

## Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionaries

SIMONETTA MONTEMAGNI

Dipartimento di Linguistica - Università di Pisa  
Via Santa Maria 36 - 51600 Pisa Italy  
e-mail: GRAMMAR@ICNUCEVM

LUCY VANDERWENDE

Microsoft Corp, Research Division  
Redmond, WA 98052  
e-mail: LUCYV@MICROSOFT.COM

### 1. Introduction

As the research on extracting semantic information from on-line dictionaries proceeds, most progress has been made in the area of extracting the *genus* terms. Two methods are being used -- pattern matching at the string level and at the structural analysis level -- both of which seem to yield equally promising results.

Little theoretical work, however, is being done to determine the set of possible *differentiae* to be identified, and therefore also the set of possible semantic relations that can be extracted from them. In fact, Wilks remarks that as far as identifying the *differentiae* and organizing that information into a list of properties is concerned, "such demands are beyond the abilities of the best current extraction techniques" (Wilks et al. 1989, p.227). However, the current state of the art in computational linguistics demands that semantic information beyond *genus* terms be available now, on a large scale, to push forward the current theories, whether that is knowledge-based parsing or parsing first with a syntactic component, followed by a semantic component.

In this paper, we will focus on analyzing the definitions not for the *genus* terms, but for the semantic relations that can be extracted from the *differentiae* (Calzolari 1984). Although many have accepted the use of syntactic analyses for this purpose for some time now (for example Jensen and Binot 1987, Klavans 1990, Ravin 1990, and Vanderwende 1990, all of which use the PLNLP English Parser to provide the structural information), many others still do not. We will demonstrate with examples why *only* patterns based on syntactic information (henceforth, structural patterns) provide reliable semantic relations for the *differentiae*. Patterns that match definition text at the string level (henceforth, string patterns) are conceivable, but cannot capture the variations in the *differentiae* as easily as structural patterns. In addition, although it is possible to parse the definition texts using a grammar designed for one dictionary (e.g. a grammar of "Longmanese," see Alshawi 1989), we have found that a general, broad-coverage grammar of English or of Italian provides a level of analysis that is as good as, and possibly superior to, a dictionary-

specific grammar<sup>1</sup>. In addition, there is no extra effort required to apply a broad-coverage text parser to the definitions of more than one dictionary, as we found for the *Longman Dictionary of Contemporary English* (henceforth, LDOCE) and *Webster's 7th New Collegiate Dictionary* (henceforth, W7) for English, and for *Il Nuovo Dizionario Garzanti* (henceforth, Garzanti) and *Italian DMI Database* (henceforth, DMI) for Italian.

The result of analyzing the *differentiae* of the definitions is presented in the form of a semantic frame; there is one semantic frame for each word sense of the entry. The contents of the frame will be any number of semantic relations (including the *genus* term) with, as values, the word(s) extracted from the definition text. Except for a commitment to the theoretical notion that a word has distinguishable senses, the semantic frames are intended to be theory-independent. The semantic frames presented in this paper correspond to a description of the semantic frames produced by the *lexicon-producer* (Wilks, op. cit., p. 217-220) and so can be the input to a knowledge-based parser. Also, these semantic frames represent the appropriate level of semantic information that is needed by a semantic component that has the task of resolving the ambiguities remaining after a syntactic component has assigned an initial analysis (see Jensen & Binot 1987, Vanderwende 1990). More generally, the result of this acquisition process is the construction of a Lexical Knowledge Base to be used as a component for any NLP system.

### 2. Semantic Relations

The semantic relations that are needed to provide a semantically-motivated analysis of the input text have not yet been enumerated by anyone. It is possible that this is due to the absence of information on a large scale that can be used to test any hypothesis of a necessary and sufficient set of semantic relations. Semantic relations associate a particular word sense with word(s) extracted automatically from the dictionary, and those words may be further specified by additional relations. The values of the semantic

<sup>1</sup>For example, the grammar for English was used, without modification, to parse over 4000 noun definitions. With a parser that forces an NP analysis, over 75% of these definitions parsed as full NPs. These are very good results, especially since many of the remaining 25% do not form complete NPs and so were parsed correctly.

relations therefore have more content than binary features and are not abstract semantic primitives, but rather representations of the implicit links to other semantic frames.

An example of a semantic relation that can be identified in the differentiae is LOCATION-OF. The definition of 'market' (LDOCE n,1) is expressed as follows:

"a building, square, or open place where people meet to buy and sell goods, esp. food, or sometimes animals."

As we will show later, it is possible from the structural description of this definition to extract the following values for the semantic relation LOCATION-OF:

```

MARKET
LOCATION-OF MEET
(HAS-SUBJECT 'PEOPLE')
BUY
(HAS-OBJECT 'GOODS,'
'FOOD','ANIMALS')
SELL
(HAS-OBJECT 'GOODS,'
'FOOD','ANIMALS')
  
```

Figure 1. Semantic frame for the definition of the noun "market."

According to this semantic frame, the verbs "meet," "buy," and "sell" are related as LOCATION-OF to the noun "market." Although the words extracted from the definitions are not disambiguated themselves according to their senses, as much information as possible is included in the semantic frame as the definition being analyzed provides. In this example, the word "meet" is further specified by a semantic relation HAS-SUBJECT that has "people" as its value. Also, since the verbs "buy" and "sell" are conjoined, both verbs have a HAS-OBJECT relation with all the syntactic objects identified in the analysis, namely "goods", "food" and "animals."

Semantic information of this type is necessary, for example, in order to automatically interpret noun compounds. Given the (partial) semantic frame above for "market" and given that "vegetable" has a purpose relation to "food" (information also automatically derived by applying structural patterns to the definition text), the noun compound "vegetable market" is interpreted automatically as:

"Market is a location for the purpose of buying and/or selling vegetables." (see Vanderwende 1992)

Examples of other semantic relations that were required to interpret noun compounds are:

SUBJECT-OF, OBJECT-OF, FOOD, MATERIAL, TIME, HUMAN, IS-FOR, LOCATION-NOUN, MADE-OF, CAUSED-BY, CAUSES, MEASURE, and MEANS.

### 3. Structural Patterns

The acquisition of semantic relations from on-line dictionaries proceeds by applying patterns to the structural descriptions of the definitions and example sentences. The patterns embody knowledge of which relations contain recurring elements and constructions convey in the context of the dictionary. For instance, the relation PURPOSE is conveyed in Italian by the phrases: "con lo/allo scopo di," "al fine di," "per," "usato per," "atto a," "che serve a," and "utile a" followed by a noun phrase or an infinitival clause. In English, this same relationship is conveyed by quite similar phrases, also followed by a noun phrase, present participle, or infinitival clause: "for (the) purpose(s) of," "for," "used for," "intended for," and past participle followed by "to."

After locating the pattern within the definition, the true extraction process consists in identifying the values to be associated with the semantic relation detected. Typically the values of the semantic relations are the heads of the pattern itself or of the complement(s) in terms of structural patterns, or the next content word(s) in terms of string patterns. However, extracting even more specific information from the differentiae, for example that the verb "meet" has "people" as its subject when it is the LOCATION-OF "market", also involves the identification of functional arguments of verbs and in the case of nouns, identification of adjectives and "with" complements.

A simple example of a structural pattern is the pattern that extracts the semantic relation PURPOSE from the parsed definition text. The pattern can be paraphrased (in part) as:

if the verb "used" is post-modified by a PP with the preposition "for," then extract the head(s) of that PP and return those as the value of the PURPOSE relation. If the PP has a verb as its head and an OBJECT attribute, return the head(s) of the OBJECT as the values of a HAS-OBJECT relation; and if it has a SUBJECT attribute, return the head(s) of the SUBJECT as the values of a HAS-SUBJECT relation.

Consider the relevant section of the parsed definition of 'cellar' (LDOCE n,1):

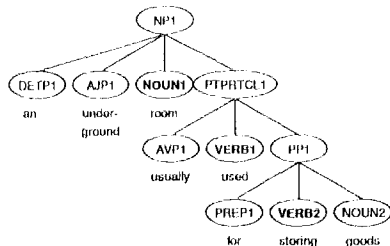


Figure 2. Parse tree for the definition of the noun "cellar."

The parse tree shown above<sup>2</sup> is but one representation of the structural description of this definition. Below is an excerpt of the record structure containing the functional information for tree node PP1 above:

NODE	'PP1'	
PRMODS	PREP1	"for"
HEAD	VERB2	"storing"
PSMODS	NOUN2	"goods"
PRP	PREP1	"for"
OBJECT	NOUN2	"goods"

Figure 3. Functional information for the prepositional phrase in Figure 2.

Following the structural pattern for PURPOSE, we see in Figure 2 that the VERB1, "used", is post-modified by a PP with the preposition "for" and so the base form of the PP head, VERB2 ("store"),<sup>3</sup> is extracted as the value of the PURPOSE relation associated with "cellar". In addition, an OBJECT has also been identified in the structural description, namely NOUN2, and so its head "goods" (in this case, the noun itself) is the value of the HAS-OBJECT of "store". The result of this pattern will be the partial semantic frame for "cellar":

CELLAR	
PURPOSE	'STORE'
	(HAS-OBJECT 'GOODS')

Figure 4. Partial semantic frame for "cellar."

#### 4. Inadequacy of String Patterns

Some patterns to identify semantic relations are relatively trivial and can be handled by string patterns. For example, no matter where the string is found in the definition text, "for (the) purpose(s) of" as well as "con lo/allo scopo di" always indicates a PURPOSE relationship between the definiendum and the head of the phrase (noun or verb) following "of/di". Markowitz et al. also discuss patterns at the string level, based on *defining formulae*, which extract such features as stative or active for adjectives, or member-set relations for nouns. These are adequate because the patterns described are generally all found at or close to the beginning of the definition text. But the most interesting patterns that identify the differentiae and the (possibly embedded) semantic relations expressed therein rely on

<sup>2</sup>The parse trees in this paper are altered representations isomorphic to actual machine output which IBM ASD has not allowed us to reproduce. Heads of constituents are directly below their parent node and the nodename is in bold.

<sup>3</sup>PPs are analyzed with a preposition premodifier and a nominal as the head.

complex structural information, information which cannot be expressed adequately in string patterns.

The following addition makes the pattern for extracting the PURPOSE relation, paraphrased in the previous section, more complete:

if the PP with "for" is not a post-modifier of a verb "used", then a PURPOSE relation between the definiendum and the head(s) of the PP can be hypothesized if the nearest noun that the PP post-modifies is the genus term.<sup>4</sup>

Consider the syntactic analysis of the relevant portion of text in the definition of "laboratory" (W7 n,1) shown below in Figure 5. Since PP2 and PP4 are coordinated, the structural relation to the rest of the analysis will be tested for the conjoined constituent, PP1. The nearest noun phrase that PP1 post-modifies is NP1, the head of which, NOUN1, is indeed the genus term (also identified automatically by structural patterns applying to this analysis.) Thus, part of the semantic frame for Sense 1 of "laboratory" will be:

LABORATORY  
PURPOSE 'STUDY,' 'TESTING,' 'ANALYSIS'

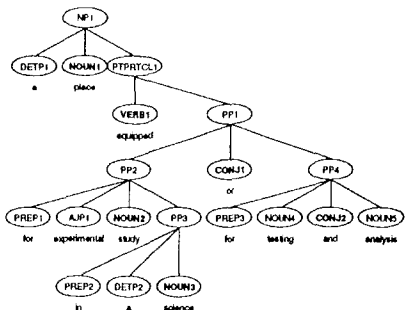


Figure 5. Semantic frame for "laboratory" and the parse from which it was derived.

Now consider the syntactic analysis of the relevant portion of text in the definition of "council" (LDOCE n): "a group of people appointed or elected to make laws, rules, or decisions, for a town, church, etc., or to give advice":

<sup>4</sup>Currently, for English, an abstract relation IS-FOR is extracted which will satisfy any searches for a PURPOSE relation.

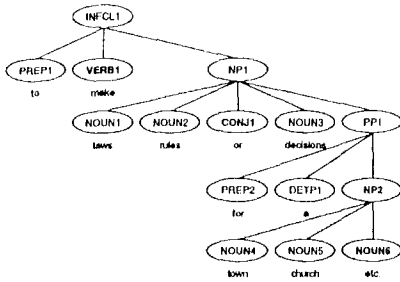


Figure 6. "for"-PP that does not create a PURPOSE relationship.

The nearest noun phrase that PP1 post-modifies is NP1, which is a coordinated construction. None of the heads of NP1, "laws", "rules" or "decisions" can be identified as the genus term, and so the pattern does not succeed in extracting a PURPOSE relation from this definition.

In order to write a string pattern that would correctly identify the semantic relations above, the pattern would have to identify conjoined heads and apply some measure of distance from the genus while counting conjoined phrases as single units. In addition, string patterns would also have to skip parentheses, identify functional arguments, and abstract from the surface realizations of the pattern, e.g. pre- and post-modification (similar observations are made in Klavans 1990). Even if the language of dictionary definitions is characterized by its regularity, variations of the defining formulae exist. These restrictions seem to be far too complex at the string level, while writing the pattern at the level of syntactic analysis describes the dependency in an intuitive manner, namely in terms of heads and modifiers.

The inadequacy of string patterns is not only evident when extracting the semantic relations directly related with the definiendum, but also when extracting those relations that show further specifications. In particular, the HAS-SUBJECT and HAS-OBJECT relations cannot possibly be extracted reliably without structural information. Wider syntactic context is also required to correctly extract the semantic features such as COLOR, SHAPE, TASTE, and SMELL not only as features of the definiendum, but also as further specifications of the words extracted as the values of semantic relations.

The structural pattern that extracts semantic features such as COLOR and TASTE would seem to be trivial: modifying adjectives or nouns that express these properties. The attachment of these modifiers, however, can be established only on the basis of syntactic information (and sometimes syntax is not enough). And only those modifiers should be extracted that relate to the definiendum or those that relate to some other word within the definition which stands in some semantic relation (for instance HAS-PART, MADE-

OF, and so forth) with the definiendum. In the latter case the information extracted still has an indirect link with the lemma being defined, but it is not expected to be interpreted as a semantic feature of the definiendum itself.

Consider these examples from the Garzanti dictionary (followed by their English glosses):

- acagiù: "albero tropicale dai frutti saporiti."  
(mahogany tree: tropical tree with tasty fruits)
- alchechengi: "pianta erbacea con bacche di color arancio racchiuse in un involucro membranaceo, commestibile."  
(winter cherry: herbaceous plant with orange berries, contained in a membranaceous covering, edible).

The TASTE and the COLOR features should not be extracted as semantic features of the definiendum. In the case of "acagiù," this is clear due to the lack of agreement between "albero" (tree) and "saporiti" (tasty); the adjective cannot modify the head noun/genus term because they do not agree in number. "Saporito", however, is the value of the semantic feature TASTE of "frutto" (fruit), which is in turn the value of the HAS-PART relation of the definiendum, also extracted by means of a structural pattern from the definition text. The semantic frame for "acagiù" is shown in Figure 7:

ACAGIÙ			
IS-A	ALBERO		
HAS-PART	FRUTTO		
	(TASTE)	SAPORITO	

Figure 7. Semantic frame for the definition of "acagiù."

In the case of "alchechengi," the PP "di color arancio" ("of orange color") does not contribute a COLOR feature to the definiendum since it cannot modify the head/genus "pianta" ("plant") given its embedded position within the syntactic structure:

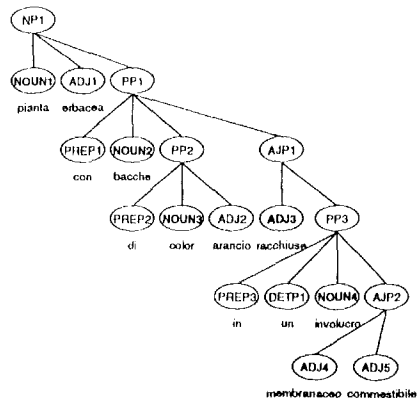


Figure 8. Parse tree for the definition of "alchechengi."

If we consider the structural description of the definition for "alcechengi," we can see clearly that the embedding of PP2 within the syntactic structure, followed by another modifier of "bacche," AJP1, makes it impossible for PP2, "di color arancio," to modify the head noun "pianta", and so the semantic feature COLOR "arancio" is extracted for "bacche", which is in a PART-OF relation with the *definicendum*.

Syntactic information is not always sufficient for resolving the correct assignment of semantic features. Consider the DMI definition for "agnolotto" (a kind of ravioli):

**agnolotto:** "involucro di pasta all'uovo rotondo o rettangolare."  
(**ravioli:** round or rectangular covering of egg pastry)

The attachment of the adjectival phrase "rotondo o rettangolare" is ambiguous and cannot be determined on the basis of syntactic information, but only based on semantic information; the correct analysis would read a "round or rectangular covering" and not "a round or rectangular egg." Despite this syntactic ambiguity, the range in ambiguity for extracting semantic relations and features is quite reduced if we start from syntactic structures instead of from simple strings.

#### 5. Why a general text parser is sufficient

There are two reasons why a general text parser is essential for providing the syntactic analyses. First, of the four dictionaries that have been explored in this research, Garzanti and DMI (for Italian) and LDOCE and W7 (for English), only LDOCE attempts to use a restricted vocabulary in the definition texts. Therefore, the scope of the vocabulary is the same as unrestricted text. Moreover, the language used in dictionaries cannot appropriately be called a specialized language given that it does not operate in a specialized domain. Second, at the syntactic level, the variety of constructions can be compared to that of textual corpora. The regularity of the language used within dictionary definitions, lexically and syntactically constrained, lies in the frequent occurrence of lexical and syntactic patterns to express particular conceptual categories or semantic relations. This regularity, which is crucial with respect to the extraction of semantic information, can be considered almost irrelevant from the point of view of parsing because of the variety of lexical choices and phrasal constructions used to express the patterns. A parser, therefore, is faced with the same range of problems in analyzing ordinary texts as in dictionary definitions and so the use of a general purpose grammar is a fundamental choice in the definition of our research framework.

One of the main disadvantages ascribed to using general text parsers is the ambiguity still remaining at the end of the syntactic analysis. It has often been observed that descriptions associated with syntactically ambiguous constructions in free text can be disambiguated in the

context of dictionary definitions. For example, within our system the default strategy in free text is to attach a prepositional phrase to the nearest available head and to keep track of the alternative possible attachment sites. In the context of dictionary definitions, the choice resulting from such a default strategy can often be overridden on the basis of lexical and/or syntactic conditions which disambiguate the potential ambiguity; for instance, with regard to the PP attachment case, there is a class of genus terms (such as "atto," act; "effetto," effect; "processo," process) that, together with given structural conditions, make the attachment decision possible.

Also, while functional assignment may be ambiguous in Italian in some cases (Chanod et al. 1991), we can assume that constructions used within dictionary definitions and example sentences are always unmarked, and consequently that the ambiguity derived from taking into account also marked orders of sentence constituents (such as Subject-Object-Verb, Object-Verb-Subject and so forth) is very unlikely to occur in the dictionary text.

Rather than taking these observations as justification for building a dictionary specific parser, we use first a broad coverage parser, followed by a post-processor which tailors the output of the parser based on the differences observed between dictionary text and general text. As it turns out, the size of the post-processor is negligible compared to the size of the grammar. This supports our claim that the variety of syntactic constructions in dictionary text is comparable to that of textual corpora. If dictionary text were substantially different from general text, we would have had to write more rules in the post-processor and it would have to be bigger than it in fact is. The structural patterns for the extraction of semantic information naturally operate on the result of the post-processor (see Montemagni 1992).

Two kinds of refinements have been devised in order to achieve more appropriate results with respect to the language used within dictionaries:

- (1) rule out ambiguity in the attachment of modifiers or in the assignment of functional roles which is not applicable in the context of dictionary definitions;
- (2) handle parses that are incomplete due to either dictionary specific constructions not occurring in free texts, or, more generally, to gaps in the lexical or grammatical knowledge of the system.

While the first refinement operates on a complete analysis but aims to reduce the high degree of ambiguity typical of free text by exploiting peculiarities of dictionary language, the second refinement concerns the robustness of the system in the absence of a complete parse.

For an example of refining the parse in order to reduce the ambiguity, consider the Garzanti definition (n.1) of "computazione" (computation) defined as "atto, effetto del computare" (the act or result of computing). The first

structural description below shows the NP parse for general text. This default analysis shows PP1 "del computare" to be attached to the closest available head, NOUN2 "effetto", while the alternative attachment site is marked with a question mark. The second parse below shows the resolution of the PP attachment ambiguity; PP1 now modifies the coordinated nouns NOUN1 and NOUN2.

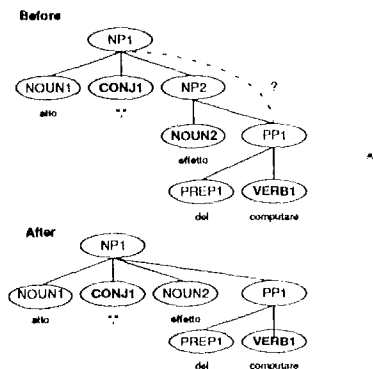


Figure 9. Resolution of PP attachment ambiguity.

This refinement is made when a prepositional phrase or an infinitival clause post-modifies coordinated head nouns that are the top nodes of the syntactic analysis. This is the typical pattern of the definitions of deverbal nouns; the PP indicates which verb the definiendum is derived from. The lexical and syntactic conditions which make the disambiguation possible are defined in the post-processor to the general text analysis.

The solution to a robust phrasal analysis while parsing dictionary text with a general grammar can be seen and faced from two different perspectives. The first perspective is dictionary specific and deals with incomplete parses due to input which would be considered ungrammatical outside of the context of dictionary definitions. The second perspective copes with incomplete knowledge of language use by exploiting the general technique of *fitted parsing* provided by the system for handling ill-formed input (Jensen et al. 1983).

Dictionary definitions are quite often formulated as condensed fragments of real texts, with elided elements which make the definition syntactically ill-formed and interpretable only by reference to a wider context. This is the case with noun definitions consisting of a noun phrase pre-modified by a prepositional phrase, where the latter specifies the usage domain of the word sense expressed by the former. The general grammar is unable to produce an NP node covering the whole input string given that the sequence PP-NP does not freely occur within ordinary texts. It is the refinement stage that should reshape the analysis and restore it as regular input on the basis of specialized

dictionary use. The analysis below of the Garzanti definition (n,1,2) "nettare" (nectar), defined as "nella mitologia classica, la bevanda degli dei" (in classical mythology, the drink of gods) exemplifies this kind of refinement.

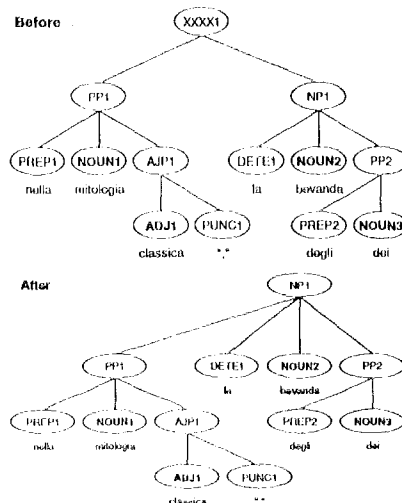


Figure 10. Refinement of fitted parse into NP.

The first of the two parses above has been generated by the general grammar; the XXXX1 label at the top node shows that the parse is incomplete. The second one has been rebuilt during the refinement stage; the XXXX1 has been replaced by the proper label NP1. In this case, knowledge of dictionary peculiarities resolves the initial partial parse and converts it into a complete and successful analysis.

Not all incomplete parses can be so easily restructured. Some are due to gaps in the system with respect to lexical as well as phrase construction knowledge. Those cases are handled by facilities in the fitting procedure, provided by the system to cope with unrestricted input. When the grammar is unable to produce a complete analysis, then a reasonable approximate but incomplete structure is assigned to the input.

Such a rough parse can still be used as input for further processing stages and for the extraction procedure itself. By allowing structural patterns to apply to incomplete parses as well, the automatic extraction of semantic information is not threatened. There is, however, a difference in the extraction procedure applied to complete (i.e. computed by the general grammar or restored during the post-processing stage) and incomplete analyses. While structural patterns are used to extract semantic information from definitions and example sentences successfully parsed, partial structural patterns and string patterns are combined when handling incomplete parses. By differentiating the

extraction procedure for the two kinds of results, the procedure becomes robust and overcomes the variability of parsing performances.

Finally, a brief account of the parsing performance of the Italian grammar for a corpus of 1000 definitions. The general Italian grammar provided complete parses for about 65-70% of these definitions. An improvement comprising 10-15% of the total was achieved during the refinement stage. For the unresolved incomplete parses, approximately 15-20%, a different extraction procedure, based on a combination of partial structural patterns and string patterns as described above, has been hypothesized. Even if this procedure is at an early development stage, it is possible to evaluate the first results. Because of the robust strategy, the extraction procedure can be applied to the entire corpus of definitions, without the worry that incomplete parses would affect the extraction of semantic information. Some information is extracted in any case; in the worst case the information is not very deep or detailed (at least the genus term is extracted). The results can be differentiated by degree of detail, but the extraction procedure never fails to produce some results.

## 6. Conclusion

Viewed ideally, the choice between structural patterns and string patterns is obviously in favor of structural patterns, because they are more suitable for achieving accuracy in the extraction of semantic information from dictionaries. Controversy rises only when considering the reliability of parsing the dictionary definitions themselves. In this paper, we show the feasibility of applying structural patterns to parsed definitions in order to extract semantic information from dictionaries, with the goal of deriving and making explicit the basic general knowledge implicitly stored in any standard printed dictionary. Structural patterns, much more than string patterns, provide the rich semantic information that makes the lexicon a relational network expanding in n-dimensions. Not only semantic features or relations directly related to the definiendum are extracted, but also further specifications of the words extracted as values of semantic features or relations.

We have also described a robust procedure for extracting this semantic information. The syntactic analysis of the definition text provided by a general text parser is evaluated automatically and, if necessary, a post-process applies to refine the parse given the context of a dictionary. The results of the structural patterns are differentiated according to the success of the parse. In this way, the use of a grammar improves the quality and the reliability of the semantic information extracted.

## References

- Alshawi, H. 1989. "Analysing the Dictionary Definitions" in Boguraev & Briscoe, eds., *Computational Lexicography for Natural Language Processing*, Longman, London, pp. 153-170.
- Boguraev, B., and T. Briscoe, eds. 1989. *Computational Lexicography for Natural Language Processing*, Longman, London.
- Calzolari, N. 1984. "Detecting Patterns in a Lexical Data Base" in *Proceedings of the 10th International Conference on Computational Linguistics*, Stanford, CA, pp.170-173.
- Chanod, J.P., B. Harriehausen and S. Montemagni. 1991. "A two-stage algorithm to parse multi-lingual argument structures" in *Proceedings of the International Conference on Current Issues in Computational Linguistics*, University Sains Malaysia, Penang, June 21-27 1991.
- Jensen, K., and J.-L. Binot. 1987. "Disambiguating Prepositional Phrase Attachments by Using On-Line Dictionary Definitions" in *Computational Linguistics* 13.3-4.251-60.
- Jensen, K., G.E. Heidorn, L.A. Miller, and Y. Ravin. 1983. "Parse Fitting and Prose Fixing" in *American Journal of Computational Linguistics* 9.3-4.123-36.
- Klavans, J., M. Chodorow, and N. Wacholder. 1990. "From Dictionary to Knowledge Base via Taxonomy" in *Electronic Text Research*, University of Waterloo, Centre for the New OED and Text Research, Waterloo, Canada.
- Markowitz, J., T. Ahlswede and M. Evens. 1986. "Semantically Significant Patterns in Dictionary Definitions" in *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, June 10-13 1986, pp. 112-119.
- Montemagni, S. 1992. "Tailoring a broad coverage grammar for the analysis of dictionary definitions" to appear in *Proceedings of EURALEX-92*, August 4-9, Tampere.
- Ravin, Y. 1990. "Disambiguating and Interpreting Verb Definitions" in *Proceedings of the 28th Annual ACL Conference*, June 6-9. Pittsburgh, pp. 260-267.
- Vanderwende, L. 1990. "Using an On-line Dictionary to Disambiguate Verbal Phrase Attachment" in *Proceedings of the 2nd IBM Conference on NLP*, March 13-15, La Defense, Paris.
- Vanderwende, L. 1992. "Understanding Noun Compounds Using Semantic Information Extracted from On-Line Dictionaries." Dissertation in preparation, Georgetown University, Washington DC.
- Wilks, Y., D. Fass, C. Guo, J. McDonald, T. Plate, and B. Slator. 1989. "A Tractable Machine Dictionary as a Resource for Computational Semantics" in Boguraev & Briscoe, eds., *Computational Lexicography for Natural Language Processing*, Longman, London, pp. 193-228..

## Note

For the specific concerns of the Italian Academy, Vanderwende is responsible for sections 1-3 and the English part of section 4, and Montemagni is responsible for the Italian part of section 4 and sections 5-6.