

Detecting semantic relations between terms in definitions

Véronique MALAISE^{1,2}
¹ STIM/AP-HP, ERM 202 INSERM
 & CRIM-INaLCO
 91, boulevard de l'Hôpital
 75013 Paris,
 France,
 {vma,pz}@biomath.jussieu.fr

Pierre ZWEIGENBAUM¹

Bruno BACHIMONT²

²DRE de l'INA
 Institut National de l'Audiovisuel
 4, avenue de l'Europe
 94366 Bry-sur-Marne Cedex,
 France,
 {vmalaise,bbachimont}@ina.fr

Abstract

Terminology structuring aims to elicit semantic relations between the terms of a domain. We propose here to exploit definitions found in corpora to obtain such semantic relations. Definition typologies show that definitions can be introduced by different semantic relations, some of these relations being likely to structure terminologies. Our aim is therefore to mine “defining expressions” in domain-specific corpora, and to detect the semantic relations they involve between their main terms. We use lexico-syntactic markers and patterns to detect at the same time both a definition and its main semantic relation. 46 markers and 74 patterns have been designed and tuned on a first corpus in the field of anthropology. We report on their evaluation on a second corpus in the field of dietetics, where they obtained 4% to 36% recall and from 61 to 66% precision, and discuss the relative accuracy of different subclasses of markers for this task.

1 Introduction

A terminology is an artifact structuring terms according to some semantic relations. Grabar and Hamon (2004) present the different semantic relations likely to be found in terminologies. These can be divided into lexical (synonymy), vertical (hypernymy, meronymy) and transversal relations (domain-specific relations). A study of definition typologies, like the one of (Auger, 1997), shows that these different relations are also present in definitions. We can then hypothesise that mining definitions along with the detection of their inherent semantic relation can help to organise terms according to the relations used in structured terminologies. We focus in this paper on the detection of terms related by hypernymy and synonymy in definitions.

The automatic detection of definitions can rely on different types of existing works. We can, first, consider the studies describing what definition is, and more particularly what definition in corpus is like. In this respect, we can cite the work of Trim-

ble (1985), Flowerdew (1992), Sager (2001) and Meyer (2001). Another type of interesting existing work is about typologies of definitions: Martin (1983), Chukwu and Thoiron (1989) and Auger (1997), amongst others, provide, in their classifications of definitions, linguistic clues to find defining statements in corpus. We propose to integrate the typologies that we mention in section 2.2, along with the linguistic clues they give: the definition markers. And, at last, some works have already focused on mining definitions from corpora, including Cartier (1997), Pearson (1996), Rebeyrolle (2000) and Muresan and Klavans (2002), mostly through the use of lexical definition markers. These works provide us with methodological guidelines and another set of lexical markers for our own experiment.

As (Pearson (1996); Rebeyrolle (2000)), our method is based on lexico-syntactic patterns, so that we can build on the work on French language by Rebeyrolle (2000). We extended her work in two respects: an analysis of the parenthesis as low-level linguistic clue for definitions, and the concomitant extraction of the semantic relation involved in a “defining expression”, along with the extraction of the definition itself. Previous works have, for instance, mined definitions to find terms specific to a particular domain of knowledge (Chukwu and Thoiron (1989)), and to describe their meaning (Rebeyrolle, 2000); we focus on the detection of the semantic relations between the main terms of a definition in order to help a terminologist to build a structured terminology following these relations.

We implemented an interface to visualise these definitions and semantic relations extractions. We tuned markers and patterns for extracting definitions and semantic relations on a first corpus about anthropology; we then tested the validity of these markers and patterns on another corpus focused on dietetics. The purpose of this test was, on the one hand, to observe whether definitions were still correctly extracted on the basis of patterns trained on a corpus differing in the domain of knowledge and in the genre of documents involved, and, on the

other hand, to detect if the semantic relation associated with each pattern was the same as the one observed in the first corpus. The markers and patterns showed to be comparable to the other experiments mentioned in terms of definition extraction: the precision reached from 61 to 66%. As for the semantic relation associated with the patterns, it obtained different scores, depending on the marker. But, in most cases, one main semantic relation is associated with a pattern in the scope of a single domain, event though a few patterns convey the same relation across our two corpora.

The remainder of this paper is organised as follows: we first present previous work (section 2), describe our method and experiment (section 3), then present and discuss results (section 4) and conclude with directions for future work (section 5).

2 Previous work

2.1 Description of definitions in corpus

As a first approach for detecting and extracting defining statements in corpora, we have to... define this object. In the literature (Trimble (1985); Flowerdew (1992),...), three categories of definitions are often mentioned: the formal definition, the semi-formal and the “non-formal” one. The formal definition follows the Aristotelian schema: $X = Y +$ specific characteristics, where X is the defined term (the “*definiendum*”), “=” means an equivalence relation, Y stands for the generic class to which X belongs (the “*Genus*”), and specific characteristics detail in which respect X is different from the other items composing the same generic class. A semi-formal definition relates the definiendum only with specific characteristics, or with its attribute(s) (Meyer, 2001). Formal and semi-formal definitions can be of simple type (expressed in one sentence), or complex (expressed in two, or more sentences). A non-formal definition aims “*to define in a general sense so that a reader can see the familiar element in whatever the new term may be*” (Trimble, 1985). It can be an association with a synonym, a paraphrase or grammatical derivation.

The common point between all these points of views on the same linguistic object, or between all these different objects sharing the same appellation “definition”, is that they all follow the same didactic purpose of disambiguating the meaning of a lexical item, that is to distinguish it from the others in the general language, or inside a specific vocabulary. These definition descriptions present them as the association between a term and its hypernym (its “*genus*”), or between a term and its specific characteristics. But there are yet other ways to express

definitions, as the works on their typology shows.

2.2 Typology of definitions

Existing definitions typologies are all dedicated to a specific purpose. We are particularly interested in those which aim at eliciting linguistic clues that can be used to mine defining contexts from corpora. We work on French, for which Martin (1983) has classified dictionary definitions in order to give guidelines for a consistent (electronic) dictionary. In the context of corpus-based research, Chukwu and Thoiron (1989) gave another classification, aiming at finding domain-specific terms in corpora. A unified typology is provided by Auger (1997), compiling both cited typologies along with three others, and from which we draw the following three categories:

- Definitions expressed by “low level” linguistic markers: punctuation clues such as parenthesis, quote, dash, colon;
- Definitions expressed by lexical markers: linguistic or metalinguistic lexical items;
- Definitions expressed by “high level” linguistic markers: syntactic patterns such as anaphora or apposition.

The definitions introduced by lexical means are divided in two branches, characterised by the lexical markers in table 1. We added elements from other studies ((Rebeyrolle, 2000) and (Fuchs, 1994) amongst others), and augmented this typology with

Definitions introduced by linguistic markers	
Copulative	“ <i>a X is a Y that</i> ”
Equivalence	“ <i>equivalent to</i> ”
Characterisation	“ <i>attribute of</i> ”, “ <i>quality</i> ”,...
Analysis	“ <i>composed of</i> ”, “ <i>equipped with</i> ”, “ <i>made of</i> ”,...
Function	“ <i>to have the function</i> ”, “ <i>the role of</i> ”, “ <i>to use X to do Y</i> ”,...
Causality	“ <i>to cause X by Y</i> ”, “ <i>to obtain X by</i> ”,...
Definitions introduced by metalinguistic markers	
Designation	“ <i>to designate</i> ”, “ <i>to mean</i> ”,...
Denomination	“ <i>to name</i> ”
Systemic	“ <i>to write</i> ”, “ <i>to spell</i> ”, “ <i>the noun</i> ”,...

Table 1: Lexical markers (English translation)

new markers, including some items introducing reformulation contexts (“*that is*”, “*to say*”, “*for instance*”, ...).

The Aristotelian definition type is presented here as a “copulative” definition, as it is linguistically marked by the copula “être” (to be). It involves a hypernymic relation (and specific differences) to describe the meaning of a term, so we consider it as a “hypernymic definition”. But we can see in table 1 that other semantic relations can also be used to define a term: synonymy (definition of “equivalence” type), meronymy (“analysis” type), causality and other domain-specific transversal relations (“function”, “characterisation” types). Mining a definition of “synonymic type” provides different denominations for the same concept; one of “hypernymy type” can help modelling the vertical structure between the “*definiendum*” and the first term of the “*definiens*” (conceptual “father” and “son” association); and definitions following transversal relations allow the expression of specific knowledge. We focus in this paper on the extraction of definitions involving hypernymy and synonymy, which are the most generally considered relations in terminology building.

2.3 Automatic definition mining

Automatic definition mining from corpora can be divided in different groups, according to the methodologies followed. We will illustrate them by describing three recent families of works: (i) Cartier (1997), (ii) Pearson (1996) and Rebeyrolle (2000), (iii) Muresan and Klavans (2002). They have used respectively “contextual exploration”, lexico-syntactic patterns and linguistic analysis and rules.

The former one extracts defining statements on the basis of the match of linguistic clues, when they are relayed in the sentence by some linguistic rules. These rules are developed by the author, withing the schema defined in the “contextual exploration” methodology (Desclés, 1996).

Pearson (1996) and Rebeyrolle (2000) have followed the methodology described by Hearst (1992), up to now mainly applied to discover hyponymous terms. It consists in describing the lexico-syntactic context of an occurrence of a pair of terms known to share a semantic relation. Modelling the context in which they occur provides a “pattern” to apply to the corpus, in order to extract other pairs of terms connected by the same relation. Pearson and Rebeyrolle have modelled lexico-syntactic contexts around lexical clues interpreted as “definition markers”. Rebeyrolle, working on French, evaluated the different pattern types she modelled, across different corpora:

she obtained a precision range of 17.95 – 79.19%, and a recall of 94.75 – 100%. The difference between the two numeric boundaries of the precision range is due to the kind of markers involved in the lexico-syntactic pattern evaluated: metalinguistic markers obtained a high precision rate, but not linguistic lexical markers.

The latter pair of authors have based their system DEFINDER (<http://www1.cs.columbia.edu/~smara/DEFINDER/>) on the lexical and syntactic analysis of a medical corpus, with semi-automatic definition acquisition. Their evaluation is focused on the usefulness of the system, as compared with existing specialised medical dictionaries. They reach a 86.95% precision and 75.47% recall, following their evaluation methodology.

We chose to follow the first methodology in our experiment (see section 3), in which we additionally explore definition mining in some cases where the definition is not introduced by lexical items. Following this methodology enables us to build on existing work dedicated to French, which showed to be interesting and efficient. The lexico-syntactic pattern methodology also enables us to access the different linguistic elements we were interested in mining: the definition itself, the main terms of the definition and the semantic relation between them. We focus this experimentation more particularly on identifying the semantic relations of synonymy and hypernymy involved in the different definitions likely to be found in corpora. We aim at testing whether a stable link can be established between the definition extraction pattern and a specific semantic relation.

3 Detecting Semantic Relations

Our goal is to automatically detect some of the semantic relations that might be found in definitions and to propose them to a human validator in charge of structuring a terminology. We focus on hypernymy and synonymy, which are the most classical relations found in terminology. If the relation is hypernymy, the terms are to be modelled in a hierarchical way, if it is synonymy, both terms can be used to express the same concept. The relations and the definitions are extracted together from corpora, by the same lexico-syntactic patterns. We present in the next subsections our two corpora (section 3.1), then the lexico-syntactic patterns we used (section 3.2) and their experimental evaluation (section 3.3): we analyse whether a relation found in connection with a lexico-syntactic pattern in the training corpus can be unchanged in the context of the same lexico-syntactic pattern, when applied to a

different corpus.

3.1 Description and preparation of the corpora

Our training corpus (76 Kwords) is focused on childhood, from the point of view of anthropologists. It is composed of different genres of documents (documentary descriptions, thesis report extracts, Web documents). Documentary descriptions were humanly collected, whereas electronic documents were automatically collected from Internet via the tools of (Grabar and Berland, 2001). Our evaluation corpus (480 Kwords), in the domain of dietetics, is composed of Web documents indexed by the CISMeF quality-controlled catalog of French medical Web sites (<http://www.chu-rouen.fr/cismef/>) in the subtrees “Dietetics” and “Nutrition” of the MeSH thesaurus. It is mainly composed of medical courses and Web pages presenting information about nutrition in different medical contexts. Both corpora were morpho-syntactically analysed by Cordial Analyser (Synapse Developpement, <http://www.synapse-fr.com/>). Cordial tags, lemmatises and parses a corpus, yielding grammatical functions (subject, object, ...) between chunks.

3.2 Lexico-syntactic patterns

A given linguistic marker (see, *e.g.*, table 1) can occur in different contexts, some of which are definitions, and can be a clue for different semantic relations. Lexico-syntactic patterns aim at reducing this ambiguity by specifying more restricted contexts in which a definition is found, and, furthermore, in which one specific semantic relation is involved.

Unlike (Hearst, 1992), we started the pattern design by analysing marker occurrences in our training corpus. We designed and tuned our lexico-syntactic patterns on this corpus, patterns dedicated to the extraction of definitions and specific relations: hypernymy and synonymy. Our patterns use the information output by the parser, including lemma, morpho-syntactic category and grammatical function. For instance: “*N(N)*” specifies that the marker “(” has to be preceded by a noun, and immediately followed by a single common noun, followed by a closing parenthesis. In this specific case, “(” introduces a hypernymic definition.

Each pattern drives different kinds of processing:

- extraction of the defining sentence on the basis of the whole pattern;
- selection of one “preferred” relation associated with the specific pattern, among the set of possible relations associated with the marker; this

relation stands between the interdefined terms of the definition;

- extraction of the interdefined terms following two strategies (contextual or based on dependencies around the marker), depending on the morphosyntactic category of the marker. When the marker is a punctuation or a noun, we usually extract its left and right syntactic contexts¹ (roughly the first chunk before the marker, and the first chunk after the marker in the sentence). When the marker is a verb, we extract its subject and object if they exist in the sentence, otherwise we extract its left and right chunks, as in the previous case.

Our patterns are implemented in XSLT and the resulting extractions are shown to a human validator through a Web interface (figure 1): an HTML form allowing the validator to complete and correct the extractions. It is possible for the validator to correct the terms extracted from the definition, in particular because the chunk often includes punctuation, which is usually not considered as part of the term, and it is possible to select a different semantic relation than the one proposed when it happens not to be the correct one. A combo box shows all the possible relations related to the marker involved in the lexico-syntactic pattern which provided the extraction of the defining sentence.

3.3 Experimental setup

We tuned our lexico-syntactic patterns to extract definitions from the test corpus. We associated with each pattern a “preferential” semantic relation, which human corpus analysis showed to be the more likely to be connected to the definitions extracted by the means of this pattern. The aim of the experiment is to test the stability of this connection, by applying the patterns to the evaluation corpus.

A random sample of the test corpus (13 texts among 132) was manually processed to tag its definitions, in order to have a standard measure for the evaluation of recall. Table 2 shows the number of definitions of synonymic and hypernymic types found in that sample, and provides the percentages of these definitions among all the different kinds of tagged definitions (“% *definitions*”) in that sample. Some definitions involved more than one semantic relation, so we also present the percentage of hypernymic and synonymic relations among all the semantic relations (“% *relations*”).

¹Depending on the position of the marker in the sentence, it might be the two following or two preceding chunks.

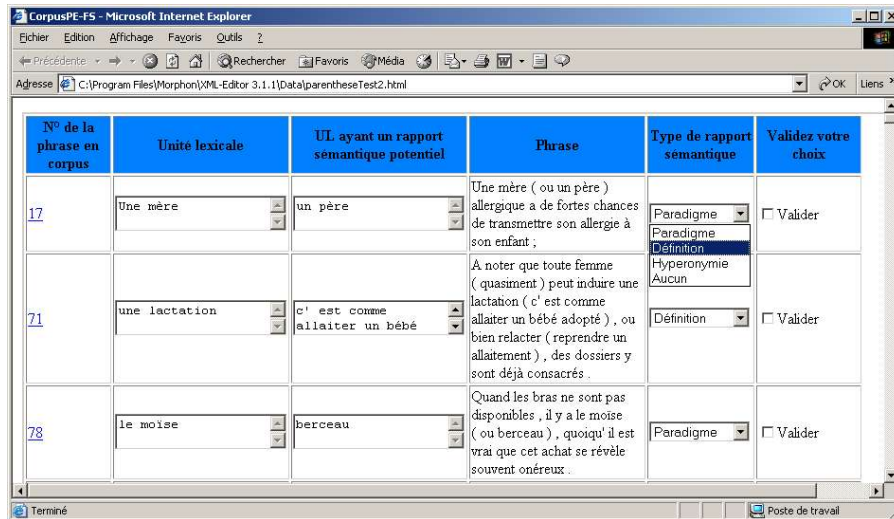


Figure 1: Human validation interface for definitions extracted with the parenthesis marker

	Hypernymy	Synonymy
# definitions	90	22
% definitions	44, 5%	10, 8%
% relations	39, 1%	9, 5%

Table 2: Number and percentages of hypernymic and synonymic definitions in a random sample of the test corpus, according to the human evaluator

In our experiment, we evaluate in turn the quality of the extracted definitions, then that of semantic relations (hypernymy and synonymy).

4 Results and discussion

Table 3 shows the number of markers and patterns prepared and tuned on the training corpus to extract definitions based on hypernymy or synonymy. Note that a given marker can be used in different patterns to extract different semantic relations. Some markers were also associated in one pattern: the metalinguistic nouns and verbs. We combined them because their individual recall was not lowered by this association and their precision score was improved. The sentences below are examples of sen-

	Hypernymy	Synonymy
# markers	3	43
# patterns	4	70

Table 3: Number of markers and patterns

tences extracted by our system; the underlined part is the marker:

- Hypernymic relation:

“*Les acides gras de la série omega-3 (MAX-epa) peuvent également être prescrits .”*,
 “[...] *les fromages à pâte cuite (tels que par exemple le fromage de Hollande) .”*”

- Synonymic relation:

“*L’ activité physique est définie comme tout mouvement corporel produit par la contraction des muscles squelettiques ,[. . .]*”
 “*une relation inverse entre l’ activité physique et l’ insulinémie ou la sensibilité à l’ insuline est habituellement observée .”*”

Table 4 presents the evaluation results: we divide them according to the semantic relation extracted. It shows the number of definitions retrieved, and the associated precision and recall. Precision is divided in two measures.

	Hypernymy	Synonymy
# extracted sentences	270	585
Precision (def)	61%	66%
Precision (rel)	26%	15%
Recall (rel)	4%	36%

Table 4: Evaluation of precision (test corpus) and recall (random sample of test corpus)

- the proportion of extracted sentences that corresponded to definitions (*def*), and
- the proportion of correct semantic relations found in retrieved definitions (*rel*).

Recall is the proportion of retrieved definitions which correctly display the semantic relation *identified in the sample corpus* among all the definitions present in this sample which were tagged as having this semantic relation by the human evaluator.²

The precision of extracted definitions is comparable to Rebeyrolle's results. The precision of semantic relations is much lower, but a global evaluation does not show the particular behavior of some of the markers. We list below the markers which were actually involved in the extraction of definitions in the test corpus.

- Markers implied in hypernymic definition retrieval: "*parenthèse*" (parenthesis), "*par exemple*" (for instance), "*sorte de*" (a kind of);
- Markers implied in synonymic definition retrieval: "*parenthèse*" (parenthesis), "*il s'agit de*" (as for), "*indiquer*" (to indicate), "*soit*" (that is), "*expliquer*" (to explain), "*préciser*" (to specify), "*marquer*" (to mark), "*enfin*" (say), "*ou*" (or), "*comme*" (as), "*à savoir*" (that is), "*autrement dit*" (in other words), "*au sens de*" (meaning), "*équivaloir*" (to be equivalent), "*c'est-à-dire*" (that is), "*définir*" (to define), "*désigner*" (to designate), "*nommer*" (to name), "*dénommer*" (to name), "*référer*" (to refer), "*expression*" (expression), "*terme*" (term).

Table 5 presents the different semantic relations found in the definitions retrieved by each marker. The first column references the markers involved in the extraction of the definition, the second ("*Expected*") presents the number of definitions, extracted by each marker, following the expected relation. "*Other*" gives the number of retrieved definitions following another semantic relation, "*Undecidable*" represents the number of definitions for which we could not determine the semantic relation,³ and "*Non definition*" presents the number of retrieved sentences that were not definitions.⁴

²The percentage of definitions of hypernymic and synonymic type among all definitions in the sample of the test corpus is given in table 2.

³Because our system extracts only one sentence, and a larger context was necessary to understand the semantic relation involved, or because of a problem in the conversion of some HTML documents to texts for the evaluation corpus.

⁴Except sentences presenting terms in a paradigm context, which is also interesting for terminology structuring. We in-

Definitions retrieved with the hypernymy patterns involved very generic markers, and they introduced a number of other semantic relations. The pattern around "*for instance*", for which 16 extracted sentences out of 95 were not definitions, can still be specified to discriminate defining contexts from others. We can notice, though, that it is one of the most productive patterns (95 extractions) and that it reaches a 47,3% precision. But the patterns around the parenthesis show that the same syntactic context can introduce different kinds of relations: in this case, the lexico-syntactic pattern cannot disambiguate the relation any further. The pattern "*N(N)*" introduced "hypernymic definitions", as well as "synonymic" or "meronymic" ones, the same syntactic context being even likely to be interpreted as a transversal relation between a treatment and a disease, for instance. It is the sentence as a whole that has to be interpreted in order to be able to define the relevant semantic relation between the terms in that syntactic context.

Some linguistic markers (as "*comme*") are reliable for detecting a semantic relation: 9 sentences out of 13 were "synonymic definitions". But surprisingly enough, some metalinguistic verbs ("*définir*", for instance) were not as effective as them in that purpose. "*Définir*" introduced only 22 "synonymic definitions" out of 68 sentences retrieved. One could think that a verb with metalinguistic function could be less polysemic than another of more "generic purpose". This naive hope happens to be wrong: "*Définir*" means "to fix (a limit)" as often as "to define". Some markers steadily introduced a semantic relation, but not the one they were supposed to: this variation is probably due to the change in domains across our two corpora. And some patterns obviously introduced a definition, but the defined element was in the previous sentence (this is the case of 92 extractions with patterns involving the marker "*Il s'agit de*"). As our system, up to now, extracts only one sentence, we could not determine whether the semantic relation was the one expected. We must address this problem, and we can hope that the precision rate will then be better than the one presented here: some sentences for which we could not interpret the semantic relation might convey the one we expected. The best precision score is reached by patterns involving two markers: a metalinguistic noun associated with a metalinguistic verb. In a more general way, analysing the defining sentences extracted, we could see that sentences that were the "best" definitions (the closest to dictionary definitions) often included this paradigm context in the "*Other*" column.

Marker	Expected	Other	Undecidable	Non definition	Total
Parenthesis (Parenthèse)	Hyp: 25	Meronymy: 1, Synonymy: 38 (+3), Transversal: 7	4	84	163
For instance (Par exemple)	Hyp: 45	Transversal: 2	32	16	95
A kind of (Une sorte de)	Hyp: 1	Transversal: 2	5	5	13
Parenthesis (Parenthèse)	Syn: 10	Paradigm: 9	2	4	25
As for (Il s'agit de)	Syn: 10	Transversal: 4, Hypernymy: 1	92	9	115
To indicate (Indiquer)	Syn: 5	Transversal: 12	6	77	100
That is (Soit)	Syn: 7	Paradigm: 31, Transversal: 13	15	1	66
To explain (Expliquer)	Syn: 1	Transversal: 21	15	28	65
To specify (Spécifier)	Syn: 1	Transversal: 5	9	26	41
To mark (Marquer)	Syn: 1	Transversal: 7	6	12	26
Say (Enfin)	Syn: 0	Paradigm: 3	2	1	6
Or (Ou)	Syn: 3	Paradigm: 23	1	0	27
As (Comme)	Syn: 9	Paradigm: 1	1	2	13
That is (A savoir)	Syn: 4	Hypernymy: 3	5	0	12
In other words (Autrement dit)	Syn: 1	0	2	0	3
Equivalent to (équivaloir)	Syn: 0	0	4	0	4
To define (Définir)	Syn: 22	Transversal: 8	19	19	68
To designate (Désigner)	Syn: 3	Hypernymy: 0	0	0	3
Term (Terme)	Syn: 1	0	1	0	2
Meaning (Au sens de)	Syn: 0	0	1	0	1
That is (C'est-à-dire)	Syn: 1	0	0	0	1
To name (Nommer)	Syn: 0	0	2	1	3
To name (Dénommer)	Syn: 1	0	0	0	1
To refer to (Référer)	Syn: 0	0	0	2	2
Expression (Expression)	Syn: 0	1	1	0	2

Table 5: Semantic relations in retrieved definitions

involved two or even three markers. This underlines the interest of introducing a relevance measure that takes into account the number of markers present in the sentence.

5 Conclusions

Our experiment tried to link the semantic relation inherent to different kinds of definitions with the marker (the heart of our lexico-syntactic patterns) and more specifically with the lexico-syntactic patterns at the origin of the extraction of the definition itself. Having a close look at some of the markers, we can observe that some linguistic items can be very reliable markers for definition extraction associated with a semantic relation. We can also find out that the polysemy of some markers is related to the domain of the corpus. In that respect, the reusability of the lexico-syntactic patterns is limited

to a set of markers which were found to be reliable across our two corpora. What is more problematic is the fact that it is sometimes not possible to make a specific distinction between different semantic relations detected with the same marker in the context of definitions sharing most of their syntactic contexts. But most of the patterns retrieve a good rate of defining sentences, some patterns being more reliable than others; and the more numerous the markers involved, the more likely it is that we have a definition. And usually these patterns retrieve definitions following one main semantic relation (this is not the case however for parenthesis and the patterns involving the marker “à savoir”). This leads to the hypothesis that if lexico-syntactic patterns may not be used to propose semantic relations that are valid across different domains, they remain a good clue for mining definitions, especially defini-

tions of one type of semantic relation *inside* a given domain. Moreover, given a new corpus, applying the existing patterns to a sub-corpus could lead to the elicitation of the associated semantic relations for that corpus, which could be a relevant methodology to discover pairs of terms following these associated relations.

References

- A. Auger. 1997. *Repérage des énoncés d'intérêt définitoire dans les bases de données textuelles*. Thèse de doctorat, Université de Neuchâtel.
- E. Cartier. 1997. La définition dans les textes scientifiques et techniques : présentation d'un outil d'extraction automatique de relations définitoires. *2e Rencontres "Terminologie et Intelligence Artificielle" (TIA'97)*, Equipe de Recherche en Syntaxe et Sémantique. Toulouse, 3-4 avril 1997:127–140.
- U. Chukwu and P. Thoiron. 1989. Reformulation et repérage des termes. *La Banque des Mots*, Numéro spécial CTN - INaLF - CNRS:23–53.
- J.-P. Desclés. 1996. Systèmes d'exploration contextuelle. *Table ronde sur le Contexte*, avril 1996, Caen.
- J. Flowerdew. 1992. Definitions in science lectures. *Linguistics*, vol.13 (2):202–221.
- C. Fuchs. 1994. *Paraphrase et énonciation*. Paris, Ophrys.
- N. Grabar and S. Berland. 2001. Construire un corpus web pour l'acquisition terminologique. *4e rencontres Terminologie et Intelligence Artificielle (TIA 2001)*, Nancy:44–54.
- N. Grabar and T. Hamon. 2004. Les relations dans les terminologies structurées : de la théorie à la pratique. *Revue d'Intelligence Artificielle (RIA)*, 18-1:57–85.
- M. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. *15th International Conference on Computational Linguistics (COLING 1992)*, Nantes:539–545.
- R. Martin. 1983. *Pour une logique du sens*. Paris, PUF.
- I. Meyer. 2001. Extracting knowledge-rich contexts for terminography. In D. Bourigault, editor, *Recent advances in Computational Terminology*, pages 279–302. John Benjamins Publishing Company, Philadelphia, PA.
- S. Muresan and J. L. Klavans. 2002. A method for automatically building and evaluating dictionary resources. *the language Resources and Evaluation Conference (LREC 2002)*, Las Palmas, Spain:231–234.
- J. Pearson. 1996. The expression of definitions in specialised texts: a corpus-based analysis. In M. Gellerstam, J. Järborg, S. G. Malmgren, K. Norén, L. Rogström, and C. Pappmehl, editors, *7th International Congress on Lexicography (EURALEX'96)*, pages 817–824. Göteborg University, Göteborg, Sweden.
- J. Rebeyrolle. 2000. *Forme et fonction de la définition en discours*. Thèse de doctorat, Université de Toulouse II - Le Mirail.
- J. C. Sager. 2001. *Essays on Definition*. John Benjamins, Amsterdam.
- L. Trimble. 1985. *English for Science and Technology: A Discourse Approach*. Cambridge University Press, Cambridge.