en effet impossible de vérifier les accords entre le nom et l'adjectif, ce qui conduit à l'extraction erronée de candidats termes de base ou de leurs variantes. Les informations morphosyntaxiques associées aux mots du texte permettent l'identification des structures de base et la description de leurs variations syntaxiques. La lemmatisation permet de regrouper les différentes flexions et variations. Par contre, aucune information dérivationnelle n'est fournie et il n'existe pas à notre connaissance d'étiqueteur dérivationnel du français, ni d'ailleurs de base lexicale fournissant les familles dérivationnelles des mots, comme par exemple CELEX pour l'anglais². Sans information dérivationnelle, nous ne pourrions pas prendre en compte les adjectifs relationnels construits sur des bases non autonomes de classe de noms (*card/coeur*), ni des racinaux supplétifs par rapport au nom (*pater/père, urb/ville*). Néanmoins, une approche par racinisation pour l'identification de variantes dérivationnelles de termes en indexation donne de bons résultats, même si ceux-ci sont moins bons qu'avec une base lexicale (Daille & Jacquemin, 1998).

4.1. Résultats quantitatifs

La Figure 3 présente le nombre de structures de base et le nombre de variations dérivationnelles identifiées. Ces structures Nom AdjR ont été extraites de l'ensemble des structures de type Nom Adj. Ces variations sont peu nombreuses, moins de 1 %, mais elles permettent d'identifier 133 hapax hautement dénominatifs.

La Figure 4 indique le nombre de regroupements dans lesquels apparaît un adjectif relationnel. Les adjectifs relationnels dérivés sont au nombre de 89 : *agronomique, alimentaire, arachidier, aromatique, ...* ; les adjectifs à emploi nominal au nombre de 28 : *acide, aminopeptidase, animal, antioxydant,*

1. PASCAL est la base documentaire scientifique développées et maintenue par l'INIST-CNRS, France

2. Une base lexicale définissant les familles dérivationnelles du français est en cours de constitution au sein du projet FRANLEX (<http://www.limsi.fr/Individu/jacquemi/Franlex>) (Namer, 1999).

| Nombre d'occurrences des structures de base | 1 | ≥ 2 | Total |
|---|--------|--------|--------|
| Nom1 (Prep (Det)) Nom2 | 17 232 | 5 949 | 23 181 |
| Nom Adj | 12 344 | 4 778 | 17 122 |
| Nom à Vinf | 203 | 16 | 219 |
| Nom AdjR | 133 | 149 | 282 |
| Total | 29 912 | 10 895 | 40 807 |

FIG. 3 – Données quantitatives des structures de base

| Candidat incluant un adjectif relationnel | Nombre |
|---|------------|
| Nom1 (Prep (Det)) Nom2 / Nom1 AdjR avec AdjR dérivé de Nom2 | 282 |
| Nom1 (Prep (Det)) Nom2 / Nom1 Adj2 with lemme(Nom2) = lemme(Adj2) | 53 |
| Total | 335 |

FIG. 4 – Nombre de regroupements incluant un AdjR

4.2. Résultats qualitatifs

Nous avons vérifié la correction linguistique des 282 variations structurelles qui permettent de regrouper une structure Nom1 Prep (Det) Nom2 et une structure Nom1 AdjR. Les erreurs sont les suivantes :

- Génération de deux monstres : *contrôle important lié à contrôles importants à l'importation et évolution normale lié à évolution des normes*;
- Deux mauvais regroupements dus à la non-reconnaissance par l'étiqueteur de la préposition composée *en fin de* : *populations finales* lié à *population atteinte en fin* and *eau finale* lié à *eau atteintes en fin*.

Nous obtenons donc une précision de correction linguistique de 95 % dans l'identification des adjectifs relationnels dérivés.

Les 52 adjectifs à emploi nominal sont tous corrects.

5. Évaluation du caractère dénominatif des adjectifs relationnels à l'aide d'AGROVOC

Le thesaurus AGROVOC³ est un taximonomie d'environ 15 000 termes associés à leurs éventuels synonymes, ce qui nous donne 25 964 termes.

À l'aide d'AGROVOC, nous avons effectué trois évaluations : en premier, nous avons vérifié l'usage des adjectifs relationnels au sein des termes du thesaurus, en second, nous avons comparé les candidats termes extraits possédant à la fois l'adjectif relationnel et sa forme concurrente en groupe prépositionnel aux termes du thesaurus, en dernier, nous avons vérifié pour 3 adjectifs relationnels dérivés et 2 adjectifs relationnels à emploi nominal apparaissant dans un candidat terme non mis en relation avec un groupe prépositionnel s'ils existaient dans le thesaurus.

5.1. Adjectif relationnel seul

Parmi les 89 adjectifs dérivés extraits de [AGR], 45 apparaissent dans des termes du thesaurus comme par exemple *enzymatique*, *fruitier*, *gazeux* et 38 uniquement sous leur forme

3. Un thesaurus multilingue développé par AGRIS.

| Type d'adjectifs relationnels | Correspondance exacte | | Correspondance partielle | | Total |
|-------------------------------|-----------------------|--------------------|--------------------------|-----------------|------------------|
| | N1 AdjR | N1 (Prep (Det)) N2 | N1 seul | Terme plus long | |
| Adjectif dérivé | 29 | 14 | 73 | 2 | 118/282 (42 %) |
| Adjectif avec emploi nominal | 6 | 0 | 15 | 0 | 21/53 (40 %) |
| Total | 35 (10 %) | 14 (4 %) | 88 (28 %) | 2 (0,6 %) | 139/335 (41,5 %) |

FIG. 5 – Comparaison des candidats termes en concurrence avec un groupe prépositionnel et AGROVOC

nominale comme par exemple *arachide* au lieu d'*arachidier*. Les adjectifs absents du thesaurus sont : *bibliographique*, *compartimental*, *miscellaire*, *neuronal*, *photonique*, *transmembranaire*. 93 % des adjectifs dérivés extraits peuvent donc être considérés comme hautement dénominatifs.

Des 28 adjectifs à emploi nominal, 22 apparaissent dans des termes du thesaurus sous leur forme adjectivale comme par exemple *continu*, *courant*, *dynamique*, et 5 seulement sous leur forme nominale comme par exemple *antioxydant*. Le seul adjectif absent du thesaurus est *cinétique*. 96 % des adjectif à emploi nominal extraits peuvent donc être considérés comme hautement dénominatifs.

Ce premier résultat corrobore l'intuition des linguistes sur le caractère dénominatif des adjectifs relationnels et l'affirmation de (Monceaux, 1997) que l'adjectif à emploi nominal peut être considéré comme un adjectif relationnel.

5.2. Adjectif relationnel au sein d'un candidat terme mis en relation avec un groupe prépositionnel

Nous avons comparé les adjectifs relationnels et leur structure concurrente avec un groupe prépositionnel avec les termes du thesaurus. Nous avons distingué deux cas : correspondance totale ou partielle. Les correspondances partielles que nous avons acceptées sont : soit le nom modifié apparaît seul dans le thesaurus comme par exemple le nom *acide* extrait du candidat *acide vinique*, soit le candidat terme appartient à un terme du thesaurus de longueur supérieure comme par exemple *ingrédients alimentaires* pour *ingrédients de produits alimentaires*. Les résultats sont présentés dans la Figure 5.

L'un des premières constatations est que le thesaurus préfère de loin les structures Nom AdjR aux structures Nom1 Prep (Det) Nom2. Néanmoins, il n'existe pas réellement de cohérence dans les choix du terme de référence quand les deux formes existent ; ainsi, on trouve *métabolisme protéique* mais pas *métabolisme glucidique*, le terme *métabolisme des glucides* étant préféré à ce dernier. Nos résultats pourraient être utilisés pour la mise à jour de thesaurus lorsque la forme Nom AdjR a été effectivement rencontrée dans un corpus. De plus, (Monceaux, 1997) précise que la forme Prep (Det) Nom2 apparentée à l'adjectif relationnel est difficilement prédictible. La présente conjointe des formes Nom AdjR et Nom1 Prep (Det) Nom2 dans un thesaurus permettrait de lever l'ambiguïté inhérente de certains de ces adjectifs relationnels.

Les structures Nom AdjR sont peu soumises aux variations syntaxiques à la différence de leurs homologues Nom1 Prep (Det) Nom2 : seules 3 % des structures Nom AdjR admettent des variations et ce sont toutes des coordinations comme par exemple : *produit alimentaire/produits agricoles et alimentaires/produit en alimentation*. Aucune variation attributive ou avec insertion de modifieur(s) n'a été rencontrée.

À l'inverse, 60 % des structures Nom1 Prep (Det) Nom2 regroupées à une structure Nom AdjR admettent de variations et principalement des modifications comme par exemple : *filtration*

| Adjectif relationnel | Correspondance totale | | Correspondance partielle | | Total |
|----------------------|-----------------------|-----------------------------|--------------------------|-----------------|-----------------|
| | Nom AdjR | Nom Prep (Det) Nom2 | Nom1 seul | terme plus long | |
| <i>animal</i> | 5/14 | 0/1 | 5/14 + 0/1 | 1/1 | 11/15 (73 %) |
| <i>bactérien</i> | 5/43 | 1/82 (sous la forme N AdjR) | 9/43 + 32/82 | 1/43 | 48/125 (38 %) |
| <i>bovin</i> | 1/18 | 0/1 | 10/18 + 0/1 | 0 | 11/19 (58 %) |
| <i>économique</i> | 16/66 | 1/7 (sous la forme N AdjR) | 11/66 + 0/7 | 0 | 28/73 (38 %) |
| <i>ionique</i> | 6/15 | 0/29 | 5/15 + 10/29 | 1/15 | 22/44 (50 %) |
| Total | 33/156 (21 %) | 2/120 (2 %) | 82/276 (29 %) | 3/276 (1 %) | 120/276(43,5 %) |

FIG. 6 – Comparaison de candidats termes sans concurrence avec un groupe prépositionnel et AGROVOC

membranaire/filtration frontale sur membrane. Certains regroupements Nom1 Prep (Det) Nom2 et Nom2 AdjR sont particulièrement intéressants : ceux où la forme Nom AdjR a été identifiée sans que la forme de base Nom1 Prep (Det) Nom2 existe. Par exemple, *acide vinique* est une forme attestée mais seules des variations syntaxiques impliquant une modification d’*acide* Prep (Det) *vin* existent comme : *acides organiques du vin*, *acide malique dans le vin*, *acides aminés des vins*, *acide tartrique dans les vins*, *acide salicylique dans les vins*. L’*acide vinique* regrouperait donc les *acides organiques*, *maliques*, *aminés*, *tartriques* et *salicyliques*, tous présents dans le thesaurus. Ces regroupements constituent une première étape à l’acquisition de connaissances à partir de textes et pourraient, par exemple, servir d’amorce à un système d’acquisition de relations sémantiques comme PROMETHEE (Morin, 1998).

5.3. Adjectif relationnel au sein d’un candidat terme sans mise en relation avec un groupe prépositionnel

Nous avons extraits les adjectifs relationnels de notre corpus grâce à leur emploi concurrent avec un groupe prépositionnel et nous avons évalué leur caractère dénominatif. Il nous a paru intéressant d’évaluer le nombre de candidats où l’adjectif relationnel n’est plus en concurrence avec un groupe prépositionnel effectivement présents dans le thesaurus. À titre indicatif, nous avons examiné trois adjectifs dérivés et 2 adjectifs à emploi nominal. Les résultats sont présentés en Figure 6. Nous remarquons que ces adjectifs relationnels non en concurrence avec un groupe prépositionnel sont encore plus dénominatifs que ceux où les deux formes sont attestées en corpus. Une interprétation pourrait être que lorsque les deux formes apparaissent conjointement dans un corpus, le processus de lexicalisation est en cours alors qu’il est terminé dans l’autre cas. En conclusion, il est nécessaire d’extraire de l’ensemble des candidats de structure Nom Adj, tous les candidats admettant un adjectif relationnel et pas uniquement ceux en concurrence avec un groupe prépositionnel.

6. Conclusion

Nous avons présenté une méthode d’identification des adjectifs relationnels en corpus grâce à la présence conjointe d’une forme Nom1 Prep (Prep) Nom2 et Nom Adj où l’adjectif est soit dérivé de Nom2, soit partage une même forme avec Nom2. Cette expérimentation corrobore les résultats des études linguistiques et prouvent que les adjectifs relationnels portent une fonction dénominatrice forte. L’identification de ce type d’adjectifs est donc importante dans un programme d’extraction de terminologie, mais aussi particulièrement intéressante dans la problématique de la mise à jour de thesaurus ou de la veille scientifique où l’apparition d’un adjectif relationnel semble caractériser une stabilisation d’un concept scientifique émergent.

Références

- ASSADI H. & BOURIGAULT D. (1995). Classification d'adjectifs extraits d'un corpus pour l'aide à la modélisation de connaissances. In *Troisième Journées Internationales d'Analyse Statistique de Données Textuelles*, p. 313–320, Rome.
- BARTNING I. (1976). *Remarque sur la syntaxe et la sémantique des pseudo-adjectifs dénominaux en français*. PhD thesis, Université de Stockolm.
- BOURIGAULT D. (1994). *LEXTER, un Logiciel d'Extraction de TERminologie. Application à l'acquisition des connaissances à partir de textes*. PhD thesis, EHESS, Paris.
- BRILL E. (1992). A simple rule-based part of speech tagger. In *Third Conference on Applied Natural Language Processing (ANLP'92)*, p. 152–155, Trento.
- DAILLE B. (1996). Study and implementation of combines techniques for automatic extraction of terminology. In J. L. KLAVANS & P. RESNIK, Eds., *The Balancing Act - Combining Symbolic and Statistical Approaches to Language*, chapter 3, p. 28–49. MIT Press.
- DAILLE B. & JACQUEMIN C. (1998). Lexical database and information access: A fruitful association. In *LREC'98*, p. 669–673. Grenade, Espagne: ELRA.
- DAVID S. & PLANTE P. (1990). Le progiciel termino : De la nécessité d'une analyse morphosyntaxique pour le dépouillement terminologique des textes. In *ICO*, volume 2.
- GUYON A. (1993). *Les adjectifs relationnels arguments de noms prédicatifs*. PhD thesis, Université Paris 7.
- JACQUEMIN C. & TZOUKERMANN E. (1997). Npl for term variant extraction: Synergy between morphology, lexicon and syntax. In T.STRZALKOWSKI, Ed., *Natural Language Processing and Information Retrieval*. Boston, MA: Kluwer.
- JUSTESON J. & KATZ S. (1995). Technical terminology: Some linguistic properties and an algorithm for identification in text. In *Journal of Linguistic Engineering*, volume 1.
- LECOMTE J. & PAROUBEK P. (1996). *Le catégoriseur d'Eric Brill. Mise en œuvre de la version entraînée à l'INALF*. Rapport interne, CNRS-INALF.
- LERAT P. (1995). *Les langues spécialisées*. Paris: PUF.
- LEVI J. (1978). *The syntax and the semantics of complex nominals*. London: Academic Press.
- MÉLIS-PUCHULU A. (1991). Les adjectifs dénominaux : des adjectifs de "relation". *Lexique*, **10**, 33–60.
- MONCEAUX A. (1993). *La formation des noms composés de structure NOM ADJECTIF*. Thèse de doctorat en linguistique théorique et formelle, Université de Marne la Vallée.
- MONCEAUX A. (1997). Adjectif de relation, complémentation et sous-classification. *Langages*, **126**, 39–59.
- MORIN E. (1998). Promethee : un outil d'aide à l'acquisition de relations sémantiques entre termes. In *TALN'98*, p. 172–181, Paris.
- NAMER F. (1999). Le traitement automatique des mots dérivés : le cas des noms et adjectifs en *et(te)*. In *Colloque "La morphologie des dérivés évaluatifs"*, Toulouse.
- REY A. (1979). *La terminologie : noms et notions, Que sais-je ?* Paris: PUF.
- TOUSSAINT Y., NAMER F., DAILLE B., JACQUEMIN C., ROYAUTÉ J. & HATHOUT N. (1998). Une approche linguistique et statistique pour l'analyse de l'information en corpus. In *TALN'98*, p. 182–191. Paris.