Using Dictionary Grammar Codes to Resolve Attachment Ambiguities

Thierry Fontenelle, Alma Kharrat

Microsoft Corporation
Natural Language Group
One Microsoft Way
Redmond, WA 98052
USA
thierryf@microsoft.com, almakhar@microsoft.com

Abstract
The resolution of prepositional phrase (PP) attachment has always been a nightmare for computational linguists. This type of attachment is part of a more general problem linked to the resolution of syntactic ambiguities. Several methods have been suggested, ranging from the use of statistical data to select the statistically most frequent constructions, to the use of machine-readable dictionaries from which the semantic information necessary to resolve ambiguities is extracted [Jensen & Binot 1987, Richardson *et al.* 1993, 1998, Richardson 1997, Vanderwende 1995]. The latter work has led to the construction of large knowledge bases in the form of networks of words linked by lexical-semantic relations extracted from dictionary definitions. In this paper, we show how grammatical codes reflecting the syntactic environment in which a lexical item can be inserted can be used by a broad-coverage parser in order to give preference to syntactic codes initiating the densest or most complex structures.

Introduction

The work which is reported on here is done in the framework of the development of a broad-coverage parser for French used in applications such as grammar checkers or command and control interfaces (i.e. computational applications that process natural language queries or commands and convert them into meaningful actions). The term 'broad-coverage parser' refers here to an automatic natural language analysis program which assigns attributes to designate the various grammatical and syntactic elements of the sentence. Such a program is not restricted to any given field or type of vocabulary, but should rather be able to process any kind of input. The grammar, based on the PLNLP English grammar described by Jensen et al. [1993], produces sentence parses for unrestricted text. It relies upon computational lexicons which drive the parsing process. The original system was developed for the English language and has been adapted to a number of other languages, viz. French, German, Spanish and Japanese. The English lexical component draws upon information extracted from the Longman Dictionary of Contemporary English (LDOCE) [Procter 1978], and the American Heritage Dictionary [AHD 1992]. The former dictionary is well-known to computational lexicographers insofar as it has been used in a very large number of research and development projects because of the wealth of syntactic, semantic and pragmatic information it contains (see [Michiels 1982; Boguraev & Briscoe 1989] and [Wilks et al. 1996] for more information on how this dictionary can be and has been used in natural language processing systems and applications).

In order to facilitate the 'localization' of the original grammars and to ease code sharing among the developers in the various language teams, it has been decided to use the LDOCE system of grammar codes for all languages, which means that all lexicographers and grammarians now use the same system in order to represent the syntactic environment of lexical items. Such codes, which are assigned at definition, i.e. at word sense level, normally consist of a letter followed by a number: the letter describes the word type (e.g. T for transitive verb, D for ditransitive verb with two different objects,.I for intransitive verb, L for linking verb, C for count noun, U for uncount/mass noun, B for adjective...) and the number describes the environment in which this item can appear:
0: need not be followed by anything;
1: followed by one or more nouns or pronouns;
2: followed by a bare infinitive;
3: followed by a TO-infinitive in English or an infinitive preceded by the semantically-depleted prepositions *à, de* or *par* in French;
4: followed by a gerund in English or a present participle in French
5: followed by a *that/que* clause.
6: followed by a WH-clause (i.e. a clause starting with *who, which, where, what, how...* in English; *si, comment, pourquoi...* in French).
The following are cases in point:

Grammar code	Explanation	Example
I0	Intransitive verb	Je dors. (= I sleep)
T1	Transitive verb followed by one Noun Phrase (NP)	Je mange une pomme. (= I eat an apple)
T5	Transitive verb followed by a QUE subordinate clause	Je sais qu'elle est malade. (= I know that she is ill)
T6	Transitive verb followed by a WH-clause	Je voudrais savoir si tu comptes venir. (= I'd like to know whether you intend to come)
X9	Verb with one object + a descriptive word or phrase	Je mets le livre sur la table. (= I put the book on the table)
D5do	Ditransitive verb followed by one (direct object) Noun Phrase + que-clause	J'ai convaincu mon chef que c'est la meilleure solution. (= I convinced my boss that it is the best solution)
C3	Count noun followed by a *de/à-* infinitive	Nous n'avons aucune raison de nous plaindre. (= We have no reason to complain)

Table 1: Grammar codes and subcategorization

The system of grammar codes designed by the LDOCE lexicographers has been found flexible enough to be used for other languages. Codes ending in [5], for example, all refer to complement clauses introduced by a standard subordinate conjunction, *that* in English, *daß/dass* in German or *que* in French and Spanish. Some codes are obviously English specific: the letter [a] which accompanies some codes in English, such as D5a or T5a, and means that the conjunction *that* can be omitted (I said [*that*] he would do it), is not used in the other languages because such omission is not possible in German, Spanish or French. Similarly, other language-specific codes were introduced to account for a subdivision of pronominal verbs in Romance languages, for instance. It must also be stressed that the distribution of grammar codes may vary from one language to the other: codes ending in [4] (+ gerund) are assigned much more frequently in English that their counterparts in French, which correspond to a present participle construction. But apart from these necessary adaptations and refinements or differences in coverage, the original system of codes has been found useful to account for the major subcategorization patterns. In an environment in which several languages are being developed simultaneously and use the same architecture and the same structure for the lexical database, this policy has proved most useful to avoid a proliferation of coding schemes and to facilitate the collaboration between the lexicographers and grammarians of all the languages covered by the system.

A lexicon-driven parser

The resolution of ambiguities is the notorious stumbling block in any NLP system. The strategy used by the parsers for all the languages we deal with is to attach pre- and post-modifiers to the closest possible head. This is arbitrary, of course, and only takes place at the shallow parsing level, i.e. the level corresponding to the first pass of the syntactic analysis, leaving certain kinds of reattachment for the next level. This means that, in the first stage of the analysis, the following two sentences, excerpted from Jensen & Binot [1987:252-253], get the same parse in which the *with* prepositional phrase is attached to the noun phrase *the fish*:

(1) I eat a fish with a fork.
(2) I eat a fish with bones.

The sentence corresponding to (1) in French (*Je mange un poisson avec une fourchette*) gets the following (wrong) analysis, in which the *avec* prepositional phrase (tree node PP1) is attached to the closest possible head (*poisson*).

```
DECL1     NP1       PRON1*    "Je"
          VERB1*    "mange"   (Subject NP1  Object NP2)
          NP2       DETP1     ADJ1*     "un"
                    NOUN1*    "poisson"
                    PP1       PREP1*    "avec"
                              NP3       DETP2     ADJ2*     "une"
                                        NOUN2*    "fourchette"
          CHAR1     "."
```

Figure 1: Parse tree for "Je mange un poisson avec une fourchette."

Jensen & Binot suggest a method based on the analysis of dictionary definitions aimed at identifying a number of lexical-semantic relations. By discovering that bones are **parts** of a fish while a fork is an **instrument** used to eat (amongst others things), it is possible to revise the preliminary analyses and make sure that the PP *with a fork / avec une fourchette* gets attached at a higher level. Our system also makes use of statistics which reflect the probability of finding specific types of environment (in terms of phrase levels, subcategorization patterns) linked to a given headword. Unlike standard cooccurrence statistics such as mutual information, which compute the probability of finding two lexical items within a given window of words, the 'statistical goodness measure' used by our system [Yedwab & Weise 1999] computes the probabilities of parse trees as the product of the probabilities of its nodes. This statistical technique is seen here as another way of giving preference to certain attachments or to certain syntactic constructions. In the following section, we describe an alternative method for dealing with such ambiguities.

Dealing with multiple grammar codes

The ambiguities described in the previous section are usually solved at a purely semantic level and require access to semantic information to be encoded in large semantic networks such as WordNet or EuroWordNet [Fellbaum 1998; Vossen 1998], MindNet [Richardson et al. 1998; Dolan et al. 2000] or the Collins-Robert lexical-semantic database described by Fontenelle [1997]. Other types of attachment problems may frequently arise due to multiple (and sometimes conflicting) grammar codes in the lexicon, however. These multiple syntactic codes can give rise to multiple readings and parses and it is essential that the NLP system be able to separate the wheat from the chaff.
Consider the following sentence, which is analyzed correctly:
(3) **Je lui donne ce livre** (= I give him this book)

```
DECL1  NP1      PRON1*    "Je"
       NP2      PRON2*    "lui"
       VERB1*   "donne"   (Subject NP1  Indobj NP2  Object NP3)
       NP3      DETP1     ADJ1*     "ce"
                NOUN1*    "livre"
       CHAR1    "."
```

Figure 2: Parse tree for "Je lui donne ce livre"

The correct analysis is based upon the assignment of D1 in the entry for *donner* (D1 = ditransitive verb with NP + indirect object PP). The NP2 corresponding to the indirect object (*lui*) is correctly attached, which is not the case in the following similar sentence:
(4) **Je donne ce livre à mon frère** (= I give this book to my brother)

```
DECL1    NP1      PRON1*    "Je"
         VERB1*   "donne"   (Subject NP1  Object NP2)
         NP2      DETP1     ADJ1*     "ce"
                  NOUN1*    "livre"
                  PP1       PREP1*    "à"
                            NP3       DETP2     ADJ2*     "mon"
                                      NOUN2*    "frère"
         CHAR1    "."
```

Figure 3: Parse tree for "Je donne ce livre à mon frère."

The problem is that the analyzer, by design, does not produce the same structure as for (3), even though the only difference is that the pronoun *lui* is now replaced by an indirect object

(*à mon frère*). The aim here is to reduce the search space for performance reasons. The PP in (4) is wrongly attached to the closest possible head, i.e. *livre*. This analysis is made possible because the verb *donner* is assigned both the D1 code (ditransitive with direct and indirect objects) and the T1 code (monotransitive verb with one object NP, as in *Il aime donner des conseils* – He likes giving advice).
Expanding on suggestions made by Michiels [1982:227ff], we start from the assumption that the analyzer should know which code to prefer. Since the system comes up with a variety of parses, we established a priority order to make sure that the parser gives priority to the codes which initiate the longest/densest parses. In (4) above, it is clear that the D1 reading should be preferred, even if a T1 interpretation is also possible from a purely syntactic point of view.
Establishing a priority order among the competing parses means that a hierarchy of grammar codes should be devised. Codes initiating the longest parsing procedure should be given priority. D1 initiating a more complex representation than T1 ("Subject_NP1 Indobj_NP2 Object_NP3" is more complex than "Subject_NP1 Object_NP2"), the reading it gives rise to should be retained.
The hierarchy, ranging from more complex to simple codes, is as follows:

Priority	**Codes**	**Syntactic structures**
To be tried first	D5 or D6	Verb + *que* (that) clause or WH-clause (*si, comment, où...*)
2nd	Other codes with 5 or 6	5 = + *que* clause; 6 = + WH-clause (*si, comment, où...*) – e.g. C5: Count noun with *que* (that) clause; I6: Intransitive verb with WH-clause
3rd	D1, V2, V3, V4, V8, X1, X7, X9	codes for verbs followed by an NP followed by something else. D1: ditransitive verb; V2: verb followed by bare infinitive; V3: Verb followed by NP + prep + infinitive (*convaincre* qn de V-inf); X1: Verb with NP + adjectival object complement; X7: Verb + NP + nominal object complement; X9: Verb with one object + a descriptive word or phrase...
4th	other codes with Capital letter + 3 or 4	3 : + Infinitive; 4: + present participle; e.g. C3: Count noun + prep + V-inf (e.g. *difficulté...*)
5th	all remaining codes except I0 (e.g. T1...)	T1: transitive verb with one object NP
Bottom	I0	Intransitive verb

Table 2: Hierarchy of grammar codes

Such a hierarchy will make it possible to correctly attach the locative prepositional phrase to verbs which are assigned the X9 code, such as *poser*, *déposer*, *fourrer*, *mettre*, *plonger*. In their X9 reading, all these verbs require a descriptive word or phrase (traditionally a PP, but

a pro-form for a locative adverbial such as *ici* (here) or *là* (there) is also possible), so that the following parse can be avoided:
(5) Mets le livre sur la table. (= Put the book on the table)

```
IMPR1      VERB1*     "Mets"   (Object NP1)
           NP1        DETP1      ADJ1*      "le"
                      NOUN1*     "livre"
                      PP1        PREP1*     "sur"
                                 NP2        DETP2      ADJ2*      "la"
                                            NOUN2*     "table"
           CHAR1      "."
```

Figure 4: Parse tree for "Mets le livre sur la table."

It is clear here that the PP "sur la table" satisfies the 9 in the X9 code. On the basis of the hierarchy above, X9 has precedence over T1 and the PP can then be attached at verb level, which yields the following parse tree:

```
IMPR1      VERB1*     "Mets"   (Object NP1)
           NP1        DETP1      ADJ1*      "le"
                      NOUN1*     "livre"
           PP1        PREP1*     "sur"
                      NP2        DETP2      ADJ2*      "la"
                                 NOUN2*     "table"
           CHAR1      "."
```

Figure 5: Revised parse tree for "Mets le livre sur la table."

Problems

It is clear that no hierarchy can be 100% reliable and natural language is (unfortunately for computational linguists) notoriously fraught with insuperable problems. The hierarchy makes it possible to give preference to the X7 reading in the following two sentences, which illustrate a pattern consisting of a Verb (*trouver* - find) followed by an NP and an object complement referring to the NP:
(6) Je trouve cette affirmation monstrueuse. (= I find this statement horrible.)
(7) Je trouve cela inadmissible. (= I find this unacceptable.)
Note that (6) is in fact ambiguous between a "small clause" (X7) reading and a purely transitive (T1) reading, while (7) is not ambiguous because the NP is realized as a pronoun (*cela*). The following sentence is equally ambiguous, but, this time, the T1 (purely transitive) reading is more likely than the X7 interpretation.
(8) J'ai trouvé ce bracelet argenté sur la route. (= I have found this silver bracelet on the road.)
The problem can be even more complex when several adjectives are found after the noun. In the following sentence, it is clear that the adjective *commerciale* collocates with *politique* and belongs to the direct object NP, while *lamentable* is the adjectival object complement which satisfies the X7 code:
(9) Je trouve votre politique commerciale lamentable. (= I find your commercial policy appalling.)

Although the concept of a hierarchy of grammar codes is obviously useful to rule out any T1 reading in the above sentence, it is also necessary to be able to identify the limits of phrases so that only *lamentable* gets recognized as an object complement:
(9') Je trouve [[votre politique commerciale] [lamentable]].
This will probably entail a sub-classification of adjectives to distinguish relational (objective) adjectives and evaluation (subjective) adjectives, starting from the hypothesis that X7 verbs are more likely to occur with an NP and evaluation adjective complements. In this respect, the distinction made by Gross & Miller [1990] between relational and descriptive adjectives is useful. Relational adjectives such as *commercial* (*relatif au commerce – related to commerce*) are normally not likely to be found in the position of attributes (* *la politique est commerciale*) and will then be directly attached to the noun they modify. Non-relational, descriptive adjectives, which are also usually gradable, are also good candidates for filling in the object complement slot. This type of distinction is all the more necessary as adjectives frequently appear after the noun they modify in French, which accounts for a greater number of potential ambiguities than in English in this respect.

Conclusion

In this paper, we have tried to show that a hierarchy of grammatical codes assigned to lexical items based on the complexity of the structures these codes initiate can be a very useful tool to disambiguate prepositional attachments and other types of modifier/argument parsing problems. Although this algorithm cannot be 100% reliable, it could be a complement to the use of statistical data or of lexical-semantic relations, which involve the development of highly complex and costly lexical resources. This does not make the latter resources less crucial, but it makes it possible to obtain useful results to be refined at the semantic analysis level.

References

[AHD 1992] *American Heritage Dictionary.* 1992. 3rd Edition. Houghton Mifflin Company.
[Boguraev & Briscoe 1989] Boguraev, B. & E. Briscoe, 1989. *Computational Lexicography for Natural Language Processing.* London and New York, Longman.
[Dolan et al. 2000] Dolan, W., L. Vanderwende & S. Richardson, 2000. Polysemy in a Broad-Coverage Natural Language Processing System, in: Y. Ravin & C. Leacock (eds) *Polysemy – Theoretical and Computational Approaches,* pp.178-204, Oxford University Press.
[Fellbaum 1998] Fellbaum, Ch. 1998. (ed.) *WordNet: An Electronic Lexical Database.* MIT Press, Cambridge, Massachusetts.
[Fontenelle 1997] Fontenelle, Th., 1997. *Turning a bilingual dictionary into a lexical-semantic database.* Max Niemeyer Verlag, Tübingen.
[Gross & Miller 1990] Gross, D. & K. Miller, 1990. Adjectives in WordNet, in: *International Journal of Lexicography,* pp.265-277, Vol.3, Number 4.
[Jensen & Binot 1987] Jensen, K. & J-L Binot, 1987. Disambiguating Prepositional Phrase Attachments by using On-Line Dictionary Definitions, in: *Computational Linguistics,* pp.251-260, Vol.13, n°3-4.
[Jensen et al. 1993] Jensen, K., G. Heidorn, & S. Richardson, 1993. (eds) *Natural Language Processing: the PLNLP Approach.* Boston, Kluwer Academic Publishers.
[Michiels 1982] Michiels, A., 1982. *Exploiting a Large Dictionary Database,* PhD Thesis, University of Liège, mimeographed.
[Procter 1978] Procter, P., 1978. (ed.) *Longman Dictionary of Contemporary English.* Harlow, Longman.
[Richardson et al. 1993] Richardson, S., L. Vanderwende & W. Dolan, 1993. Combining Dictionary-based and Example-based Methods for Natural Language Analysis, in: *Proceedings of the Fifth*

International Conference on Theoretical and Methodological Issues in Machine Translation, pp. 69-79, Kyoto, Japan.

[Richardson 1997] Richardson, S., 1997. *Determining Similarity and Inferring Relations in a Lexical Knowledge Base,* Ph.D. thesis, The City University of New York.

[Richardson et al. 1998] Richardson, S., W. Dolan, & L. Vanderwende, 1998. MindNet: acquiring and structuring semantic information from text, in: *Proceedings of COLING '98.*

[Vossen 1998] Vossen, P. (ed) 1998. *EuroWordNet: A multilingual database with lexical-semantic networks.* Kluwer Academic Publishers, Dordrecht/Boston/London.

[Vanderwende 1995] Vanderwende, L., 1995. *The Analysis of Noun Sequences using Semantic Information Extracted from On-Line Dictionaries,* PhD dissertation, Georgetown University.

[Wilks et al. 1996] Wilks, Y., B. Slator, & L. Guthrie, 1996. *Electric Words - Dictionaries, Computers and Meanings.* MIT Press, Cambridge, MA and London.

[Yedwab & Weise 1999] Yedwab, L. & D. Weise, 1999. *Microsoft's Statistical Goodness Measure – Patent Disclosure.* Internal document.

Maira Jimenez – Microsoft Machine Translation Group

- mnogi uporabljajo predlog, potem hierarhijo · hod