# PARSING DIFFICULTIES & PHONOLOGICAL PROCESSING IN ITALIAN

Rodolfo Delmonte
Istituto di Linguistica e Didattica
Ca' Garzoni-Moro - S.Marco 3417
Università degli Studi di Venezia(I)

A B S T R A C T

A recognition grammar to supply information to a text-to-speech system for the synthesis of Italian must rely heavily upon lexical information, in order to instantiate the appropriate grammatical relations.
Italian is an almost free word order language which nonetheless adopts fairly analysable strategies to move major constituents: some of these can strongly affect the functioning of the phonological component. Two basic claims will be made: i. difficulties in associating grammatical functions to constituent structure can be overcome only if Lexical Theory is adopted as a general theoretical framework, and translated into adequate computational formalisms like ATN or CHART; ii. decisions made at previous point affect focus structure construal rules, which are higher level phonological rules which individuate intonation centre, build up adequate Intonational Groups and assign pauses to adequate sites, all being very sensitive to syntactic and semantic information.
We will concentrate on Subject/Object function association to c-structure in Italian, and its relation to ATN formalism, in particular HOLD mechanism and FLAGging. Then we will show how syntactic decisions interact with an intonation grammar. We shall also introduce two functional notions: STRUCTURE REVERSIBILITY vs. FUNCTIONAL REVERSIBILITY in Italian.

## 1. INTRODUCTION

In a recent paper we presented (Delmonte, 1983) a phonological processor for Italian which has been implemented at the University of Venice and is used in a text-to-speech system (Delmonte et al., 1984) for the synthesis of Italian at the Centre of Computational Sonology of the University of Padua. Recently the system has been equipped with a lexicon and a morphological analyser (Delmonte et al., 1985) while the parser is on its way to be built, which, since we adopt Lexical-Functional Grammar (LFG) (Bresnan, 1982), as background linguistic theory, should take the form of a chart, much in the vein of Kay's (1977,1979,1980) and Kaplan's (1973) functional and general syntactic parsers. At present we are working at the context-free grammar and the semantic information to be associated with each lexical entry. As it appears, Italian is a much more complex language to be analysed when compared with English, German and French. As we shall discuss in the paper, difficulties arise basically because Italian has a relatively much higher freedom in the order of constituents than the above mentioned languages. Also, the phenomenon of the unexpressed Subject or Null Subject, makes the working of a parser a much harder task. In this sense, a chart being unbiased as to what procedure to adopt in the course of the analysis, will allow the parser to benefit both from top-down and bottom-up procedures in an efficient way (plus the obvious back-up and parallel processing operations usually required). Besides, both semantic and grammatical features need to be present throughout the parsing process, and they will be used to guide the overall parsing strategy.

## 2. PASSIVE STRUCTURES AND REVERSIBILITY

Assuming that the ultimate goal of a parser is that of accomplishing the analysis of the input text in terms of underlying grammatical relations, we are usually fronted with the task of assigning thematic roles to functional representations mapped onto constituent structures, as well as defining other non trivial semantic relations including ellipsis, predication, coordination, quantifier/negation and modality scope, head to modifier/complement/adjunct relations. All these aspects are relevant to a constructional rule of focus structure which addresses directly the informational structure of the text.

The intermediate level of grammatical function assignment is in this perspective a relevant level of representation in that it allows the mapping from c-structure to θ-roles: in this sense, it contributes to speed up the recognition procedure.

In English, a completely reversible structure is

the following:

1. The secretary has been killed by the director.

in which, either NP arguments of the predicate KILL can assume the grammatical functions of SUBJect or OBJect of the sentence. On the contrary, in non-reversible passive structures like,

2. The book has been read by John.

only an animate NP argument can be interpreted as SUBJect of the predicate READ; thus, inanimate NP arguments can only be interpreted as OBJect of the sentence.

It is clear that non-reversible passive structures contain additional grammatical cues to speed up comprehension, but these cues are only available from lexical entries in which selectional restrictions are listed. Semantic features are thus called into question, and are used to constrain θ-role assignment in recognition grammars, in order to derive from functional structure adequate mapping for focus structure.

In a more strictly computational perspective, verb morphology is accessed first for Agreement tests; when passive morphology is detected, this local cue is sufficient to reverse grammatical function assignment carried out so far to the previously analysed NP SUBject, and assign it Object function. Also transitivity test is necessary not to get entangled with intransitive verbs taking Auxiliary BE.

If we regard constituent discontinuities as the major issue to be addressed in the grammatical perspective so far outlined, passive structures are the canonical case of NP movement in Transformational Grammar (TG), in which traces or gaps are left behind by displaced constituents; or within LFG theory of control, the coindexing performed on f-structures between metavariables and empty nodes. In a strictly configurational language like English there does not seem to be such a strong motivation for adopting LFG theoretical framework and introducing the intermediate representation in terms of f-structures. It might as well be sufficient to inspect precedence and dominance relations as instantiated by constituent structure and relate them to PSR of a context-free grammar in which canonical constituent order is encoded. Since in tensed clauses either a lexical or a pronominal SUBJect must be expressed in preverbal position - or else a dummy pronoun like "there, it" - it could be possible to label NP1, or the one dominated directly by S, as SUBJect, whereas postverbal NP2 if present, as OBJect of the clause, or the one dominated directly by VP.

Unfortunately, what applies to English or other fixed word order languages, does not apply to romance languages and in particular to Italian or Spanish, which have been called Null Subject Languages (NSL). One of the distinguishing properties of NSL is that they do not have a canonical position for NP Subject: it can either appear in preverbal position as in English, or

i. in postverbal position as a case of Subject Inversion;
ii. be unexpressed as a case of obviative or extrasentential pronominal, in tensed clauses;
iii. be stranded or extraposed, usually in tensed clauses; NP Subject has been moved out of its matrix clause and placed after an embedded clause, which it controls (Subject must be unexpressed), or not - no intervening lexical NP Subjects are allowed, however.

Before going into the analysis of Italian with more detail, it is worth while noticing that not always NP1 entertains Subject function, nor NP2 can be interpreted as Object of a finite clause in English, as the following examples show:

3. Computers have been given no consideration whatsoever by linguistis in Italy.

4. Her father Mary hates.

5. The latest book by Calvino sells well.

6. The logical operator .NOT. applies to the parenthesized statement.

7. Geneva is easy to reach from Italy.

in which we have cases of fronted NP2 detectable only by having access to NPs inherent semantic features. Thus, in 3, it is OBJect2 which has been passivized and not NP2; in 4. we have a topicalized sentence with fronted NP2; in 5. SELL is used in ergative structural configuration, in which NP2 is raised to Subject; the same applies to 6, a case in which Subject NP would be always omitted (subjectless impersonal structures are frequently used in technical and scientific English); also 7. is a subjectless structure, in which "tough predicate" appears and Object NP2 is raised to Subject position. And now briefly, NP2 need not always be interpreted as Object of its clause, as shown below:

8. There came the magician with his magic rod.

9. But the real murderer is the landlord.

10. Mary gave John a beautiful present.

where 8. is a presentation sentence with a dummy pronoun "there" and the Subject NP is in postverbal position; 9. is a predication sentence in which something is predicated about the NP Subject "the landlord" in postverbal position; and in 10. the postverbal NP is OBJect2 of ditransitive Verbs constructions, which has undergone dative shift.

### 3. WH- CLAUSES AND FUNCTIONAL REVERSIBILITY

In Italian, reversible structures are also present sometimes obligatorily, always optionally, in wh- clauses. Let us quote first the following example,

11. This is the lion that ate the man that ate the

11. rabbit that ate the carrot.

Each embedded clause can only be interpreted as containing an NP argument of EAT assuming Subject function when in preverbal position and Object function when in postverbal position, the complementizer "that" relativizing only the left-adjacent NP and interpreted as Subject of the following clause. No such interpretation is allowed in 12.

12. This is the man (that) the lion ate.

in which the intervening NP "the lion" prevents the complementizer from occupying strictly preverbal position, thus being assigned Object function and as such it is possibile to omit it; Subject function being thus assumed by "the lion". But in Italian also 13. would have to be allowed,

13. *This is the carrot that ate the rabbit that ate the man that ate the lion. (Questa è la carota che mangiò il coniglio che mangiò l'uomo che mangiò il leone).

This sentence is absolutely symmetrical semantically to 11., except for the fact that 13. predicates something about "a carrot" - the head - whereas 11. predicates the same concept though with a different informational structure, focus on the "lion". Conceptually an operation recalling the passive.

Let us now reformulate the two notions that we have introduced so far, structural reversibility vs. functional reversibility and repeat example 1,

1. The secretary has been killed by the director.

where what we want to stipulate is the possibility to interchange Subject/Object function between the two arguments of the predicate KILL; thus, we could also have,
1.i The director has been killed by the secretary.
besides the active forms,
1.ii The director has killed the secretary.
1.iii The secretary has killed the director.

Structural reversibility involves basically the possibility to use the same constituent order and to freely alternate the instantiation of grammatical functions, while the underlying grammatical relations intervening between the arguments of the predicate, change. What is implied is that: although both NP arguments of the predicate are eligible to be interpreted as Subject or Object, only one interpretation will result in each case a grammatically valid configuration results. Thus, even though thematic roles can be attached interchangeably to either preverbal or postverbal NPs without violating selectional restrictions or semantic compatibility conditions, it is the final constituent order and structure that decides on the final interpretation.

In this sense, non-reversible passives contain cues such that their verb's selectional restrictions permit only a single well-formed mapping between NP positions in phrase structure tree and functional structure positions. Other cues will help producing the final interpretation besides structurai syntactic ones: since either NP1 or NP2 in (surface) c-structure won't match their selectional restrictions with the requirements of functional structure mapping, the parser will compute directly one or the other interpretation disambiguating the resulting sentences on the basis of conditions and tests on the arcs, rather than on its context-free grammar. Thus in 2.

2. The book has been read by John.
we are not allowed to build,
2.i* John has been read by the book.
2.ii* The book has read John (OK in Italian)
but only,
2.iii John has read the book.

without violating selectional restrictions. Going back to our previous examples, 11. and 13., what we have then is an example of non-reversible functional structures. In this case, both preverbal and postverbal positions in constituent structure could be freely accessed by the two arguments of the predicate EAT as was the case with example 1, and contrary to what happened with example 2.:

A. what is blocked in structural reversibility - as non reversible passives show - is the availability of one of the structural constituent positions to be accessed by both arguments of the predicate;
B. in functional reversibility - as non reversible wh- clauses show - it is the availability of one of the arguments of the predicate to be assigned any grammatical function, that is blocked.

In B. constituent order is irrelevant, and it is crucial in A.; B. is typical of NSL, while A. is typical of configurational languages in which grammatical functions can be associated in a reliable way to fixed or canonical constituent orders. In Italian, no such canonical order exist, essentially because both preverbal and postverbal constituent positions constitute an unmarked case for Subject/Object functional assignment.

The consequences of this analysis of Italian in terms of functional reversibility will be explored when analysing functional reversible structures. For now it suffices to remark that a parser is unable to rely on constituent order alone to produce reasonable predictions on the underlying grammatical relations: it will be obliged to make available semantic information in all cases of tensed active clauses.

**4. SUBJECT EXTRAPOSITION IN WH- CLAUSES**

In English, "that/which/who" restrictive relatives and indirect questions, as well as wh- questions, are easily computable in that the extraction site - the wh- word/phrase can be extracted either from Subject or Object position - is readily avail-

able to the parser by looking up lexical subcategorization frames and phrase structure so far computed - roughly the Agenda and the Chart. In the case of wh- questions, do-support will trigger Object function assignment to the fronted wh- word/phrase; besides, also wh- questions without do-support or subject auxiliary inversion are allowed, only when the question element is the Subject. In the remaining embedded structures, an intervening NP in preverbal position will trigger Object function for the wh- word/phrase. Let us look at some examples,

14. Dove (VP e ha sepolto ($NP_{object}$ il tesoro (S' che e ha rubato e)))($NP_{subject}$ l'uomo (S' di cui e parlavi e))

where we only marked major constituents and empty positions in f-structure. This example translates literally example n.120 from Ritchie, 1980:

14.i Where did ($NP_{subject}$ the man (who you mentioned e) bury ($NP_{object}$ the treasure (which he stole e))?

In Italian we always have this elaborate structure when heavy NPs are involved in wh- questions. If we coindex NPs with empty nodes we get,

14.ii Dove(VP $e_i$ ha sepolto($NP_j$ il tesoro(S' $che_j$ $e_i$ ha rubato $e_j$)))($NP_i$ l'uomo(S' di $cui_i$ $e_k$ parlavi $e_i$))?

where NP Subject "l'uomo" has been displaced beyond two bounding nodes - in Italian NP and S' count as such (see Rizzi, 1980) - and also binds the empty NP Object position of the lower relative clause, whereas the null subject position in front of "parlavi" is assigned obviative or disjoint reference,to an extrasentential antecedent. Example 14 is a replica of the simple structure of a yes/no questions:

15. Ha finito i compiti tua sorella?
(Has your sister finished her homework)

where postverbal position is again reserved for NP Object and the NP Subject "tua sorella" has been stranded or "extraposed". In wh- questions, the problem with Italian syntactic structure is due to the absence of a clear surface indicator for grammatical function assignment, even though, as a rule, it is the Object NP that is questioned, as in 14.15. But the following examples do not follow this rule:

16. Quale pesce ha pescato la segretaria?
17. Quale segretaria ha pescato il pesce?

which can be translated respectively as,

16.i Which fish did the secretary catch?
17.i Which secretary caught the fish?

where the underlying grammatical functions can easily be recovered due to the presence of do-support in 16.i - thus inducing Object function on the questioned element, and the lack of do-support in 17.i thus inducing Subject function assignment on the wh-phrase.

Unfortunately in Italian 16. and 17. contain no structural cues indicating that what is being questioned is either a Subject or Object, in other words these structures are fully functionally reversible. Grammatical functions are assigned when selectional restrictions for the predicate CATCH are accessed and semantic inherent features of the arguments are detected and compared. Further difficulties arise with embedded structures in wh- questions, as shown below:

18. Chi era la persona che Gino ha incontrato e ieri?
19. Chi era la persona che e ha incontrato e Gino?
20. Chi e ha detto che e avrebbe assunto e il capo?
21. Che cosa e ha detto che e avrebbe acquistato e al mercato Gino?
22. Chi e intendeva mettere in imbarazzo e Mario?
23. Quale segretaria e conosceva e il direttore?

translatable as,

18.i Who was the person that John met yesterday?
19.i Who was the person who met JOhn yesterday?
20.i Who said that he/she would have engaged the chief?
ii Who did the chief say that he would have engaged?
21.i What did he/she say that John would have bought at the market?
ii What did John say that he would have bought at the market?
22.i Who did Mario intend to upset?
ii Who intended to upset Mario?
23.i Which secretary knew the director?
ii Which secretary did the director know?

We only marked structural gaps at functional level with the underlined e; here the first difficulty is constituted by the ambiguity naturally associated to all these structures, with the exception of 18. In this case, no ambiguity arises because we have a predicative structure followed by a restrictive relative in which Subject preverbal position is appropriately filled by the proper noun Gino/John. However, in 19, another interpretation is available: "la persona" is the head NP of the following relative and controls the empty subject position of the Verb INCONTRARE, while "Gino" is Object NP. This interpretation, though, is not the only one available in 19, since in Italian, Gino might as well have been extracted from Subject position via Subject Inversion - or rather, it might occupy postverbal position, another canonical position for Subject function in Italian.

In 20. then, three gaps are available, consequently three alternative interpretations as follows,

a. Chi ha detto che il capo avrebbe assunto ieri
b. Chi ha detto che avrebbe assunto il capo ieri
c. Chi ha detto il capo che avrebbe assunto ieri

where in a. we have the higher clause Subject posi-

tion controlled by "chi", and "il capo" controlling the lower Verb; in b. "chi" is the Subject of the nigher clause and "il capo" the Object of the lower one; in c. "il capo" is the Subject of both the higher and lower clause, and "chi" is made to fill Object position of the lower verb. For 21, the following two alternative structures, though, are only available:

a. Che cosa (x) ha detto che Gino avrebbe acquistato al mercato.
b. Che cosa ha detto Gino che avrebbe acquistato al mercato.

no other structure is available since "che cosa" is usually extracted from Object postverbal position, and Italian does not allow double filled Object positions. In b. Gino controls both empty subject position in the higher and lower Verbs, and the Object postverbal position is reserved for the wh- word: so, only a. can alternatively be generated.

These interpretations are generated also because the predicate-argument structures of the Verbs allow it: INCONTRARE is an only transitive verb, while ASSUMERE can be intransitivized, and ACQUISTARE is again an only transitive verb. Transitive verbs require an Object NP while intransitivizable ones don't. With intransitive verbs only one interpretation would be allowed as in:

24. Chi ha detto che sta arrivando Gino?
(Who said that John was arriving?)

where ARRIVARE does not allow an Object NP, thus "Gino" must be analysed as Subject; besides, also "chi" could not possibly be analysed as Object NP of ARRIVARE, so it is made to occupy Subject position of the higher clause and inserted in the empty slot adjacent to the wh-word.

## 5. HOLD MECHANISM AND WH- CLAUSES

It appears thus, that a minimal requirement for producing adequate parses for these complex wh-clauses is access to predicate-argument structure in the Lexical Form - roughly subcategorization frames - of Verbs. These would be entered in the Agenda as expectations to be fulfilled by the parser. It is also clear that we would like to have a rule for functional control induced structurally, by means of which, empty Subject positions in tensed embedded clauses and in matrix clauses would be bound by lexically filled adjacent Subject position (corresponding to c-command dominion in terms of syntactic binding - See Zaenen, 1983).

The problem now is the following: how do we get extraposed/stranded NP Subject or Object to climb up to fill the appropriate gaps?

In ATN formalism, a question element register HOLD, is used to contain the questioned element which is stored temporarily until the rest of the clause is processed. In wh- clauses the element is then passed down to any constituent that might use it or that in turn could pass it down to one of its constituents. If we follow Winograd's suggestion, we might: "put the Held item into a special role register associated with every type of constituent that uses it"(1983, 233).

In particular, "chi" in examples 19.20.24. could be associated with NP constituents/empty NP nodes in f-structure. Since it could be made to fill either Subject or Object positions. When transitivity and agreement have been checked, and the Verb of the lower clause has been parsed we will be left with the following parallel structures, schematically represented:

i. chi e / ha detto / che pro / avrebbe assunto e /
ii. chi e/ ha detto / che chi / avrebbe assunto e /
iii. e pro/ ha detto/ che pro / avrebbe assunto chi/

| COMP | NP / VP | /COMP | NP / VP | NP / |
|---|---|---|---|---|
| | SUBJECT | | SUBJECT | OBJECT |

when the lexical NP "il capo" il reached, it must be made to climb up into the empty registers (e) or pro the obviative extrasentential pronominal, of the already parsed phrase structure. Since ATN grammars are usually made for top-down processing strategies, and this example would clearly constitute a case of bottom-up data-driven processing strategy it would seem that a CHART could perform better, being unbiased as to what strategy to follow. Anyhow, this is the structural representations that we would like to get:

iv. chi / ha detto/che/il capo/avrebbe assunto e/
v. chi / ha detto/che chi/avrebbe assunto il capo/
vi. chi il capo/ ha detto/che/il capo/avrebbe assunto e/

where the empty node in iv. represents the extraction site of the wh- word; in v. chi is Subject in both clauses; and in vi. il capo is Subject of both clauses and chi is extracted from e. We must remember that the lexical NP "il capo" might as well be lacking, without affecting the grammaticality of these interpretations: in this case i.ii.iii. would have to be preserved and accounted for on a discourse level, i.e. antecedents of empty nodes/pro positions be recovered from text or discourse.

If we look in more detail into the HOLD mechanism, we can easily see that it embodies a particular linguistic phenomenon: i. it individuates wh-words or phrases displaced leftward from their original locations, and stores them temporarily in a register; ii. it inspects forward its right context in search of a hole in constituent structure, using lexical information; iii. the hole must have the same constituent label of the stored item, and must be in a lower network, where the contents of the Hold register will have to be passed down. A

copy of the NP (carrying SUBJect or OBJect function) or PP (carrying INDirect OBJect, ADJunct - of time, location and direction ) constituent parsed will be stored to be used later on by another network where the corresponding hole is detected.

The problem with HOLD mechanism consists in the fact that the linguistic phenomena of Italian we are discussing about the Extraposed Subject are of a different nature: basically they differ from the ones dealt with HOLD in the non availability of the constituent to be stored in a register at the beginning of the analysis, since usually with our set of phenomena, first comes the "hole" and then the constituent to fill it with.

In other words, this is not the procedure that we envisage to use in order to parse our null/empty subject Italian sentences. In fact, our recognition mechanism shall have to deal with the following series of events:

1. wh- words/phrases will be available first and followed by their extraction sites, hence the corresponding holes will have to be detected;
2. NP displaced leftward, either in terms of grammatical function assignment - the OBJect comes before the SUBJect in constituent structure - or as topicalized/left-dislocated NPs will be available first and the appropriate function reassignment will have to be performed as soon as the verb is reached: it can either show passive voice or be checked by agreement and transitivity tests on arcs;
3. null/empty NP positions in preverbal structure will give rise to INFLection features LIFTing to the empty slot, and then Subject inversion or null subject will have to be accounted for. In this case, only INFLection features will be available, and will possibly be followed by the NP they belong to.

What we need then is the inverse procedure envisaged for wh- movement, i.e. the HOLD-VIR mechanism; basically this amounts to saying that the cases we are dealing with are simply cases of NP movement like passive structure, the only difference consisting in the fact that no NP is available at the start.

If we look into passive NP function assignment mechanism, we can see that what triggers the procedure is verb morphology: once passive voice is detected in the main verb, DIRect OBJect is set to SUBJect, which must have been already properly parsed (see Winograd, 217). This setting procedure is like an assignment statement in a programming procedure: the first NP encountered by the parser is assigned SUBJect function at first; when the verb is met, its label is changed to that of DIRect OBJect. Subject is subsequently set to a dummy NP,which as Winograd comments, is used to indicate an NP node with no register contents, constructed to represent un unknown subject. When the PP with preposition "by" is parsed, it is taken as the phrase specifying the agent; thus the dummy subject NP will be set to the (by) NP, now the deep subject, or stay empty if the sentence is an agent-deleted passive like "the fish have been caught". In order to parse our null subject sentences we would adopt the same procedure used for passives, except for the fact that our Subject NP need not be present in the same S/NP network in which the hole is detected. As we noted in extraposed subject sentences, the NP could appear rightward beyond two bounding nodes (even more are allowed as long as no intervening NP Subject appears). All networks where a hole has been FLAGged should be accessed by the parser whenever an "exceeding" NP is parsed, or simply an NP eligible to be interpreted as Subject of the higher predicates already encountered in the analysis.

## 6. FOCUS STRUCTURE AND FUNCTIONAL REVERSIBILITY

We already discussed elsewhere (Delmonte, 1983, 1984) rules for Focus Assignment, where Focus was directly compared to Intonational Centre, thus prosodic focus rather than semantic or informational focus. Focus structure addresses directly the second notion of Focus which need not be symmetrically computable at a phonological and syntactic level. There are clear asymmetries which can be detected in the following Italian structures, we shall discuss.

Basically what is addressed structurally by focus structure is the VP in c-structure, which must include the last argument in the c-dominion of the predicate or f-controlled by it; complements/adjuncts Syntactically-bound are adjoined into focus structure. Also the first of optional arguments adjacent to the VP - like PP, complements, adjuncts possibly controlled by Strong Lexical Form of the predicate, or adjoined to it by means of a Theory of Syntactic Closure - can constitute focus structure.

Let us first go back to our examples, 22. 23., which can be defined completely functionally reversible structures, and see how they interact with focus structure construal rules. In particular, they can be analysed as follows,

22.i Chi intendeva mettere in imbarazzo Mario?

23.i Quale segretaria conosceva il direttore?

what in English is achieved by means of syntactic structure, in Italian is achieved via focus structure, which we have represented here as underlining, at the end of which a pause may be produced. The phonological focus or intonational centre must be included in focus structure on one of its constituent, usually the last on the right. According to the constituents, in condition of adjacency, focus structure can thus be expanded and produce two different focus assignment: with focus on the question-

ed element this will be interpreted as Subject of the clause; with focus on the VP, the questioned element will be interpreted as Object. Usually the VP delimits focus structure in wh- questions in Italian.

As we said previously, in functional reversibility, even though both positions are available to be filled by the two arguments of the verb, only one will produce the desired grammatical relations. Nonetheless both positions in c-structure are grammatically viable to instantiate meaningful sentences, in keeping with structural and semantic restrictions of the grammar of Italian.

As with reversible passives, in 22. 23. both arguments of the verbs can be interpreted as either SUBJect or OBJect, but differently from reversible passives, reversible wh- clauses don't make available to the parser any constituent structural hint as to which NP argument enacts which grammatical function.

If this is the situation with wh- clauses, more complex configurations will result in declaratives, given our fourfold classification of phenomena related to Subject function location in constituent structure. In the following examples the two basic simple declarative sentences will produce nine permutations each with 20 different structural relations in c-structure, but only two possible underlying grammatical relations in terms of functional associations.

25. Il sindaco sposerà mia sorella.
  1. Mia sorella sposerà il sindaco.
  2. Mia sorella il sindaco sposerà.
  3. Sposerà il sindaco mia sorella.
  4. E' mia sorella che il sindaco sposerà.
  5. Il sindaco la sposerà mia sorella.
  6. Mia sorella la sposerà il sindaco.
  7. La sposerà mia sorella il sindaco.
  8. La sposerà il sindaco mia sorella.
  9. Quale sindaco sposerà mia sorella?

26. Mia sorella sposerà il sindaco.
  1. Il sindaco sposerà mia sorella.
  2. Il sindaco mia sorella sposerà.
  3. Sposerà mia sorella il sindaco.
  4. E' il sindaco che mia sorella sposerà.
  5. Mia sorella lo sposerà il sindaco.
  6. Il sindaco lo sposerà mia sorella.
  7. Lo sposerà il sindaco mia sorella.
  8. Lo sposerà mia sorella il sindaco.
  9. Quale sindaco sposerà mia sorella?

27. Il sindaco sposerà Marco.

28. (?) Marco sposerà il sindaco.

29.i Il sindaco si è sposato con mia sorella.
  ii Mia sorella si è sposata con il sindaco.

These focus structures can be induced by the following lexical entries for SPOSARE:

30. Lexical entries for SPOSARE

1. "SPOSARE ((SUBJ), (OBJ))"

$$\begin{pmatrix} \text{SUBJ} & \text{OBJ} \\ \text{agent} & \text{patient} \\ \alpha\,\text{sex} & -\alpha\,\text{sex} \end{pmatrix}$$

2. "SPOSARE ((SUBJ), (OBJ))"

$$\begin{pmatrix} \text{SUBJ} & \text{OBJ} \\ \text{civil} & \text{patient} \\ \text{official} & \\ \text{priest} & \end{pmatrix}$$

$\text{PRED}_{caus}$ : CAUSE (x, BECOME (PRED (y)))

$\text{PRED}_{inch}$ : BECOME (PRED (y))

$$(\ \text{REFL}) =_c +$$

3. "SPOSARE ((SUBJ)) (CON OBJ)"

$$(\ \text{REFL}) =_c +$$

If we look at these entries, we are presented with a causative verb meaning "officially marry two people (cause people to get married), usually of different sex", and a normal active agentive verb meaning "get married". Thematic roles associated to NPs occurring with the verb vary according to the grammatical functions associated with c-structure configurations, apparently. In fact, lexical and semantic restrictions will be paramount in deciding focus structure and θ-roles association to NP positions.

In the permutations under 25. "il sindaco"(the mayor) is assigned Subject function and two ambiguous readings may be generated: either the mayor is the civil official who causes my sister "mia sorella", the NP OBJect to contract marriage with someone else, or he is himself the affected agent of the marriage. In the permutations listed under 26. "mia sorella" is assigned Subject function, thus only one interpretation is allowed, that is the agentive reading and the mayor is going to become my sister's husband.

If we look at the permutations, we have in 2. topicalization, with OBJect NP in focus structure (FS); in 1. the grammatical relations of 25 are preserved only if emphatic reading is assigned with contrast-

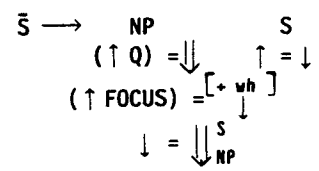
ive meaning. This also applies to 26.1; 3. is a case of inverted subject thus being included in focus structure; 4. is a cleft structure in which the NP fronted has Object function, and FS includes the copulative or predicative sequence, excluding though the following completive; 5. is a right dislocated structure in which the Object has a copy pronoun in preverbal position and FS only delimits the VP thus resulting; 6. is a left-dislocation and FS only includes the VP with the resumptive pronoun; 7. is a right dislocated structure very much like 5. except that the subject has been stranded to the end of the sentence; the same applies to 8. which is a right-dislocated structure with inverted subject, the NP appearing right after the verb and thus available - optionally though - for inclusion in the VP. Finally in 9. we have a wh- question in which the questioned element is included in FS according to which grammatical relations have to be instantiated: SUBJect function in narrow FS, OBJect function in wider FS.

These processes in informational structure, where FS is computed, are made possible when all levels of analysis are integrated and LFG representation schemata and lexical rules are made to apply. In particular, since "causer" thematic role can only be assigned to Subject NPs and not to OBJect NPs - the more so in functionally reversible structures - the appropriate grammatical relations will be altered if grammatical functions are not properly assigned. Inchoative lexical redundancy rule allows only the agentive meaning to be instantiated, simply because this lexical form derives via a lexical rule applied not to causative but to active transitive lexical form of predicate SPOSARE. Thus 28. will be marked as semantically deviant, whereas both 29.i/ii are interchangeable in meaning.

In other words, if no information is available as to the grammatical functions being entertained by the NPs argument of the predicate, the opposite meaning may well be instantiated, and this will affect the phonological representation which in turn will affect the phonetic realization of the sentence. This information will necessarily have to be derived from the lexical form associated with the predicate, and eventually be adequately coupled with annotated PSRs as represented within the framework of LFG.

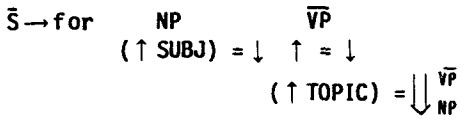
Thus we propose to couple PSRs with phonological marking of focus when relevant: this representation will interact with lexical representations and lexical redundancy rules to filter out c-structures and produce the appropriate f-structures. Semantic focus is also annotated when non-ambiguous structures result. For instance no phonological marking is indicated in wh- questions since as we already noticed, when the questioned element is in narrow FS it will have to be analysed as SUBJect, whereas in wider FS as OBJect. Also, TOPIC does not give rise to phonological marking, apart from comma intonation assignment to XP in right/left dislocation.

1. **Wh- Questions**

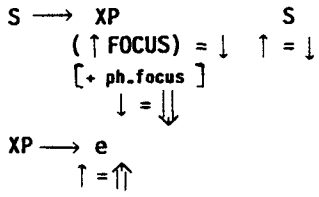



2. **Tough Predicates**

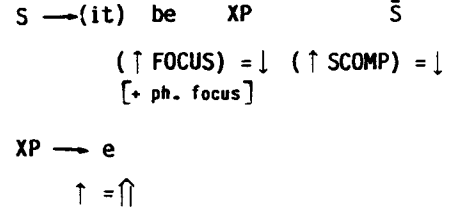
tough : A (↑ PRED) = "TOUGH ((↑ SCOMP))"
(↑ SCOMP TOPIC) = (↑ SUBJ)



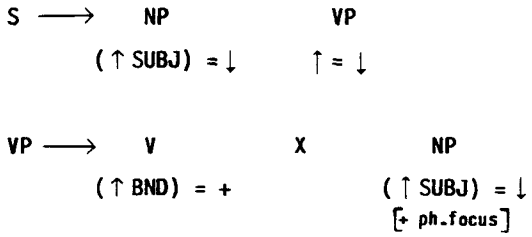

3. **Topicalization**




4. **Cleft constructions**




5. **Subject inversion**



6. Right/Left Dislocation

In the configuration

$$\bar{S}\,[_{TOP}\,[XP]\,\bar{S}\,[\ \ldots\ \underset{[+\,acc]}{pronoun}\ \ldots]]\ (XP)$$

assign comma intonation to the XP out of S, whose coreferent pronoun is inside S

LFG Formalism is quite easy and straightforward to be elucidated, though we feel this is not the right place to explain it (but see Bresnan, 1982). Why this formalism should be relevant in the description of non strictly configurational languages like Italian it is intuitively apparent from the examples we reported above. The coupling of annotated PSRs with Lexical Forms in which grammatical functions are specified as arguments of the predicate is enormously advantageous in view of parsing. In fact, this mechanism will allow the parser to reduce drastically parallel structure analyses since derivated structures like Subject Inversion and Topicalization will eventually be assigned their grammatical relations in a straightforward manner, by simply looking up selectional restrictions associated to each argument position in f-structure.

For instance, in 25. 26. there is no duplication of lexical entries with NPs c-structure positions apart from permutations under 1. As we said previously this would be treated as a contrastive emphatic structure when opposite f-structure mappings have to be recovered; otherwise NP1 and NP2 would be assigned their canonical f-structure. If the parser is allowed to produce all possible analyses with the remaining permutations, a great number of duplicated structural configurations will result - as far as f-structure is concerned. This is not a desirable result, however, given that LFG formalism allows the parser to restrict its hypotheses to just those cases permitted by Italian grammar.

In particular, it is simply a case that examples 5. 6. 7. 8. are disambiguated by the presence of a resumptive pronoun marking gender differences between the two arguments of the predicate. Whenever no such information is available, the parser will again duplicate analyses - both arguments belong to the same gender.

Other approaches have been proposed (see Stock, 1982; Cappelli et al., 1984) - disregarding exclusively semantic approaches (Schank & Abelson, 1977) - for Italian which put forward global hypothesis for the availability of a semantic space (Stock) in which to manipulate syntactic structures so far analysed; or a syntactic space (Cappelli et al.) limited to the Left Context, in which to perform a small set of abstract operations "on the current hypothesis about the analysis of the whole parsed segment of the input" (ibid.,42). We believe, however, that LFG formalism together with CHART mechanisms for alternating bottom-up with top-down processing strategies, while keeping all major constituents previously completed, should be sufficient in reducing the number of alternative paths that the parser might have to follow.

## REFERENCES


**ANTINUCCI F., G.CINQUE**(1977), Sull'ordine delle parole in italiano: l'emarginazione, **Studi di Grammatica Italiana VI**, 121-146

**BATES M.**(1978), The Theory and Practice of Augmented Transition Network Grammar, in L.Bolc(ed), **Natural Language Communication with Computers**, New York, Springer, 191-260.

**BRESNAN J.**(ed)(1982), **The Mental Representation of Grammatical Relations**, Cambridge Mass., The MIT Press.

**CAPPELLI A., L.MORETTI, I.PRODANOF, O.STOCK**(1978), **Parsing an Italian Text with an ATN Parser**, Laboratorio di Linguistica Computazionale, C.N.R. Pisa, Agosto.

**CAPPELLI A., G.FERRARI, L.MORETTI, I.PRODANOF**(1984), A Framework for Integrating Syntax and Semantics, in B.G.BARA, G.GUIDA, (eds), **Computational Models of Natural Language Processing**, Elsevier Pub., North Holland, 33-57.

**CHOMSKY N.**(1972), Deep Structure, Surface Structure and Semantic Interpretation, in **Studies in Generative Grammar**, The Hague, Mouton.

**CHOMSKY N.**(1981), **Lectures on Government and Binding**, The Hague, Dordrecht.

**DELMONTE R.**(1983a), Sintassi, semantica, fonologia e regole di assegnazione del fuoco in italiano, **Atti SLI**, XVII Congresso Internazionale di Studi su Sintassi e Morfologia della Lingua Italiana d'Uso. Teoria e Applicazioni Descrittive. Roma, Bulzoni.

**DELMONTE R.**(1983b), A Phonological Processor for Italian, in **Proceedings** of the 1st Meeting of the European Chapter of the Association for Computational LInguistics, Pisa, 26-34.

**DELMONTE R.**(1983c), **Regole di Assegnazione del Fuoco o Centro Intonativo in Italiano STandard**, Padova, CLESP.

**DELMONTE R., A.MIAN, G.TISATO**(1984), A text-to-speech system for the synthesis of Italian, **Proceedings of ICASSP 84**, San Diego(Cal.)

**DELMONTE R., A.MIAN, M.OMOLOGO, C.SATTA**(1985), Un riconoscitore morfologico a transizioni aumentate, **Atti A.I.C.A.**, Firenze.

**DELMONTE R.**(1984b), La Syntactic Closure nella Teoria della Performance, **Quaderni Patavini di Linguistica** 4, 101-131.

**DELMONTE R.**(1984c), **On Certain Differences between English and Italian in Phonological Processing and Syntactic Parsing**, Istituto di LIngue Straniere MOderne, Università di Trieste.

**GUERON J.**(1980), On the syntax and semantics of PP Extraposition, **Linguistic Inquiry** 11, 637-677.

**JACKENDOFF R.**(1972), **Semantic Interpretation in Generative Grammar**, Cambridge Mass., The MIT Press.

**KAPLAN R.M.**(1973), A general syntactic processor, in R.Rustin(ed) **Natural Language Processing**, New York, Algorithmic Press,193-241

**KAY M.**(1977), Morphological and syntactic analysis, in A.Zampolli(ed), **Linguistic Structure Processing**, New York, North-Holland,131-234
**KAY M.**(1979), Functional Grammar, **Proceedings** 5th Annual Meeting of the Berkeley Linguistic Society, 142-158.
**PEREIRA F.**(1983), Logic for Natural Language Analysis, Tech.Note 275, SRI International, Menlo Park(Cal.)
**RITCHIE G.**(1980), **Computational Grammar**, Sussex(UK), The Harvester Press.
**RIZZI L.**(1982), **Issues in Italian Syntax**, Dordrecht, Foris Pub.
**SELKIRK E.O.**(1984), **Phonology and Syntax**, Cambridge Mass., The MIT Press.
**STOCK O.**(1982), Parsing on Wednesday, Rapporto Tecnico n.312, Istituto di Psicologia del C.N.R., Roma.
**WANNER E., M. MARATSOS**(1978), An ATN Approach to Comprehension, in M.Halle, J.Bresnan, G.Miller(eds), **Linguistic Theory and Psychological Reality**, Cambridge Mass, The MIT Press, 119-161.
**WILLIAMS E.**(1980), Remarks on stress and anaphora, **Journal of Linguistic Research**, 1, 1-16.
**WINOGRAD T.**(1983), **Language as a cognitive Process; Vol.I: Syntax**, Menlo Park(Cal.), Addison-Wesley.
**ZAENEN A.**(1983), On Syntactic Binding, **Linguistic Inquiry** 3, 469-504.