

THE ASSIGNMENT OF GRAMMATICAL RELATIONS IN NATURAL LANGUAGE PROCESSING

Leonardo Lesmo, Vincenzo Lombardo

Dipartimento di Informatica - Università di Torino

C.so Svizzera 185 - 10149 Torino - ITALY

e-mail: lesmo.vincenzo@di.unito.it

1. Introduction

One of the main goals of an interpreter is to map the syntactic descriptions found in the sentence into the correct roles that the elements (described by the nominals) play in the situation at hand (described by the verb). For instance, we must be able to state that in

1) The cat ate the mouse

the cat is the "eater" and the mouse is the "eaten thing". Of course, if we only talk about roles and situations we miss some significant generalizations. In

2) The boy drank the water,

if we say that the boy is the "drinker" and the water is the "drunk thing", we disregard the evident similarity of the roles of "eater" and "drinker" in the two situations. The notion of *deep case* arises as the common ground underlying a number of "apparently" different roles. Upon this notion some frameworks, that stand at the core of semantic representation and natural language processing, are built (see [Fillmore 68], [Bruce 75] and [Somers 87]).

The hard task is to devise a mapping between the surface descriptions and these deep cases. The complexity of some syntactic phenomena, like passivization, subject and object raising, long distance dependencies, has led many researchers to pose an intermediate level between the linear string of words and the case system. The concept involved is that of "grammatical relation", such as "subject", "direct object", "indirect object". It is claimed, for example, that "passivization" is universally (cross-linguistically) explained if one says that the "object" of an active sentence becomes the "subject" in the passive form, rather than by saying that the NP in the VP is moved to replace the NP in S (that is a direct mapping). In the latter case it is implicit that the particular language under examination has a Subject-Verb-Object structure (SVO), as it usually happens in configurational languages such as English. In the example

3a) Lo hanno visto gli amici di Piero

(Him have seen the friends of Piero)

3b) E' stato visto dagli amici di Piero

((He) has been seen by Piero's friends)

the passive form does not obey the law of direct mapping. The example is, however, easily accounted for by the relational theories. The passivization rule induces only changes of function: the SUBJ becomes the BY-complement and the OBJ becomes the SUBJ.

The importance of grammatical relations, taken as primitives for a universal grammar, is stated by a number of formalisms often collected under the label of Relational Grammar. The problem is to map the surface constituents into their correct roles. With languages as Italian, which stands in the middle between configurational and freely ordered languages [Stock 89] some flexibility is required to accomplish this task. One possibility is to adopt a neutral syntactic structure, open to several alternatives in the interpretation process. The head & modifier approach seems to feature this kind of neutrality, and has effectively been used for dealing with free word order languages, like the Slavonic languages [Sgall et al. 86] and Finnish [Jappinen et al. 86].

The dependency formalism we have adopted is presented in [Lesmo, Lombardo 91]. An example is reported in fig.1, and concerns the sentence:

4) La ragazza che lavora al guardaroba fu persuasa da un cliente a comprare una enciclopedia

(The girl who works at the wardrobe was persuaded by a customer to buy an encyclopedia).

The daughter nodes that stand on the left of their head precede it in the linear order of the sentence, while daughter nodes on the right follow it. The arcs that link the nodes in the dependency tree are of three types: arcs of structural and logical dependency (D&S arcs, represented by bold arrows in the figure), arcs of only structural dependency (STR arcs, simple arrows in the figure), and arcs of only logical dependency (DEP arcs, dashed arrows in the figure). D&S arcs link two words that stand in a "both structural and logical" relation. STR and DEP split these two functions of arc: an STR individuates a purely superficial

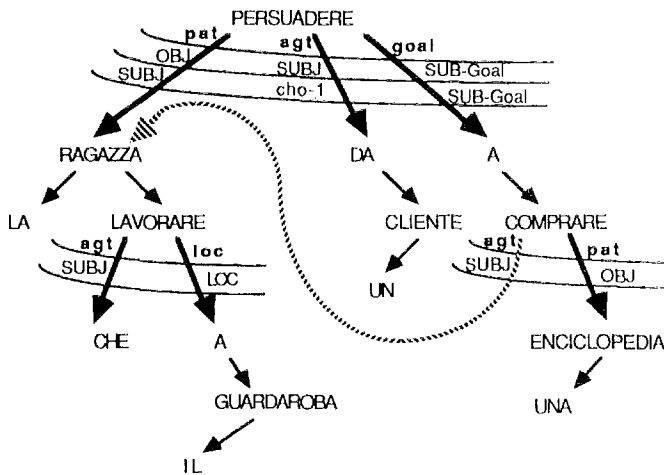


Fig.1 - An example of dependency tree. Because of space constraints, the figure already includes the grammatical relations that will be described below. The bold labels (e.g. agt, pat) refer to deep cases. The labels immediately below them refer to the initial stratum of grammatical relations. The lowest labels are the last stratum (surface relations). SUB-Goal stands for a GOAL relation expressed via a subordinate sentence. Cho-1 is a chomeur (see text), expressed in Italian via a BY (DA) complement.

attachment, DEP represents a deep dependency between two words that are structurally independent. DEP arcs enable us to represent long distance dependencies, the sharing of dependent nodes (i.e. multiple heads, see fig.1) and to represent coordinative and comparative constructions without violating the adjacency principle [Hudson 84], that applies only to STR and D&S arcs¹. An arc involving dependency (of DEP or D&S type) is labelled with the grammatical relation that exists between the two nodes that it links (the arrangement in strata is explained below).

The goal of this paper is to show that the formalism of Relational Grammar can be integrated in a useful way in a general NL interpreter, in particular if the surface structures are represented via the dependency formalism. The paper examines the problems associated with the use of RG in an interpretive (as opposed to generative) framework, where the phase of surface relation hypothesization is critical. The partial configurationality of Italian can be exploited as heuristic information aiding the interpreter in selecting the preferable initial

hypothesis. On the contrary, the RG rules governing the mapping between strata aim at confirming the hypotheses: they are applied on the basis of the lexical and morphological information associated with the verb, where the lexicon provides the first stratum and possible constraints on rule applicability.

2. The assignment of grammatical relations

We start this section by providing a short overview of the main ideas of RG. Such ideas are shared by many formalisms, but we will mostly refer to the work described in [Perlmutter 83] and [Perlmutter, Rosen 84], where it can be found a comparison with other RG formalisms.

Grammatical relations are arranged in a hierarchy and are usually referred to by numbers: 1, which is the highest, corresponds to SUBJECT, 2 to DIRECT OBJECT, 3 to INDIRECT OBJECT. The key principle of RG is the promotion of relations to higher levels in the hierarchy. The passive can be described as a *promotion* of 2 to 1 (i.e. DIR-OBJ to SUBJ), leaving the previous 1 element "unemployed". The relation "unemployed", which is technically indicated by the corresponding French word

¹ The adjacency principle intuitively states that a word B, that stands between the words A and C in the sentence, results in the same position if we project the related nodes in the dependency tree onto a line.

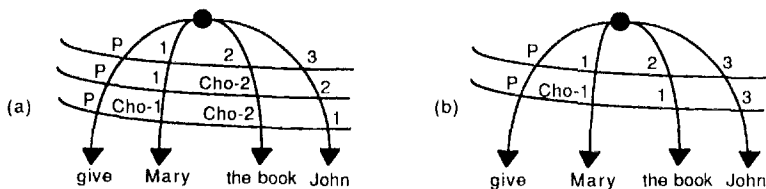


Fig.2 - The Relational Networks associated with the sentences "John was given the book by Mary" (a) and "The book was given to John by Mary" (b).

chomeur, is assigned to an element that cannot be involved in any other promotion. Consider 5a) **Mary gave the book to John**, where **Mary** is the 1-element, **the book** is the 2-element and **John** is the 3-element. If we apply the rule for passivization described above, **the book** must be promoted to the 1 relation, **Mary** becomes a chomeur, while **John** is still the 3 element. The **chomeur-1** element, i.e. a chomeur element that results from the "unemploying" of a 1-element, assumes the surface form of a by-complement in English, thus yielding

5b) **The book was given to John by Mary**

where **Mary** cannot be involved in any other promotion, because of its chomeur condition. A similar rule applies to double-accusative constructions, as shown in fig.2a.

At the same level of promotional rules we can posit the *lexical* rules, that account for the determination of grammatical relations within subordinate untensed sentences, as in 4). Such information is stored within the lexical entry of the verb that governs the subordinate clause. For example, **to promise** forces the SUBJ of the subordinate clause to be the SUBJ element of the governing clause, as we can see in

6) **Mary promised John to write him a letter,**

where the SUBJ of **write** is **Mary**, the same of **promise**. On the contrary, **to persuade** forces the SUBJ of the subordinate to be the OBJ element of the governing clause, as in 4). It must be noted that the lexical rules are related to the assignment of relations in the initial stratum, even if they are subsequently changed by promotional rules. For example, in

7) **The girl was persuaded by a customer to buy an encyclopedia,**
the girl is the element which is still shared by the two clauses, even if it is the SUBJ now.

The semantic interpretation process takes advantage of the functional analysis, i.e. the analysis in terms of grammatical functions: relational structures are easily mapped onto

logical representations, because of the resemblance between a Relational Network and a Predicate-Argument structure: the initial stratum states which are the grammatical relations (actually the elements at the sentence level) that act as arguments of the predicate identified by P.

From a computational point of view, syntactic and semantic clues must be taken into account, in order to map the grammatical relations onto the surface descriptions. The mapping is carried out incrementally, i.e. as soon as the nominal head of the complement is parsed: it is highly language-dependent, and considers features like inflectionality, configurationality and deep underlying structures. The mapping must also take into account the changes on the surface form that are induced by the rules on grammatical relations, discussed informally above².

Unfortunately the mapping raises some difficulties. The bias of the rules at the RG level is of a generative kind. Rules start from the initial stratum, the one which is closer to the deep cases (arguments of the predicate), to produce the final stratum of the surface arrangement; on the contrary, in an interpreter of language, the task is to trace what rules have been actually applied (and in what order) to the initial stratum (and the subsequent strata) in order to achieve the surface realization of the grammatical relations. Useful heuristics are devised to identify the surface clues that evidence the application of a particular rule. For example, a passivization is accompanied by a passive form of the verb.

Starting from the surface descriptions, the

² Of course, the application of such rules involves also changes in focus. Two expressions that result to be derivable from each other (3a and 3b), according to phenomena that are explained in terms of grammatical relations, are therefore not strictly equivalent, even if both of them involve the same roles to be played by the individuals in the ground sentence (or, better, in the sentence that has been claimed to be ground).

interpreter is not always able to uniquely determine the assignment of grammatical relations on the basis of syntactic features. Consider

8a) Giovanni il vino lo ha bevuto

(John the wine [it] has drunk)

8b) Il vino Giovanni lo ha bevuto

(The wine John [it] has drunk)

Only semantics allows a hearer to realize that in both cases the "drinker" is John and not the wine. Hence a flexible interaction between syntactic and semantic information (selectional restrictions) must be devised.

Since in our system the analysis is incremental, in the sense that parsing and interpretation are synchronous processes, as soon as the dependency tree is extended with a head of a substructure the semantic interpreter is triggered to interpret it: in the case of verbs, all the complements that precede the verb are interpreted when the verb is found, while each complement that follows the verb is interpreted as soon as it is attached to it.

The association of the grammatical relations with the descriptions in the sentence is accomplished by the rules at the relational level (GR rules), which are divided into three groups: the first of them deals with the initial proposal of relations based on syntactic features (Proposal Rules - PR), the second concerns the movement across the strata (Stratal Rules - SR), and the third, of a lexical kind (then Lexical Rules - LR), accounts for the sharing of relations in untensed subordinate clauses, as in 4. The verbal lexical entry contains, among other information, its initial stratum of grammatical relations. Once the verb has been found, the GR rules are triggered in order to find out the roles that are played by the elements that precede the verb in the input sentence (incremental interpretation). It is the actual input that determines which of the three groups must be applied. For example, if we have a single active sentence (without subordinate embedded clauses or passive forms), the Proposal Rules are triggered. SR and LR rules are activated only in presence of special features: a passive form (**was eaten**), for instance, activates the Passivization rule (belonging to SR), if we have the pair <SUBJ,OBJ> in the current stratum, while lexical rules are associated with verbs that govern subordinate clauses (e.g. **to persuade**). The result of the application of one or more rules is the final stratum, against which the assignment of relations guessed by the PR group is matched.

In the PR, the first feature that is taken into account is the syntactic form of the participants. SUBJ and DIR-OBJ require that the

corresponding nominals are not preceded by a preposition, and pronouns be inflected appropriately. For example, in 3a the pronoun **lo** features an accusative case, thus a DIR-OBJ. If two nominal descriptions are not inflected and, hence, they cannot be associated with a particular relation via this marking, as in

9) Il gatto mangio' il topo

(The cat ate the mouse),

the position of the nominals can be useful, since, in a partially configurational language such as Italian, grammatical relations are usually connected with the canonical positions of the SVO order: with a transitive verb, the SUBJ precedes the verb and OBJ follows it; with intransitive verbs, the position of the nominal without a preposition does not affect the grammatical relation assigned to, since it will be surely the SUBJ wherever it is. If the order too does not give an unambiguous assignment of grammatical relations, the last resources are the number agreement for the SUBJ relation and the semantic check. In situations such as

10) Le ragazze Giorgio le ha viste

(The girls Giorgio [them] has seen)

even if the nominal descriptions stand on the same side of a transitive verb, the latter agrees only with **Giorgio** in number. On the contrary, only semantics can solve a situation as 8; moreover, the semantic check can also reject an assignment made on the basis of the syntactic features that we have described. Consider, for example, the sentence

11) Un sasso calcio' il vitello

(A rock kicked the calf)

Even if the order rules assign **un sasso** the SUBJ relation and **il vitello** the OBJ, such an assignment is rejected on the semantic ground. Notwithstanding a system that works correctly cannot be based only on semantics, since a sentence like 11 sounds really strange to a native speaker, if we are not in a particular focussing situation.

3. An example

In figure 1, we can find the result of the interpretation of sentence 4. When the analysis arrives at **lavora**, in the relative sentence, its initial stratum <SUBJ> is retrieved from the lexicon. Since **che** (*who*) is a nominal without a prepositional marker, it (or better the element referred to) is the SUBJ of **lavora**, as stated by the Proposal Rules. **Lavora** has also an adjunct, **al guardaroba** (**at the wardrobe**), a non-term relation of type LOC. The structure for the nominal description **la ragazza che lavora al guardaroba** has already been built,

when the input word is the verb **persuadere**. Its lexical entry provides the parser with an initial stratum of grammatical relations that consists of: a SUBJ, an OBJ and a subordinate sentential Goal (SUB-Goal)³, i.e. a persuader, a persuadee and the persuasion. This basic assignment can be related to the deep cases of AGT, PAT and GOAL respectively. Moreover, a lexical rule is contained in the lexical entry:

*The SUBJ of the subordinate untensed clause governed by **persuadere** is the OBJ element of the governing clause.*

Since the verb is in the passive form and the current stratum features a SUBJ and an OBJ relations, the Passivization rule in the SR group is triggered, in order to find the actual arrangement of relations in the input sentence. The new stratum is <SUBJ, Cho-1, SUB-Goal>, against which the proposals made by the PR group are matched. Since the nominal description already found is not inflected and is not marked by a preposition, the positional rules suggest that, since it precedes the verb and agrees with it, a possible assignment of relation is SUBJ. The semantic check, which is activated on the basic relation (i.e. OBJ) validates such an assignment, because *a girl that works at the wardrobe* may happen to be persuaded. The analysis proceeds to the next nominal description, with the set of relations (Cho-1, SUB-Goal) not assigned yet. **Da un cliente (by a customer)** has exactly the form of a Cho-1 in Italian. The Proposal Rules, whose hypothesis is confirmed by the semantic check, are sufficient to deal with this situation. When we find the verb **comprare (buy)**, the PR group assigns to such a description the SUB-Goal relation and consequently the lexical rule associated with **persuadere** assigns the SUBJ relation of the initial stratum of **comprare to la ragazza che** The initial stratum of **comprare** features also an OBJ relation, that will be assigned to **enciclopedia**, when it is found.

The completeness of the set of found grammatical relations is checked when the node corresponding to the verb is "closed", i.e. when it cannot have further modifiers.

4. Conclusions

The paper illustrates how RG can be used to map in a principled way surface dependency relations into thematic roles.

³ The Goal relation, such as Instrument or Location, is a non-term relation and participates only to special kinds of promotional rules (see [Perlmutter 83] for details).

The main feature of the approach is the strict cooperation among different knowledge sources (lexicon, RG rules and semantics) in carrying out the task: this cooperation is made necessary by the partial configurationality of Italian, where the ordering of constituents can only be considered as the basis for plausible suggestions, but not as the source of strict constraints. The adoption of an unmarked input (an unlabelled dependency tree) makes available a flexible starting point that leaves the RG module the task of making the required inferences.

The ideas expressed herein are implemented in the GULL system (see [Lesmo, Torasso 83, 85a] for the syntactic part [Di Eugenio, Lesmo 87] for the basic ideas about semantics): both levels of grammars are represented via condition-action rules. The system is implemented in Common Lisp and runs on SUN workstations.

References

- [Bruce 75] Bruce B., *Case Systems for Natural Language*, Artificial Intelligence 6, 1975, pp. 327-360.
- [Di Eugenio, Lesmo 87] B.Di Eugenio, L.Lesmo, *Representation and Interpretation of Determiners in Natural Language*, Proc. 10th IJCAI, Milano, Italy, 1987, pp.648-654.
- [Fillmore 68] Fillmore C., *The case for case*, in "Universal in Linguistic Theory", (Bach and Harms eds.), Holt, Rinehart and Winston, 1968.
- [Hudson 84] Hudson R., *Word Grammar*, Basil Blackwell, Oxford, 1984.
- [Jappinen et al. 86] Jappinen H., Lehtola A., Valkonen K., *Functional Structures for parsing dependency constraints*, Proc. COLING 86, Bonn, Germany, 1986, pp.461-463.
- [Lesmo, Torasso 83] Lesmo L., Torasso P., *A Flexible Natural Language Parser based on a two-level Representation of Syntax*, Proceedings of the 1st Conference ACL Europe, Pisa, Italy, 1983, pp.114-121.
- [Lesmo, Torasso 85a] Lesmo L., Torasso P., *Analysis of Conjunctions in a Rule-Based Parser*, Proceedings ACL 85, Chicago, USA, 1985, pp.180-187.
- [Lesmo, Lombardo 91] Lesmo L., Lombardo V., *A Dependency Syntax for the Surface Structure of Sentences*, Proc. of WOCFAI, Paris, July 1991.
- [Perlmutter 83] Perlmutter D.M. (ed.), *Studies in Relational Grammar 1*, Univ. of Chicago Press, 1983.
- [Perlmutter, Rosen 84] Perlmutter D.M., Rosen C.G., (eds.), *Studies in Relational Grammar 2*, The Univ. of Chicago Press, 1984.
- [Sgall et al. 86] Sgall P., Haijcová E., Panevová J., *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*, D. Reidel Publ. Co., 1986.
- [Somers 87] Somers H.L., *Valency and Case in Computational Linguistics*, Edinburgh Univ. Press, 1987.
- [Stock 89] Stock O., *Parsing with Flexibility, Dynamic Strategies, and Idioms in Mind*, Computational Linguistics 15, 1989, pp.1-18.