Qualitative terminology extraction: Identifying relational adjectives

Béatrice Daille

Qualitative terminology extraction:

Identifying relational adjectives

Béatrice Daille
IRIN

This paper presents the identification in corpora of French relational adjectives, phenomena considered by linguists as highly informative. The approach uses a termer which is applied on a tagged and lemmatized corpus. Relational adjectives and nominal compounds which include a relational adjective are then quantified and their informative status is evaluated thanks to a thesaurus of the domain. We conclude with a discussion of the interesting status of such adjectives and nominal compounds for terminology extraction and other automatic terminology tasks.

1. Introduction

Identifying relational adjectives such as *malarial*, and nominal phrases in which they appear such as *malarial mosquitoes*, could be interesting in several fields of NLP, such as terminology acquisition, topic detection, updating of thesauri, because they hold a naming function acknowledged by linguists (Bartning 1976), (Levi 1978), (Mélis-Puchulu 1991), etc. The use of relational adjectives is so favored in scientific fields (Monceaux 1993).

Paradoxically, terminology acquisition systems such as TERMINO (David and Plante 1990), LEXTER (Bourigault 1994), TERMS (Justeson and Katz 1995), have not been concerned with relational adjectives.

But, even if the relational adjective hods an heavy naming function, it remains ambiguous in the same way than an uniterm because of its possible migration from a scientific domain to another: the adjective *planetary* employed with the noun *right* belongs to legal domain, with the noun *system* to astronomy, with the noun *electron* to chemistry, etc. This ambiguity is raised only when the noun which goes along the adjective is identified. A nominal phrase including a relational adjective is interpreted by (Lerat 1995) as:

> "(…) a way to condense information under a justifiable form rather than an explicit one."

The relational adjective, even inside a nominal phrase, remains ambiguous as long as the prepositional phrase with which it could be paraphrased is not clearly enounced: does the nominal phrase *animal flour* mean *flour for animals or flour made with animals?*

Our concern is:

1) To identify nominal phrases in which appear relational adjectives, as well as the prepositional phrases by which they could be paraphrased. We will see by another source presented in section 2 that this property of paraphrase is used to identify these adjectives.

To check the naming character of these adjective and to evaluate the naming character of the nominal phrase in which they appear.

Identifying both the adjective and the prepositional phrase could be useful in several type of NLP applications and allow us:

- In the field of terminology acquisition, to group synonym forms referring to an unique concept such *produit laitier (dairy product)* and, *produit au lait (product with milk), produit de lait (product of milk), produit issu du lait (product made of milk),* etc. Assadi and Bourigault (1995) have proposed a clustering model for the grouping of adjectives inside nominal phrases extracted by LEXTER (Bourigault 1994) in order to help terminology validation. Contrarily to our approach, their objective was to group

adjectives appearing in the same context, and more precisely those which were sharing the same head nouns such as the class composed of the adjectives *annuel (yearly), correspondant (corresponding), total (total), réel (real)* that appears with the noun *coût (cost)*;

- In the field of thesaurus updating to replace the nominal phrase with a prepositional phrase by the one with a relational adjective, this later being more characteristic of a scientific denomination;
- In the field of monolingual or bilingual dictionary, to disambiguate the meaning of the relational adjective thanks to its prepositional form.

To purchase this identification, we use shallow parsing (Abney 1991), then for morphological processing, a dynamic method which takes as input a corpus labeled with part-of-speech and lemma tags. This method does not use any lexical database, and, in particular, no derivational information is provided. This, because it does not exist for French a derivational morphology tagger, nor a lexical database such CELEX for English[i] where to each lemma is attached its derivational morphological structure. Without derivational information, nor relational adjectives built from non autonomous bases of noun classes such as *cœur/card* (*heart/card)*, nor from Latin noun bases such as *père/pater (father/pater), ville/urb (town/urb)* will be identified. However, a dynamic approach for identifying derivational links of terms by stemming gives good results for document retrieval (Daille et Jacquemin 1998), even if these results are not as good as those obtained with a lexical database.

In this study, we first mention the definition and some linguistic properties of relational adjectives (AdjR). We then present the termer and the modifications that we made in order to allow the identification of AdjR in texts. We quantify the results obtained from a technical corpus of the field of agriculture [AGRO] and compare the AdjR and the nominal compounds in which they appear with the terms of a thesaurus on the same field. We conclude by the effective informative character of such adjectives for terminology extraction and also other NLP fields.

2. Definition and Linguistic properties of relational adjectives

According to linguistic and grammatical tradition, there are two main categories among adjectives: epithetic such as *important (significant)* and relational adjectives such as *laitier (dairy)*. The first ones cannot have an agent interpretation contrarily to the second ones: the adjective *laitier (dairy)* within the nominal phrase *production laitière (dairy production)* is an argument to the valent noun *production (production)* and this is not the case for the adjective *important (significant)* within the phrase *production importante (significant production).* The term of "adjective of relation" or "relational adjective" have been introduced by (Bally 1965) and enables it to convey this idea of "relation" usually expressed by a preposition. These adjectives are so-called "pseudo-adjective" by Government-Binding and Transformational linguistic schools (Postal 1969), (Zribi-Hertz 1972), (Bartning 1976). Epithetic adjectives (AdjE) and relational adjectives (AdjR) share the properties of agreement in number and gender with the noun with which they come along and the possibility to be in an attributive position. But, these two categories differ from each other in regards to morphological, paraphrastic, syntactic and semantic properties which applied either to the single adjective or to the nominal phrase within it appears.

3. Morphological properties

Relational adjectives are either denominal adjectives ---morphologically derived from a noun thanks to suffix---, or adjectives having a nominal usage such as mathématique (mathematical/mathematics). But, for the first ones, all the adjective-forming suffixes do not lead to relational adjectives. The following suffixes are considered by (Dubois 1962) as appropriate: -ain, -aire, -al, -el, -estre, -ien, -ier, -il(e), -in, -ique. These favourable suffixes attach themselves either to a noun : *cellule (cell)* → *cellulaire (cellular)*, or a scientist nominal base: *hôpital (hostipal* (noun)) → *hospitalier (hospital* (adjective)).
However, (Guyon 1993) remarks that a suffix even the most appropriate is never necessary nor sufficient. Several adjectives wearing a favourable suffix are not relational: it is the case of the adjectives ending with *–ique*, which characterize chemistry field and which are not derived from a noun, such as

désoxyribonucléique (desoryribonucleic), dodecanoique (dodecanoic), etc. Other suffixes inappropriate are sometimes used such as the suffixes *–é* and *-eux*:
carbone (*carbon*) → *carboné* (*carbonaceous*), *cancer* (*cancer*) → *cancéreux* (*cancerous*), etc.

3.1. *Paraphrastic properties*

A relational adjective is usually paraphrasable by a prepositional phrase. The preposition employed, as well as the presence or not of a determiner, depends on the head noun of the nominal phrase:
acidité sanguine (blood acidity) ≈ acidité du sang (acidity of the blood)
conquête spatiale (space conquest) ≈ conquête de l'espace (conquest of space),
débit horaire (hour rate) ≈ débit par heure (rate per hour),
expérimentations animales (*animal experimentation*) ≈ *expérimentations sur les animaux* (*experimentation with animals*).

3.2. *Syntactic properties*

Relational adjectives are subject to syntactic constraints mentioned in (Monceaux 1997) on the contrary to epithetic adjectives. These properties apply to the nominal phrase formed by the noun and the relational adjective:

- Non-predicativity, saying the ban of predicative position, apart from in specific conditions forcing a typical interpretation:
  AdjE: *cette production est importante (this production is significant)*
  AdjR: ?*cette production est laitière* (?*this production is dairy*) (strictly)
  cette production est laitière (*this production is dairy*) (typically)
- The incompatibility with a degree adverbial modification, including in predicative sentences with a contrastive interpretation:
  AdjE: *une production très importante (a very significant production)*
  AdjR :**une production très laitière* (**a very dairy production*)
  cette production est très laitière* (this production is very dairy*)
- The non-fronting position, saying the ban to encounter relation adjectives in a prenominal attributive position:
  AdjE : *une importante production*

AdjR : **une laitière production*

Other properties are also mentionned such as the non-coordination, neither between relational adjectives nor with epithetic adjectives (Guyon, 1993), the direct contact with the head noun in a sequence of postnominal adjectives (Melis-Puchulu 1991), etc.
These syntactic properties should be carefully handle. Indeed, the relational adjective when it reflects a "typological" or a "typical" property, accepts the predicative position: *une pollution agricole (an agricultural pollution)/cette pollution est agricole (this pollution is agricultural)*, *un problème mathématique (a mathematical problem)/ce problème est mathématique (this problem is mathematical)*, detachment, modification and enumeration*: les réformes politiques et économiques* (*the political and economical reforms*), *une production surtout céréalière* (*a mainly cereal production*), etc.

3.3. *Semantic properties*

By definition (Bally 1965), relational adjectives express a link of "relation" although the epithetic adjective a link of "inherence". For Kleiber (1990), the relational adjective contains a "categorematic" notion --- which builds its own reference--- contrary to the epithetic adjective which contains a "syncategorematic" notion ---which is referencally dependent---. For (Tamba-Mecz 1980), the relational adjective reveals a "external-global" vision although the epithetic adjective reveals a "internal-partial" one. In short, relational adjectives point notions with a defined reference out and give cause for putting in relation two independent notions inside the nominal phrase: one carried by the head noun and the other contained in the relational adjective.

4. Identifying relational adjectives

Among all the linguistic criteria that we have presented, few are operating on automatic identification: this is the case of semantic criteria, but also of negative syntactic criteria. Indeed, the fact that an predicative construction between the noun and the adjective is not encountered in the corpus, does not allow to deduce

that the construction is impossible. The presence of a forbidden construction could allows us to refuse the relational status to an adjective. But, these constructions being possible in some semantic interpretations of the adjective, we prefer, in a first time, not to reject adjectives which accept them. We study in section 5.3 the syntactic constructions encountered and see how they could be exploited. The morphological criterion is not sufficient and uses it alone will induce noise (see section 2.1) and might alter our study.

Syntactic and semantic properties being turned down, morphological property being insufficient alone, we use their paraphrastic property which includes the morphological property. This paraphrastic criterion is a strong one, which will produce few noise, but which will not afford to identify exhaustively relational adjectives on account of:

- The absence of paraphrases in the corpus;
- A non-paraphrasability or a complex paraphrasability;
- The large derivational distance between the adjective and the noun, in particular for all adjectives built from non autonomous bases.

We will see in section 5.3 how acquiring still with reliable criteria other relational adjectives.

Identifying Noun Adj sequences characterized by the paraphrastic criterion requires a program able to extract all nominal phrases from a corpus. First, we present the termer that we choose, then , the modifications that we made in order to be able to identify relational adjectives.

4.1. *Terminology Extraction Program*

ACABIT (Daille 96), the termer used for this experiment, cases the task of the terminologist by proposing, for a given corpus, a list of candidate terms ranked from the most representative of the domain to the least by using a statistical score. Candidate terms which are extracted from the corpus belong to a special type of cooccurrences:

- the cooccurrence is oriented and follows the linear order of the text;
- it is composed of two lexical units which do not belong to the class of functional words such as prepositions, articles, etc.;

- it matches one of the morphosyntactic patterns of what we will call ``base terms'', or one of their possible variations.

The patterns for base terms are:

[Noun1 Adj] *emballage biodégradable (biodegradable package)*

[Noun1 (Prep (Det)) Noun2] *ions calcium (calcium ion)*, *protéine de poissons (fish protein)*, *chimioprophylaxie au rifampine (rifampicin chemoprophylaxis)*

[Noun1 à Vinf] *viandes à griller (grill meat)*

These base structures are not frozen structures and do accept several variations. Those which are taken into account are:

1. Flexional and Internal morphosyntactic variants:
   - graphic and orthographic variants which gather together predictable flexional variants: *conservation de produit* (*product preservation), conservations de produit (product preservations)*, or not: *conservation de produits (products preservation)* and case differences.
   - variations of the preposition: *chromatographie en colonne (chromatography in spine)*, *chromatographie sur colonne (chromatography on spine)*;
   - optional character of the preposition and of the article: *fixation azote (nitrogen fixation)*, *fixation d'azote (fixation of nitrogen)*, *fixation de l'azote (fixation of the nitrogen)*;
2. Internal modification variants: insertion inside the base-term structure of a modifier such as the adjective inside the Noun1 (Prep (Det)) Noun2 structure: *lait de chèvre (goat's milk)*, *lait cru de chèvre (milk straight from the goat)*;
3. Coordinational variants: coordination of base term structures: *alimentation humaine (human diet)*, *alimentation animale et humaine(human and animal diet)*;
4. Predicative variants: the predicative role of the adjective: *pectine méthylée (methylate pectin)*, *ces pectines sont méthylées(these pectins are metylate)*.

The corpus is tagged and lemmatized. The program scans the corpus, counts and extracts collocations whose syntax characterizes base-terms or one of their variants. This is done with shallow parsing using local grammars based on regular expressions (Silberztein 1994; Basili et al. 1993). These grammars use the morphosyntactic informations associated to the words of the corpus by the tagger. The different occurrences are grouped as a pair formed by the two lemmas of the candidate term.

Figure 1 demonstrates the output for a candidate term: the first line indicates the base structure of the candidate, a numerical identifier (13375), the lemmas of which the pair is composed (*produit*, *surgeler*) *((product, deep-frozen)),* its frequency (frq=4), the value of the statistical score (stat=29.16) and a summary of the variations encountered (VAR=1110).

```
Na 13375 produit surgeler frq=4  stat=29.16 VAR=1110
Flexion
    b007  nbr=1
    text1=produits surgelés
Modification
    i001  nbr=2
    text1=produits alimentaires surgelés nbo1=1
    text2=produits halieutiques surgelés nbo2=1
Coordination
    c001  nbr=1
    text1=produits congelés ou surgelés nbo1=1
```

Figure 1. Example of a candidate associated to a base structure

The variation summary is a string of four characters, each character receiving a value of 1 if the variation has been encountered, 0 if not. The first character represents flexional and internal morphosyntactic variants, the second internal modification variants, the third coordination variants and the last one predicative adjective variants. Next we find, for each type of variation, the rules which have been used (for example b007), the number of occurrences recognized by the rule, the initial text. The candidate term proposed to the expert is by default the most

frequent base form between the inventoried occurrences, *produits surgelés (deep-frozen products)* for the example in Figure 1.

4.2. *Modification of the termer*

To identify relational adjectives, we use their paraphrastic property. We group base-terms of Noun1 Prep (Det) Noun2 structure with terms of Noun1 Adj structure where either, Adj is derived from Noun2, such as *production de céréales* (*cereal production*), *production céréalière* (*cereal production*), or Adj owns a nominal use, such as *muscle bovin* (*bovine muscle*), *muscle de bovin* (*bovine's muscle*).
The identification of relational adjective takes place after the extraction of the occurrences of the candidates terms and their syntactic variations. Each candidate coming with its base structure and a summary of different variations encountered, it would have been possible to check that the relational adjective is not used in a predicative position, or with a degree modification if (Monceaux 1997) had not demonstrated that these properties are not always true.
Each candidate sharing a Noun Adj structure is examined thanks to its ending, relational adjectives owning determined endings.

Suffix	Number of rules	Example of rule	
-al	5	-al/	*national/nation*
-aire	8	-aire/	*dentaire/dent (dental/tooth)*
-atif	2	-atif/+e	*normatif/norme (normative/norm)*
-é	2	-é/+e	*carboné/carbone (carbonaceous/carbon)*
-el	2	-el/+e	*industriel/industrie (industrial/industry)*
-er	1	-er//+e	*paysager/paysage (landscaped/landscape)*
-eux	3	-eux/+e	*veineux/veine (venous/vein)*
-ien	1	-ien/+ie	*bactérien/bactérie (bacterial/bacterium)*
-ier	2	-ier/	*fruitier/fruit (fruit)*
-if	2	-if/	*sportif/sport (sport)*

-in	1	-in/	*porcin/porc (porcine/pork)*
-ique	15	-ique/+ie	*graphique/graphie (graphic/written form)*
-iste	1	-iste/+isme	*nationalisme/nationaliste*
-oire	1	-oire/+ion	*inflammatoire/inflammation (inflammatory)*

Table 1. Number and example of transformational rules by suffix

The most common suffixes in French applying to relational adjectives have been enumerated by (Guyon 1993): the favourable suffixes are: *-ain, -aire, -al, -el, -estre, -ien, -ier, -il, -in, -ique*, and the non-favourable ones are: *-esque, -eux, -é, -if, -oire*. We have left aside the suffixes: *-ain, -il, -in, -esque*. The suffixe *–ain* characterizes adjectives derived from proper nouns such as *toulousain* (*native of Toulouse*); *-il* and *–estre* are rare; *-esque* carries a disparaging sense such as *livre* (*book*), *livresque* (*bookish*).

For each suffix, we have create transformational rules in order to generate from the adjective the most predictable forms of the base noun. These rules have been established by hand thanks to examples of compound nouns of structure Noun AdjR found in (Guyon 1993) and (Monceaux 1993), and other readings. We do not pretend that they cover all the existing derivational forms but only the most frequent ones. Table 1 shows the number of rules written for each suffix and an example of rule. These rules generate one or several possible base nouns for a given adjective. This overgeneration method used in information retrieval by (Jacquemin and Tzoukermann 1997) gives few noise because the base noun must not only be an attested for in the corpus, but also must appear as an extension of a head noun. For example, with the adjective *ionique (ionic)*, we generate both *ionie (ionia)* and *ion (ion)*, but only *ion (ion)* is an attested form; with the adjective *gazeux (gaseous),* the nominals forms *gaz (gas)* and *gaze (gauze)* are generated and the two of them are attested;

```
3076 BASE=0101 nbcand=16 stat=32.36
npn 1693 fibre aliment frq=1  VAR=1000
Flexion
b014  nbr= 1
text1= fibres dans les aliments nbo1= 1
```

```
nar 1826 fibre alimentaire frq= 15  VAR= 1000
Flexion
b007 nbr= 15
text1= fibres alimentaires nbo1= 14
text2= Fibres alimentaires nbo2= 1
```

Figure 2. Example of a candidate associated to several base structures

but, the adjective *gazeux (gaseous)* appears with the noun *échange (exchange)* which is paraphrased in the corpus by *échange de gaz (gas exchange)* and not by *échange de gaze (gauze exchange)*.

The algorithm below resumes the successive steps for identifying relational adjectives:

1. Examine each candidate of Noun Adj structure;
2. Check that the adjective ends with one of the listed suffixes, with some adding constraints for some suffixes, such as for the suffix *–er* that the identified adjective is not a past-participle;
3. Apply a transformational rule in order to generate all the possible corresponding base nouns;
4. Search the set of candidate terms for a pair formed with Noun1 (identical between a Noun1 (Prep (Det)) Noun2 and a Noun1 Adj structures) and Noun2 generated from step 3.
5. If step 4 succeeds, group the two base structures under a new candidate term. Take out all the Noun Adj structures owing this adjective from the set of Noun Adj candidates and rename them as a Noun AdjR structure.

For non-ethnical adjectives with a nominal function, as for example *problème technique (technical problem)* and *problème de technique (problem of technics)*, we have accepted that a candidate term could share several base structures: one of type Noun1 (Prep (Det)) Noun2 and another of type Noun1 Adj.

No computation is needed for guessing Noun2 as Noun2 and Adj share the same lemma.

An example of the grouping of two base structures is given in Figure 2. The first line identifies the candidate thanks to an unique numerical identifier (3076), a

summary of the encountered base structures (BASE=0101), its frequency (nbcand=16) and its statistical score value (stat=32.36).

Number of occurrences of base structures	**1**	**≥ 2**	**Total**
Noun1 (Prep (Det)) Noun2	17 232	5 949	23 181
Noun Adj	12 344	4 778	17 122
Noun à Vinf	203	16	219
Total	**29 779**	**10 743**	**40 522**

Table 2. Number of base structures extracted from the [AGRO]

The base structure summary is a string of four characters, each receiving a value of 1 if the base structure has been encountered, or 0 if not. The first character represents the Noun Adj structure, the second the Noun1 (Prep (Det)) Noun2 structure, the third the Noun Prep Vinf structure and the last one the Noun AdjR structure. The other lines supply the same information as before (cf. Fig 1 and section 2.1) with the difference that the statistical score now applies to the grouping rather than the base structure.

5. Results and Evaluation

Our corpus, called [AGRO], is made up of 7272 abstracts (2.5 Mbytes) from French texts in the agriculture domain and extracted from PASCAL[ii]. We have used the Brill part-of-Speech Tagger (Brill 1992) trained for French by (Lecomte and Paroubek 1996) and the lemmatizer developed by F. Namer (Toussaint *et al.* 1998). This chain of corpus treatments is satisfactory, except for the treatment of agreements inside the sentence or the sentence phrases. It is impossible to check the agreement between the noun and the adjective and this leads to the extraction of erroneous candidate terms.

5.1. *Quantitative results*

Table 2 resumes the number of base structures extracted from [AGRO]. From these base structures, 395 grouping have been done. The linked presence of

nominal phrases of which the extension is fulfilled either by a relational adjective, or be a prepositional phrase the number is rare ---a little bit more than 1 % of the total of occurrences---. But, these groupings permits us to extract from the numerous hapax ---more than 70 % of the total of occurrences--- candidates which, we presume, will be highly denominative.

The number of relational adjectives which have been identified is 129: *agronomique (agronomical), alimentaire (food), arachidier (groundnut), aromatique (aromatic)*, etc.

6. Qualitative results

We checked the linguistic accuracy of the 395 structural variations which group a Noun1 Prep (Det) Noun2 structure and a Noun1 AdjR structure. Reported errors are the following:

- Generation of a "monster": *évolution normale (normal evolution)* linked to *évolution des normes (standard evolution)* (1 grouping);
- Three wrong grouping because of the homography, and the non homonymy, of the adjective and the noun: *fin (thin (Adj)/end (Noun)), courant (ordinary(Adj)/current(Noun)), potentiel (potential).*
- Three wrong groupings because of tagging errors where the frozen preposition *en fin de* has not been identified: *populations finales* linked to *population atteinte en fin*.

This lead us to a linguistic precision of 98 % in the identification of relational adjectives.

7. Evaluation with AGROVOC

The thesaurus AGROVOC[iii], is a taximony of about 15 000 terms associated with synonyms in a SGML format, which leads to 25 964 different terms. The thesaurus AGOVOC is used for indexing with data fitting agricultural retrieval systems and indexing systems.

We made two comparisons with AGROVOC: we first checked whether these relational adjectives were really part of terms of the thesaurus and second, we compared the candidate terms extracted with a relational adjective with the terms and the thesaurus. We consider that the presence of the relational adjective in AGROVOC confirms its informative character, and that the presence of a candidate term attests its terminological value.

7.1. *Relational adjectives alone*

From the 124 correct relational adjectives, 68 appear inside terms of the thesaurus in epithetic position, such as *continu (continuous), dynamique (dynamic), enzymatique (enzymatic), fruitier (fruit), gazeux (gaseous),* and 15 only under their nominal form in an extension position, for example the relational adjective *arachidier (groundnut)* does not appear but *arachide* is used in an extension position. Moreover, among the 124 adjectives, 73 appear in AGROVOC under their nominal term as uniterms. Among these 73 uniterms, 43 of them, such as *texture*, share a relational use, such as *textural*; the other 30, such as *estérase (esterase)* do not have a relational use.

The adjectives which are not present in the thesaurus in an extension position under either their adjectival or nominal form are a number of 11: *bibliographique (bibliographic), compartimental (compartment), coûteux (costly), haplotypique (haplotypic), logarithmique (logarithmic), miscellaire, neuronal (neuronic), opératoire (operating), photonique (photon), polyphénoloxydasique,* and *transmembranaire (transmembrane)*.

So 93 % of them are indeed highly informative.

This first result corroborates the assumption of the linguists that relational adjectives owns a denominative character.

8. Candidate term with a relational adjective

If relational adjective are indeed highly informative, we still have to demonstrate that the nominal phrase in which they appear are informative too.

For 9 relational adjectives, we have compared the nominal phrases in which they appear with terms of the thesaurus AGROVOC. We have done also this comparison for the nominal phrases with a prepositional extension including a

noun from which has been derived the relational adjective. Pour each adjective, we compute the following indexes:

T_A the number of terms in AGROVOC in which the relational adjective appears in an epithetic position, saying the terms of Noun AdjR structure. For example T_A = 15 for the adjective *cellulaire (cellular)* because it appears in 15 terms of AGROVOC such as *différenciation cellulaire (cellular differenciation)*, *division cellulaire (cellular division),* etc.

T_N the number of terms in AGROVOC in which the noun from which has been derived the relational adjective appears inside a prepositional phrase, saying the terms of Noun1 Prep (Det) $Noun_{AdjR}$ structure. For example T_A = 4 for the noun *cellule (cell)* because it appears in 4 terms of AGROVOC such as *banque de cellules (cell bank)*, culture de cellules *(cell growing),* etc.

	Noun AdjR	Noun1 Prep (Det) $Noun_{AdjR}$
Precision	0,34	0,04
Recall	0,46	0,14

Table 3. Averages of precisions and recalls

C_A the number of candidate terms of Noun AdjR structure. For example C_A = 61 for the adjective *cellulaire (cellular)* because it appears in 61candidate terms such as *acide cellulaire (cellular acid)*, *activité cellulaire (cellular activity), agrégat cellulaire (cellular aggregate),* etc.

C_N the number of candidate terms of Noun1 Prep (Det) $Noun_{AdjR}$ structure. For example C_N = 58 for the noun *cellule (cell)* because it appears in 58 candidate terms such as *ADN de cellule (cell DNA)*, *addition de cellules (cell addition),* etc.

Then, for each candidate term of C_A and C_N , we have checked its belonging to AGROVOC. The only matches that we have accepted are exact matches. With this comparison, we obtained the following indexes:

a the number of candidate terms of Noun AdjR structure which have been found in AGROVOC under the Noun AdjR structure.

b the number of candidate terms of Noun AdjR structure which have been found in AGROVOC under the Noun1 Prep (Det) $Noun_{AdjR}$ structure.

c the number of candidate terms of Noun1 Prep (Det) $Noun_{AdjR}$ structure which have been found in AGROVOC under the Noun AdjR structure.

d the number of candidate terms of Noun1 Prep (Det) $Noun_{AdjR}$ structure which have been found in AGROVOC under the Noun1 Prep (Det) $Noun_{AdjR}$ structure.

These indexes allow us to compute precision P and recall R for each Noun AdjR structure and each Noun1 Prep (Det) $Noun_{AdjR}$ structure with the help of the following formula:

$$P_{\text{Noun AdjR}} = (a + b) / C_A \qquad R_{\text{Noun AdjR}} = (a + b) / T_A$$

$$P_{\text{Noun1 Prep (Det) NounAdjR}} = (c + d)/ C_N \qquad R_{\text{Noun1 Prep (Det) NounAdjR}} = (c + d)/ T_N$$

The averages of precisions and recalls for the two structures are summarized in table 3. This comparison of the average of precisions computed shows that candidate terms with a Noun AdjR structure have 10 times more chance to be terms than their equivalent in Noun1 Prep (Det) $Noun_{AdjR}$. The analysis of the average of recalls is too impressive: it is generally difficult to obtain a recall superior to 25 % when comparing candidate terms extracted from a corpus and a thesaurus of the same domain (Daille *et al.*1998). The average of recalls obtained thanks to the identification of relational adjective shows that nearly half of the terms built with the defined relational adjectives are identified. These good values of precision and recall have been obtained on linguistic criteria only without taking into account frequency.

Moreover, we notice that even if AGROVOC integrates quite the same number of terms of Noun Adj structure and of Noun1 Prep (Det) $Noun_{AdjR}$ structure, there is not real coherence in the thesaurus encoding; for example, we *find métabolisme protéique (proteinic metabolism)*, but not *métabolisme glucidique (carbonhyfrate metabolism)*, the term *métabolisme des glucides (metabolism of carbonhydrate)* being preferred. Our results could be used to update the thesaurus when the form Noun AdjR has been encountered in a corpus.

8.1. *Analysis of Syntactical Variations*

The Noun AdjR structures are few submitted to syntactical variations contrarily to their equivalent in Noun1 Prep (Det) $Noun_{AdjR}$ structures when they exist: only 10 % of Noun AdjR structures accept variations and those are mainly coordination such as *produit alimentaire/produits agricoles et alimentaire (food product/agricultural and food product)* and some modification such as *composition lipidique cellulaire (cellular lipid composition).* We have examined all the coordinations and we notice that they involve other relational adjectives. This result could be used to identify new relational adjectives, in the same way that (Jacquemin 1996) did for term acquisition through indexing. Modifications are more difficult to exploit in the case of an inserted adjective: either the adjective is indeed relational such as *lipidique (lipid)* in *composition lipidique cellulaire (cellular lipid composition)*, or, the adjective is part of a compound such as *gras (fat)* in *matière grasse industrielle (industrial fat content).* No predicative variants have been encountered. An interpretation could be that such structure are few used in technical domain.

On the other hand, 60 % of the Noun1 Prep (Det) Noun2 structures grouped to a Noun AdjR structure accept syntactic variations, mainly modifications, *as filtration membranaire (membrane filtering)/filtration frontale sur membrane (frontal filtering on membrane)*. Some groupings are of particular interest: those where the Noun AdjR structure has been identified, but lacks the Noun1 Prep Noun2 base form. For example, whereas *acide vinique (vinic acid)* is attested, only syntactic variations of *acide (acid)* Prep (Det) *vin (wine)* are identified: *acides organiques du vin (organic acid of the wine)*, *acide malique dans le vin (malic acid of the wines), acides aminés des vins (amino-acid of the wines)*, *acide tartrique dans les vin (tartaric acid in the wines), acide salicylique dans les vins (salicyclic acid in the wines).* The term *acide vinique (vinic acid)* groups all these different acids as *acides organiques, maliques, aminés, tartriques et salicyliques (organic, malic, amino-, tartaric, and salicylic acids)* which are all kinds of acids existing in AGROVOC. These groupings constitute a first step in knowledge acquisition from texts, and, for example, could be used as the bootstrap in a system for acquiring semantic relations, such as PROMETHEE (Morin 1998).

9. Conclusion

We succeeded in the identification of relational adjectives by finding both Noun1 (Prep (Det)) Noun2 and Noun1 AdjR structures in texts. This experience corroborates the linguist's studies and their intuition about the informative character of the relational adjectives. Identifying relational adjectives could thus be used to recover uniterms from corpora. We have also proved that nominal phrase including a relational adjective are by far more informative than their equivalent in Noun1 Prep (Det) $Noun_{AdjR}$ structure. The method presented is robust even if it does not allow us to identify exhaustively all relational adjectives appearing in a corpus. But, we saw how improving such method by exploiting coordination variants. Taking into account such lexical units is interesting for terminology extraction, but also for updating thesauri or for technological development where the occurrence of a relational adjective represents a stabilization of an emerging scientific concept.

References

Abney, S. 1991, Parsing with Chunks. *Principle-Base Parsing*. Berwick, R. and Tenny, C. (ed.). Kluwer Academic Publishers, pp. 257—278.

Assadi, S. and Bourigault, D. 1995. Classification d'adjectifs extraits d'un corpus pour l'aide à la modélisation de connaissances. Troisièmes Journées Internationales d'Analyse Statistique de Données Textuelles, pp. 313—320, Roma.

Bally, C. 1965. *Linguistique générale et linguistique française*. Berne:Francke Verlag, 4th edition.

Bartning, I. 1976. *Remarque sur la syntaxe et la sémantique des pseudo-adjectifs dénominaux en français*. Ph.D. thesis, University of Stockholm.

Basili, R., Pazienza, M.T. and Velardi, P. 1993. Acquisition of Selectional Patterns in Sublanguages, *Machine Translation*, vol. 8, pp. 175—201.

Bourigault, D. 1994. LEXTER, un Logiciel d'Extraction de TERminologie. Application à l'acquisition des connaissances à partir de textes. Ph.D. thesis, EHESS, Paris, France.

Brill, E. 1992. Asimple rule-based part of speech tagger. *Third Conference on Applied Natural Language Processing (ANLP'92).* pp. 152—155.

Daille, B., Gaussier, E. and Langé, J.M. 1998. An evaluation of Statistical Scores for Word Association. *The Tbilisi Symposium on Logic, Language and Computation: Selected Papers*, chap. 11, pp. 177—188, CSLI Publications.

Daille, B. and Jacquemin, C. 1998. Lexical database and information access: a fruitful association. *LREC'98.*pp. 669—673.

Dal, G., Hathout, N. and Namer, F. 1999. Construire un lexique dérivationnel : théorie et réalisations. *TALN'99.* pp. 115—124.

David, S. and Plante, P. 1990. Le progiciel TERMINO : de la nécessité d'une analyse morphosyntaxique pour le dépouillement terminologique des textes. *ICO,* vol. 2(3), pp. 140—155.
.

Dubois, J. 1962. *Etude sur la dérivation suffixale en Français moderne et contemporain,* Paris, Larousse.

Guyon, A. 1993. Les adjectifs relationnels arguments de noms prédicatifs, PhD thesis, University Paris 7.

Jacquemin, C. 1996. A symbolic and surgical acquisition of terms through variaition, *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, Heidelberg, Springer, pp. 425—438.

Jacquemin, C. and Tzoukermann, E. 1997. NPL for term variant extraction: Synergy between morphology, lexicon and syntax, *Natural Langague Information Retrieval*, chap. 2, T. Strzalkowski (eds.), Kluwer Academic Publishers, Dordrecht.

Justeson, J. and Katz, S. 1995. Technical terminology: Some linguistic properties and an algorithm for identification in text. *Journal of Linguistic Engineering*, vol. 1(1), Cambridge University Press.

Kleiber, G. 1990. *La sémantique du prototype – Catégories et sens lexical*. PUF, Linguistique nouvelle.

Lecomte, J. and Paroubek, P. 1996. *Le catégoriseur d'Eric Brill. Mise en œuvre de la version entraînée à l'INALF.* Technical report. CNRS-INALF.

Lerat, P. 1995. *Les langues spécialisées*. Paris, PUF

Levi, J. 1978. *The syntax and the semantics of complex nominals*. Academic Press, London.

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*. PhD. Thesis, University of Marne la Vallée, France.

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. *TALN'98*, pp. 172-181.

Postal, P.M. 1969. Anaphoric island. *Fifth Regional Meeting of the Chicago Linguistic Society.* R.B. et al. (eds.), University of Chigago.

Silberztein, M. 1994. Intex: a corpus processing system. *COLING'94*, pp. 579—583.

Tamba-Mecz, I. 1980. Sur quelques propriétés de l'adjectif de relation. Travaux de Linguistique et de Littérature, 18-1:119—132.

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. *TALN'98*, pp. 182—191.

Zribi-Hertz, A. 1972. Sur un cas de construction pseudo-predicative*, Recherches Linguistiques,* 1:159—168.

[i] A derivational lexical database is being built for French inside the FRANLEX project (http://www.limsi.fr/Individu/jacquemi/FRANLEX) (Dal et al. 1999)

[ii] PASCAL is the Scientific Documentary Database maintained by INIST-CNRS, France.

[iii] A multilingual thesaurus developed by AGRIS (International Information System for Agricultural Sciences and Technology)