BINDING PRONOMINALS WITH AN LFG PARSER

Rodolfo Delmonte°

Dario Bianchi ∗

° University of Venice - Department of Linguistics - Ca' Garzoni Moro - San Marco 3417 - 30124 VENICE(It)

*University of Parma - Department of Physics - Viale delle Scienze - 43100 PARMA(It)

°E-mail: delmonte@IVEUNCC.bitnet

°VAX:MG9VEVB3@icineca

°Phone: 39-41-5298459/5204477/5287683

°FAX: 39-41-5211559

Abstract

This paper describes an implemented algorithm for handling pronominal reference and anaphoric control within an LFG framework. At first there is a brief description of the grammar implemented in Prolog using XGs(extraposition grammars) introduced by Pereira(1981;1983). Then the algorithm mapping binding equations is discussed at length. In particular the algorithm makes use of f-command together with the obviation principle, rather than c-command which is shown to be insufficient to explain the facts of binding of both English and Italian. Previous work(Ingria,1989;Hobbs,1978) was based on English and the classes of pronominals to account for were two: personal and possessive pronouns and anaphors - reflexives and reciprocals. In Italian, and in other languages of the world, the classes are many more. We dealt with four: a.pronouns - personal and independent pronouns, epithets, possessive pronouns; b.clitic pronouns and Morphologically Unexpressed PRO/pros; c.long distance anaphors; short distance anaphors. Binding of anaphors and coreference of pronouns is extensively shown to depend on structural properties of f-structures, on thematic roles and grammatical functions associated with the antecedents or controller, on definiteness of NPs and mood of clausal f-structures. The algorithm uses feature matrixes to tell pronominal classes apart and scores to determine the ranking of candidates for antecedenthood, as well as for restricting the behaviour of proforms and anaphors.

1. The parser

A parser is presented which works on Italian and German, and binds pronominals within their utterance leaving unsolved the reference of free pronouns. It is divided into two main modules, the grammar and the binding algorithm. The grammar is equipped with a lexicon containing a list of fully specified inflected word forms where each entry is followed by its lemma and a list of morphological features, organized in the form of attribute-value pairs. Once the word has been recognized, lemmata are recovered by the parser in order to make available the lexical restrictions associated to each predicate. Predicates are provided for all lexical categories, noun, verb and adjective and their description is a lexical form in the sense of LFG. It is composed both of functional and semantic specifications for each argument of the predicate: semantic selection is operated by means both of thematic role and inherent features. Moreover, in order to select appropriately adjuncts at each level of constituency semantic classes are added to more traditional syntactic ones like transitive, inaccusative, reflexive and so on. Semantic classes are meant to capture aspectual restrictions which are crucial in deciding for the appropriateness and adequacy of adjuncts, so that inappropriate ones are attached at a higher level.

Grammatical functions are used to build f-structures and processing pronominals. They are crucial in defining lexical control: as in Bresnan (1982), all predicative or open functions are assigned lexically or structurally a controller. Lexical control is directly encoded in each predicate-argument structure.

Structural information is essential for the assignment of functions such as TOPIC and FOCUS. Questions and relatives, (Clitic) Left Dislocation and Topicalization are computed with the Left Extraposition formalism presented by Pereira(1981;1983). Procedurally speaking, the grammar is implemented using definite clauses. In particular, Extraposition Grammars allows for an adequate implementation of Long Distance Dependencies: restrictions on which path a certain fronted element may traverse in order to bind its empty variable are very easily described by allowing the prolog variable associated to the element in question - a wh- word or a relative pronoun - to be instantiated in a certain c-structure configuration. Structural information is then translated into functional schemata which are a mapping of annotated c-structures: syntactic constituenty is now erased and only functional attribute-value pairs appear; also lexical terminal categories are erased in favour of referential features for NP's determiners, as well as temporal and modal features. Some lexical elements disappear, as happens with complementizers which are done away with and substituted by the functional attribute SCOMP or COMP i.e., complement clause.

From a theoretical point of view, using Prolog and XGs as procedural formalism we stuck on to LFG very closely (see Shieber(1985); Pereira & Shieber(1984); Pereira(1985))even though we don't use functional equations: in particular the Fusion mechanism can be performed straightforwardly and the Uniqueness

Condition respected thanks to Prolog's unification mechanism. It differs from LFG's algorithm basically for dismissing functional equations: however, functional schemata can encode any kind of information in particular annotated f-structures, keeping a clear record of all structural relations intervening between constituents. In particular, long distance dependencies are treated using XGs, since they can easily encode paths from a controller to its controllee, as well as restrictions to prevent "island violations". In this case, we don't rewrite an empty category by means of a rewriting rule, as in LFG, rather, we activate a procedure as in Pereira(1983): moreover, the bindee or controllee to be bound by its controller or binder is assigned semantic and functional features by its predicate so that semantic compatibility can be checked when required, or else features transmitted to the controller once binding has taken place: Italian is a highly structurally ambiguous or undetermined language (see Delmonte, 1985), so that semantic or thematic checking seems necessary at this level.

2. Theoretical Background

Italian has three reflexive elements, one of which is a possessive anaphoric pronoun, "proprio", than a short distance reflexive pronoun, "se stesso", and a long distance one "sè". The short distance reflexive "se stesso" has a distribution that is somewhat similar to the English reflexive "himself", though there are differences between the two. It may corefere with a coargument and its antecedent must appear in the same minimal finite domain. On the contrary with the long distance reflexive "sè" the antecedent must be a subject: however it must be "governed" by a preposition, i.e. it must be contained in an OBLique or an ADJunct PP. As to the long distance possessive anaphoric pronoun "proprio", it is subject oriented and clause bound, but in lack of an adequate antecedent it may look out of its clause (complement or adjunct or coordinate) for its antecedent. In addition, there is the multivalued clitc "si" which may be assigned the following functions "passivizing", "reflexive", "impersonal or arbitrary": its behaviour is determined strictly by the verb predicate to which it is bound. None of the reflexive elements may be used as SUBJects.

Italian has also four pronominal elements, one of which a possessive pronoun, "suo", than a Null Subject pronoun which behaves very closely to the English personal pronouns; finally a set of lexical independent pronouns which are used for contrastive or emphatic aims. All these pronouns look for their antecedent outside their minimal containing clause. As to the possessive "suo", it behaves quite differently from the corresponding English "his". "His" can be bound by an OBJect coargument, when it is contained in the SUBJect NP as for instance in "His daughter loves John". This is not allowed in Italian, the SUBJect being a strong domain for reference. The same applies to "proprio", which being a possessive anaphoric pronoun is sensitive to the grammatical function it is contained in. However, there is one exception, and this is the case constituted by psychic verbs, whose SUBJect is characterized by a thematic role which is very low in the hierarchy of theta-roles: it is an (emotional) Theme, as for instance in "La propria salute preoccupa ognuno/Gianni". Coreference between "proprio"

and "ognuno" is allowed, but is banned with "Gianni" as antecedent. Clearly this does not apply to the corresponding "La sua salute preoccupa ognuno/Gianni" where no such coreference is allowed.

As Dalrymple(1990) comments, "constraints on anaphoric binding are lexically associated with each anaphoric element. In fact generalizations have been noted that deal specifically with the lexical form of the anaphoric element: elements of a particular morphological form are usually or always associated with particular sets of anaphoric binding constraints"(ibid.,2). It is interesting to note that such functional notions like "subject", "tense" and "predicate" are essential in defining these constraints, they all "denote some syntactically or semantically 'complete' entity"(ibid.3). As Dalrymple comments, "In a complete, consistent f-structure, a PRED denotes a syntactically saturated argument structure; presence of a SUBJ entails a predication involving some property and the subject; and presence of TENSE indicates an event that has been spatiotemporally anchored. The 'complete' entities are the relevant domain for binding conditions"(ibid.3)

The grammatical function of the antecedent is part of the antecedent constraints: an anaphor must be bound or may be bound to a SUBJect. Also the domain in which an anaphor must find its antecedent is always constrained relatively to either the syntactic predicate of the the anaphoric element is an argument, the minimal domain with a subject containing the anaphor, or the minimal tensed domain containing the anaphor. These can be regarded as domain constraints. Moreover, we may think of two kinds of binding constraints: positive and negative constraints. In line with Binding Theory of Chomsky(1981), 'reflexive' is an element which must be bound or must have an antecedent within some syntactically definable domain. On the contrary, 'pronominal' is an element that must be free, or be noncoreferent with elements in some syntactically definable domain.

However if we look at "proprio", we see that it must be bound in its minimal tensed domain, but in case no suitable antecedent is available locally, it may look outside and be assigned an antecedent or even receive arbitrary reading at certain semantic conditions, definable in terms of tense, subject, aspect. As Dalrymple suggests, there may a typology of constraints rather than a typology of anaphoric elements(ibid.,4). In previous works(Hobbs, 1978; Ingria, 1989) only syntactic constituency and c-command was considered, but recent work in linguistics has clearly proven this approach to be insufficient. In particular, both Chomsky's(1981) and Manzini's(1983) theory wrongly predict the grammaticality of sentences such as,

1) *I persuaded/told the boys$_i$ that[S1 each other's$_i$ pictures were on sale.

1i) The boys$_i$ thought that each other's pictures$_i$ were on sale.

were the reciprocal anaphor each other lacking an accessible subject in its Domain Governing Category(we will not enter into a discussion of Chomsky's binding principles nor in Manzini's modifications - see Giorgi(1984)), its Sentence (S1) is predicted to corefer freely, hence the object NP of the matrix clause is treated

as a possible antecedent on a par with the subject in 1i. Since it is wrong to say that anaphors can corefer freely, what is needed is a theory of Long Distance Anaphor, which is able to explains how the anaphor is still subject to a number of binding constraints.

Here crucially, the terms long-distance and short-distance are not used in the way in which Ingria does, and do not apply to pronouns: in particular personal pronouns, cannot be treated as long-distance anaphors(see, ibid.263) since they can pick up an antecedent in any domain whatsoever, outside their minimal domain, the clause in which they are contained - including their matrix clause and the discourse. On the contrary possessive anaphors and reflexive anaphors which count as long-distance anaphors must be bound by an antecedent before leaving their matrix clause - in other words they cannot be bound by a discourse-level antecedent. This applies to lexical personal pronouns as well as to morphologically unexpressed personal pronouns like PRO/pro which can be bound in a superordinate clause or in the discourse. However reflexives in constructions involving picture noun phrases allow non-local antecedents, and rather than being subject to syntactic constraints they seem to obey discourse constraints as Pollard and Sag(1989) discuss in their work.

In the same way it is possible to explain why in the example 2) below, with an experiencing verb, the anaphor contained in the subject NP can be bound by the object which does not c-command it, showing that this notion is not sufficient in itself to tell it apart from 3) where the same structural conditions do not apply:

2) Each other's pictures$_i$ pleased the boys$_i$.

3) *Each other's wives$_i$ murdered the men$_i$

In other words each other seems to behave like a long distance anaphor, i.e. a possessive pronoun like proprio in Italian, with some exceptions. The lack of c-command is clearly shown in case a quantifier appears as experiencer,

2i. La propria$_i$ salute preoccupa ognuno$_i$/One's health worries everyone

In the same way the two Italian anaphors *sè*, which must always be governed by a preposition and *se stesso* which can also be governed by a verb, seem to behave: *sè* is differentiated from *se stesso* by the fact that it can look for a subject in a superordinate clause and by being subject-oriented, i.e. [+SUBJECTIVE]. On the contrary se stesso can be bound also by other grammatical functions and is strictly local. *Proprio*, being a mixture of both, can be bound by other grammatical functions besides the subject, and can look for a binder in a superordinate clause.

In addition, with psychic and experiencing verbs the anaphor contained in the theme/subject can be bound by the experiencer/object - the same does not apply to the pair agent/subject & theme/object of transitive verbs. In other words, candidates for antecedenthood must be selected in accordance with their status as grammatical function and thematic role. The same applies whenever the experiencer is the subject of raising verbs - when better antecedents lack - like seem/sembrare.

12)a. ?*La propria$_i$ salute preoccupa Marco$_i$/self's health worries Mark

b. La propria$_i$ salute preoccupa ognuno$_i$/self's health worries everybody

c. La malattia della propria$_i$ moglie preoccupa molto Marco$_i$/The illness of self's wife worries Mark a lot

d. *La propria$_i$ moglie odia Gino$_i$/ *La figlia della propria$_i$ moglie odia Gino$_i$

13)a. His$_i$ wife hates John$_i$

b. *His father$_i$ hates/worries everybody$_i$

As these examples clearly show, quantifier status is a very important parameter to assess the status of candidates for antecedenthood. Also, language dependent differences are clearly visible from the paradigm: Italian possesses a wider range of pronominals and anaphors and allows binding of a possessive within the same clause as embedding becomes more deeply embedded. However deep embedding does not rescue 12d: thematic relations are the relevant criterion in this case. In the corresponding English examples, binding is performed at reversed conditions: not by a quantifier is the only requirement.

Belletti and Rizzi(1988) propose for these kind of examples and for others that Principle A of the Binding Principles be an "anywhere" principle (ibid.,314), in the sense that it can apply at D-structure, where the subject NP is contained withing the VP, thus justifying the fact that the anaphor contained in the Subject is bound before it moves to its S-structure position. Obviously, this is also relevant for sentences like

14) Which picture of himself$_i$ do you think [that Bill$_i$ likes e best]?

where Move-a has destroyed the well-formed binding configuration by extracting (the constituent containing) an anaphor from the c-domain of its antecedent. In a framework like LFG, however, no such "anywhere" principle could be made to work since categories which must be bound are only visible at one level of representation. In particular, syntactic variable are visible at c-structure and this is where they must be bound by their controller; lexical anaphors are only visible at f-structure where they must be given an antecedent in their nuclear f-structure. For an example like 14 above, there is a variable binding operation that takes place at c-structure level between the FOCus wh- phrase and the empty element in the embedded clause; when we get to the next levet of representation, the anaphor contained in the FOCus is part of a syntactic chain, i.e. is included in a non-argument function, the discourse function FOCUS, and is bound to an argument function the OBJECT of the predicate "LIKE" which also assigns it its theta-role. Since the argument function is the place in which the FOCUS will be interpreted, they bear the same index they can be bound under f-command, as we shall see.

3. F-command, operator binding and backward pronominalization

As we said, in order to perform binding procedures, all functional structures are transferred into a tree with arcs and nodes, where arcs contain grammatical function. Arcs also relate each function to its mother node, allowing in this way to compute all functions contained in an upper function: this is the crucial notion for the definition of f-command dominia(see Bresnan,1982).

The algorithm uses f-command rather than c-command and obviation to prevent clitics and lexical pronouns to look for antecedents in the same f-structure in which they are contained. Formally it is expressed as follows:

F-command

For any occurrences of the functions ,   in an f-structure F,   f-commands   iff   does not contain   and every f-structure of F that contains   contains

It is worth while reminding that f-structures coincide with lexical forms, i.e. a predicate-argument structure paired with a grammatical function assignment; in other words an fname PRED whose fvalue is a lexical form. Usually clause nuclei are the domain of lexical subcategorization, in the sense that they make available to each lexical form the grammatical functions that are subcategorized by that form (see Bresnan, 1982:304). In case also nouns are subcategorized for, the same requirement of coherence and completeness may be applied. Not all nouns however take arguments(see Grimshaw, in publication). As a consequence, "...an f-structure is locally coherent iff all of the subcategorizable functions that it contains are subcategorized by its PRED; an f-structure is then (globally) coherent iff it and all of its subsidiary f-structures are locally coherent. Similarly, an f-structure is locally complete iff it contains values for all of the functions subcategorized by its PRED; and an f-structure is then (globally) complete iff it and all of its subsidiary f-structures are locally complete."(ibid,.305) In this sense f-structure is a notion absolutely parallel to Chomsky's(1986) Complete Functional Complex, with the difference that in LFG grammatical functions are all made available in the lexical form - in particular the SUBJect -, whereas in a CFC this must be stipulated.

As for obviation, it applies to big PROs, to little pros, and to lexical pronouns: it is expressed as follows and has been incorporated in our feature system:

Obviation Principle

If P is the pronominal SUBJ of an obviative clause C, and A is a potential antecedent of P and is the SUBJ of the minimal clause nucleus that properly contains C, P is or is not bound to A according to whether P is + or - U, respectively.

Two things must be noted: first, the principle predicts that disjoint reference applies only with subject and not with nonsubject antecedents in the matrix. To distinguish reflexive pronouns which are subject-bound clause internally, in a later paper(Simpson,Bresnan, 1983), the principle has been substituted by the presence of a lexical feature [+SUBJECTIVE]. However, the conditions that must be met to bind "long anaphors" - that is reflexive pronouns which can be bound from a higher clause, and not necessarily by a subject - include mode consideration [±UNREAL], as well as the notion of f-command. In particular, the f-structure which contains the Antecedent may be the same of the one containing the Pronominal, or else be the one containing it.

A more elaborate framework results from Bresnan et al.(1985) where pronouns which must obey the Coargument Disjointness Condition (i.e. they may not be bound to an argument of the same predicate) are obviative and are marked [±NUCLEAR], thus meaning that they may or may not appear in the same syntactic nucleus as

their antecedent - an ADJunct is never part of the nucleus so that a pronoun is allowed,

16a. John wrapped a blanket around him.

b. John wrapped a blanket around himself.

The English reflexive pronoun "himself" is [+NUCLEAR] and must find an antecedent within the same nucleus containing the pronominal and a subjective function; while "him" is [-NUCLEAR]. The ADJunct "around himself" however lacks a subjective function and the anaphor must look for an antecedent in the closer higher domain. However, English pronoun "him" is not obviative like the corresponding Italian one,and this fact, when added to the presence of two sets of anaphoric pronominals, gives the rather different distribution in the corresponding Italian sentences:

16i. Gino$_i$ ha visto un serpente$_j$ vicino a lui$_{k/*i/*j}$ (John has seen a snake near him)

ii. Gino$_i$ ha visto un serpente$_j$ vicino a sè$_{i/*j}$ (John has seen a snake near "sè")

iii. Gino$_i$ ha visto un serpente$_j$ vicino a se stesso$_{j/*i}$ (John has seen a snake near himself)

Thus, the relevant domain for anaphors and pronouns contained in nominal f-structures is not the f-structures directly containing them: this is due to their functional nature and not simply to structural reasons. As to reciprocals, reflexives and possessives anaphors are all assigned SUBJECT function thus counting as possible candidates for antecedenthood: but a conflict is raised here by the referential nature of anaphors which is marked as nonreferential in their feature matrix, hence unable to become antecedents of themselves. This conflicting result works as a filter for anaphors at the structural level, erasing their ranking as candidates for antecedenthood but raising them out of their subordinate f-structure into the upper one: in this way, anaphors cannot be bound within their minimal f-domain but must be bound in the upper one, pronouns are left free to corefer.

At clause level, reflexive pronouns look for binders in the same f-structure in which they are contained:. as we said, two kinds of anaphors must be taken care of: long anaphors like "sè", and short anaphors like "se stesso". Only short anaphors can be bound by non-subjects and only long anaphors can be bound in an upper clause if no suitable binder appears in the local minimal one. The possessive anaphor "proprio" on the contrary partakes of features belonging to both short and long anaphors: it can use both a short and a long distance strategy; it is not SUBJective. We have established then that the lexical feature [-SUBJCTV] distinguishes short anaphors from long anaphors, which are marked [+SUBJCTV]. Summarizing, we have two sets of reflexive pronouns,

a. non-subjective reflexive pronouns[-SUBJCTV] "sè"

b. subjective reflexive pronouns[+SUBJCTV] "se stesso"

In addition, long distance anaphors like the possessive "proprio", non specified as to SUBJectivity, behaves both as a long and a short anaphor, according to the domain in which it can be bound, and is posaitively marked for [+pro, +ana].

3.1 Our Proposal

Our proposal takes into account the facts of Italian in particular but also those of English, Norwegian and other

languages as discussed by Enç(1989) or Dalrymple(1990). Binding is expressed by coindexation of a controller   and a controllee , just like coreference between antecedent and pronoun, in a domain F - a complex f-structure, at the following conditions:

1.   is an f-structure [+anaph] and is bound in its F-domain

2.   is an f-structure [+pron] and is not bound in its F-domain

The first part of the formulation accounts for the fact that an anaphor is in complementary distribution with a pronoun, i.e. that in the domain in which the anaphor must be bound the pronoun must be free, or not be bound. Now, the smaller domain, is an f-structure with a SUBJect, be it an open or a closed f-structure. Obviation could be used to tell pronouns or pronominals obviative in a certain domain, an obviative proposition, that is a clause nucleus; however either formulations of obviation do not account for the behaviour of NPs. No mention seems required for referential expressions at this level, where no mention is made about the antecedent.

3. F is an F-domain iff
        f-commands   in F and I is licensed

The second part of the formulation, says that the structure in which the antecedent and the anaphor must be bound is the one containing a SUBJect function - this is derived from the licensing condition: in an NP the F containing the head, in a clause, the F containing the SUBJect of the clause, in an ADJunct the one containing the PRO, in an open function, the open function itself.

4. F-command:

A function a f-commands a function   in F iff

a.   is not contained in , and   is not directly contained in ,   = SUBJect

b. every f-structure of F that contains   contains

b1.   may contain   in F iff   is in a weak RD

c. a function   is directly contained in a function   if   is a subsidiary f-structure of a function

{the subject is not accessible to itself - the remaining arguments/adjuncts of the head Noun may be bound by the subject; as well as the i-within-i reformulated}

In a., the antecedent/binder cannot be contained in the f-structure of its bindee, in other words, the relation is asymmetric; also the bindee cannot be directly contained in the f-structure of the antecedent but it must constitute a separate f-structure. This is trivial, but requires the formulation of a notion, "directly contained", which divides f-structures contained in complements and adjuncts of a head from their governors.

The b. clause only applies when the bindee is contained in the same F that contains the binder, but the binder is down in a separate f-structure which is open. However, for the licensing conditions on F given below, obliques are not regarded as possible F-domains.

5. Licensing conditions for an Indexing I of   with
   : 1. i. must be lexically free;
        ii. it is the SUBJect
        iii. it is in a strong RD
        iv. its   -role is superior in the following hierarchy:
   agent > benefactive > recipient/experiencer/goal > instrumental >theme/patient > locative

(iii. differentiates between an ADJunct PP and a predicative one, in the sense that the anaphor contained in an adjunct PP is bound to the SUBJect of the higher strong RD, whereas an anaphor contained in an open PP is bound locally to the closer function).

   2. otherwise,
   A. a function   is free in the discourse if F is a
      weak RD,
   B. a function   is coreferent/cospecified in the
      discourse if   is in a strong RD.

6. A function is lexically free iff,
   - it is argumental
   A function is lexically bound iff,
   - it is Ø - empty, existentially bound argument
   - it is an expletive (no PRED, but FORM)
   - it is a quasi-argument

7. A R(eferential)-D(omain) is an f-structure specified for referential energy:
   i. it is strong iff a. it is a closed function;
                       b. it is referentially transparent
   ii. it is weak iff   a. it is an open function;
                       b. it is referentially opaque.
   iii. Referential energy :
      a. for clause nuclei(where a SUBJect is obligatory) is expressed by atomic attribute-value pairs: TENSE=[±REF]{past tense individuates a specific reference time}, MODE [±REAL]{real mode is assertive and implies the truth of the proposition-at least on part of the speaker}, CLASS[±IMPLIC] {implicative verbs imply the truth of their complements and may be interpreted referentially - also factivity is included}, ASPECT [±PERF] {perfective aspect implies the existence in the world of the object predicated by the verb};
      b. for NP heads of relative predicative adjuncts CARD= [±DEF/Ø], INDIV [±SPEC], [±ref].
   c. transparency obtains whenever the features have positive value.

4. The algorithm for anaphoric control

Two structuress are built from the output of the grammar: annotated c-structures, i.e. a directed graph which can be traversed primarily through syntactic constituents; and a list of the functional schemata associated with semantic forms - in other words, all PRED expressions with a list of semantic attribute-value pairs, i.e. the f-structure mapped from the previous structure, where pronominal binding is computed. The algorithm applies to a completely parsed structure which is a graph translating the annotated c-structure of LFG into the f-structure. The algorithm uses the notions of domains used in LFG as well as functional information as to the grammatical function associated with a certain constituent, and its thematic role. The definition of domains is based crucially on the notion of f-structure and governors are derived from grammatical function and thematic role, as we shall describe in details below.

When a pronoun is encountered, the algorithm moves up to the left of its minimal domain, the closest f-structure containing it and stops in the first superordinate f-structure; on the contrary, with anaphors, the search is to

the left within the same f-structure containing it, unless it is contained in a SUBJect. It is worthwhile reminding that at f-structure level the VP node disappears and an OBJect NP appears at the same level of a SUBJect NP. F-structures contained in a nominal f-structure behave differently due to their grammatical function as discussed below.

In line with Bresnan et al(1985) and contrary to the proposal contained in Dalrymple(1990) we use functional features as lexically specified properties of individual anaphoric elements. These features both account for and translate lexical category, in this way directly triggering the binding algorithm that fires a certain procedure whenever a [+anaph] feature is met in the referential table associated to a certain f-structure. Features also serve to restrict the type of possible antecedents in terms of reference to the SUBJect; to set up a hierarchy for antecedenthood in which possible antecedents are ranked according to their associated grammatical function and thematic role; to unify morphological features checking for agreement in person and number, and selectional restrictions imposed by inherent semantic features; to tell apart quantifiers and quantified NPs which cannot be used as antecedents in backward pronominalization. A complete list of features is given below.

Whenever an antecedent is found - selected by the presence of the feature [+ref] - its ranking is checked as well as its features for agreement: the interaction with binding principles determines the possibility for an OBJect referential expression to act as binder of long distance anaphors. In other words, binding works by default according to the principle "bind anaphors as soon as possible". On the contrary pronominal coreference imposes the algorithm to pick up a certain referential expression as possible candidate and to reject other referential expressions owing to their ranking in the hierarchy. Only one antecedent is selected for [+ana] elements; with [+pro] more than one antecedent is selected according to the rules and to the antecedents available.

Whenever a pronoun is left unbound the algorithm adds an instruction "resolve(x)", which is used to trigger the anaphoric binding algorithm at discourse level(see Bianchi & Delmonte, 1989). The remaining pronouns and anaphors are assigned a couple of indexes: their own and the one of their antecedent and binder. Following recent work by Enç(1989) who discusses a pronominal system for natural languages made up of seven classes, we built one made up of four classes for Italian - Chomsky's system based on two classes, anaphors and pronouns is insufficient. To be added to these four classes - which include anaphors and nouns(common, proper) - there is one class for pleonastic lexically unexpressed pronouns constituted by a verbal agreement in Italian, deprived of deictic import. Pronouns can be lexically specified or not, this being expressed by a feature introduce in Bresnan(1982), [±MU] (Morphologically Unexpressed). Thus, big PRO's resulting from tense specification which can be subject to anaphoric control - in LFG PROs are structurally or lexically functionally controlled - are differentiated from little pro's by the fact that the former are marked [+ana], and the latter are marked [-ana]. These are differentiated from clitics and independent lexical pronouns by the fact of being [+MU], whereas the latter

are [-MU]. Besides, clitics are marked [+ana], whereas tonic personal pronouns are [-ana]. Epithets contain a deictic or a determiner feature specification. Pronominal quantifiers are marked [+pro] [±PART]. We give below a complete classification in features of all pronominal and nominal expressions as computed by the system, as a translation of lexical category together with features from SPEC, and NUMBER.

Table 2. Classification of pronouns anaphors and referential expressions
1. PROs[+ref,+pro,+ana,-def,+MU]
2. pros[+ref,+pro,-ana,+def,+MU]
3. clitics[+ref,+pro,+ana,+def,-MU]
4. lexical pronouns[+ref,+pro,-ana,+def,-MU]
5. epithets[+ref,+pro,-ana,±def,-MU]
6. common nouns[+ref,-pro,-ana,+class,±def,±sing]
7. partitive nouns[+ref, -pro, -ana, +class, +part, ±def, ±sing]
8. proper nouns[+ref,-pro,-ana,-class,±sing]
9. quantified NPs[+ref,-pro,-ana,±def,±part,±sing]
10. pron. quantifiers[+ref,+pro,-ana,±def,±part, ±sing]
11. null det. nouns[+ref,-pro,ana, +class, 0def, ±sing]
12. long anaphors [-ref,+pron,+ana,+SUBJCTV]
13. short anaphors [-ref,-pron,+ana, -SUBJCTV]

Other features will be attributed to nouns by their determiner: in particular articles are translated into [±DEF], numbers into [±CARD], quantifiers into [±PART]. The lack of determiner or the null determiner is marked by the presence of the feature [0 DEF]. The feature [±PART] is also assigned when a prepositional marker "di" is used to indicate an indefinite or a definite unspecified quantity (corresponding to the English "some, a (little) bit of". This information is recorded under a different functional node, the one named SPECifier, and are listed here only for convenience.

In addition, common nouns are differentiated from proper nouns by the feature +CLASS for the former and -CLASS for the latter, indicating that common nouns are used to denote classes or properties of individuals, as opposed to proper nouns which should pick out individuals. Moreover, common nouns are specified in reference by definiteness, whereas proper nouns use definiteness only redundantly - in Italian a proper noun may be preceded by a definite article. When a noun is recognized as proper, this feature is discarded. Proper nouns are assigned a higher score than common nouns, as candidates for antecedenthood. Cardinality is marked by Number, which adds the information that a Singular, Definite, Specific noun phrase is to be interpreted as a unary set of the class of objects or individuals denoted by the noun, i.e. there is only one member referred to by the noun phrase in universe of discourse that we want to pick up. Plural noun phrases are treated differently, i.e. as quantified NPs.

5. The Basic Algorithm

We list here below the basic algorithm in its Prolog formulation: as we said previous it applies on f-structures which are compiled as a directed graph, and accessed by an algorithm with performs graph search. The complete

algorithm is made up of about 4000 lines of program in Prolog.

F-structure
```
f_structure(Index,F_R,Node) :-
        node(Node):F_R:index:Index.
```
F-command
```
f_command(Alpha,Alpha_Funct,Beta,Level) :-
    f-structure(Beta,F,N), F=subj/_,
    node(N1):F1:node(N), F1 = subj/_,
    node(N2):F2:node(N1),
    f_c(N2,F2,Alpha,Alpha_Funct,O,Level_x),
    Level is Level_x + 2.
f_command(Alpha,Alpha_Funct,Beta,Level) :-
    f-structure(Beta,F,N), F=subj/_,
    node(N1):F1:node(N), F1 \ subj/_,
    f_c(N1,F1,Alpha,Alpha_Funct,O,Level_x),
    Level is Level_x + 1.
f_command(Alpha,Alpha_Funct,Beta,Level) :-
    f-structure(Beta,F,N),
    F1 \= subj/_,
    f_c(N,F,Alpha,Alpha_Funct,O,Level_x).
f_c(N,F,Alpha,Alpha_Funct,0,0) :-
            node(N):Alpha_Funct:index:Alpha,
            Alpha_Funct \= F.
f_c(N,F,Alpha,Alpha_Funct,Lev,Lev) :-  Lev > 0,
            node(N):Alpha_Funct:index:Alpha.
f_c(N,F,Alpha,Alpha_Funct,Lev,Level):-
    node(N1):F1:node(N),Lev1 is Lev + 1,
f_c(N1,F1,Alpha,Alpha_Funct,Lev1,Level).
```

And this is how the main algorithm is triggered by the presence of a certain feature in the referential table associated to a certain f-structure node:
```
resolve_anaphoric(Net,Index,WeightedList) :-
            node(Node):index:Index,
            node(Node):ref_tab:List,
             member(+ana,List),
bagof(Outref,refer(Node,List,Outref),Listref),
            maplist(scoring,Listref,WeightedList).
resolve_pronoun(Net,Index,WeightedList) :-
            node(Node):index:Index,
            node(Node):ref_tab:List,
             member(+pro,List),
bagof(Outref,refer(Node,List,Outref),Listref),
            maplist(scoring,Listref,WeightedList).
```

Now, consider how "se stesso" is bound:
```
refer(Node,[-ref,-pro,+ana,+me],Ante/N) :-
            node(Node):index:Ind,
            f-command(Ante,F_ante,Ind,N),N = 0,
             F_ante = subj/_,
                !.
refer(Node,[-ref,-pro,+ana,+me],Ante/N) :-
            node(Node):index:Ind,
            f-command(Ante,F_ante,Ind,N),N = 1.
```

Two examples are shown here: the first is a simple case of a possessive anaphor contained in a SUBJect NP of a psychic verb: f-command is used to raise the "proprio" out of the SUBJect f-structure and the presence of an OBJect Experiences triggers binding. In the second example the long-distance anaphor "proprio" is contained in the SUBJect NP of a sentential complement: only the SUBJect of the higher clause is chosen as antecedent; the nuclear NP OBJect is discarded from the list of possible candidates because it is an Unaffected Theme (in case it were an Experiencer it would have been included).

EXAMPLE 1. La salute della propria moglie preoccupa Mario (the health of "propria" wife worries Mario)
f-structure
```
Net   ex33
index:f2
pred:preoccupare
mode:ind
tense:simple/pres
sem_cat:psych/emot
subj/causer_emot:ref_tab:[+ref,-pro,-ana,+class]
        index:np34
        pred:salute
        sem_cat:state
        gen:fem
        num:sing
        spec:def:+
        subj/posses:ref_tab:[+ref,-pro,-ana,+class]
                index:np35
                pred:moglie
                sem_cat:human
                gen:fem
                num:sing
                spec:def:+
            subj/posses:ref_tab:[-ref,+pro,+ana,-mu]
                    index:np36
                    pred:proprio
                    gen:fem
                    num:sing
obj/experiencer:ref_tab:[+ref,-pro,-ana,-class]
        index:np37
        pred:mario
        sem_cat:human
        gen:mas
        num:sing
        spec:def:0
```
OUTPUT OF THE ANAPHORIC BINDER
Net index: ex33
TO RESOLVE: np36
CONTROLLED: nil
PRONOMINALS: np36[-ref,+pro,+ana,-mu]
F-COMMAND: np37/2
Possible antecedent/s of np36: [np37/101]

EXAMPLE 2: lui ritiene che la propria sorella ami Gino (he believes that "propria" sister loves John)
f-structure
```
Net ex42
index:f2
pred:ritenere
mode:indic
tense:simple/pres
sem_cat:attitude
subj/agent:ref_tab:[+ref,+pro,-ana,-mu]
        index:np4
        pred:lui
        sem_cat:human
        pers:3
        gen:mas
        num:sing
```

```
                case:[nom]
            spec: def:+
obj/prop:index:f4
        pred:amare
        mode:subjunct
        tense:simple/pres
        sem_cat:state/emot
        subj/experiencer:ref_tab:[+ref,-pro,-ana,+class]
                index:np11
                pred:sorella
                sem_cat:human
                gen:fem
                num:sing
                spec:def:+
            subj/posses:ref_tab:[-ref,+pro,+ana,-mu]
                    index:np12
                    pred:proprio
                    gen:fem
                    num:sing
        obj/theme_unaff:ref_tab:[+ref,-pro,-ana,-class]
                index:np13
                pred:gino
                sem_cat:human
                gen:mas
                num:sing
                spec:def:0
```

OUPUT OF THE ANAPHORIC BINDER

Net index: ex42

TO RESOLVE: np12,np4

CONTROLLED: nil

PRONOMINALS:[np4/[+ref,+pro,-ana,-mu],np12/[-ref,+pro,+ana,-mu]]

EXTERNAL(ex42,np4)

Possible Antecedent/s of np4: none

Possible Antecedent/s of np12: [np4/30]

6. More complex structures

6.1 Assigning Antecedents to Obviative Pronouns

Obviative pronouns in Italian can be subdivided into three different kinds: clitics, null Subject pronoun, lexical pronouns. Clitics are to be differentiated from lexical pronouns by two basic properties: they are unstressed and they can be bound in the syntax by a TOPic function. In case they are unbound at c-structure, they can be assigned an antecedent at f-structure. Lexical pronouns are always stressed, and can never be long-distance bound in the syntax. However, they can be used in doubling a local NP, as follows,

20) Il presidente ha promosso un candidato che lui, da semplice commissario, aveva bocciato.

/ The president passed a candidate which he, as a mere commissioner, had failed.

Lexical pronouns can also be used accross sentences or within the text, for contrastive or emphatic aims(see Bresnan & Mchombo(1987) on Chichewa). Finally, the Null Subject is lexically empty and behaves very closely to clitic pronouns: it can be bound in the syntax or be unbound and be assigned an antecedent at f-structure. Obviously, it cannot be stressed nor be used for emphatic, contrastive use nor for doubling. Being lexically empty makes it somewhat different from clitics in relation to the binding domain: it can be bound from within a complement clause or an adjunct clause by a lexical pronoun, but not by a common or proper Noun.

21) a. pro Ha detto che lui non verrà. / pro said that he will not come.

b. pro Ha detto che Mario non verrà.

c. pro Ha parlato di guerra perché lui ama le armi. / He has told about war because he likes weapons.

d. pro Ha parlato di guerra perché Mario ama le armi.

Only the a.- c. examples allow for coreferentiality between little pro and the lexical pronoun in the COMP - the lexical pronoun being also free to look for an external antecedent in the discourse. The same would happen in case a clitic was introduced in place of the lexical pronoun,

22) pro Ha parlato di guerra perché Mario lo conosce. / He told about war because Mario knows him.

If we front the adjunct clause, both the lexical pronoun and the clitic are available as antecedents of little pro; and also the common or proper Noun is available, since it f-commands it. However, the lexical pronoun is only available if a list of referents is intended and not to continue the discourse topic.

22) a. Poiché pro ama le armi, lui ha parlato di guerra.

b. Poiché pro ama le armi, la polizia lo controlla. / Since pro loves weapons, the police controls him.

c. Poiché pro ama le armi, Mario ha parlato di guerra.

It is a well known fact that adjunct clauses can be attached to a lower level, within a complement clause or they can be fronted therein, as in the following examples:

23) a. Gino ha detto che Maria verrà all'incontro dopo PRO aver parlato a Tom. / John said that Mary will come to the meeting after having talked to Tom.

b. Dopo PRO aver parlato a Tom, Gino ha detto che Maria verrà all'incontro. / After having talked to Tom, John said that Mary will come to the meeting.

The difference between a. and b. lies both in semantic interpretation and in the availability of antecedents for big PRO. As to semantic interpretation, the adjunct clause modifies the complement predicate in the a. example, and the matrix predicate in the b. example. As to binding of big PRO Mary will be the antecedent in a. example and John in the b. example. The skeletal f-structures for the two examples captures the different behaviour of f-command in a straightforward way:

```
23a.    SUBJECT: Pred: Gino
        PRED: DIRE <SUBJ, COMP>
        SCOMP: Pred: VENIRE <OBJ> SUBJ
                OBJ: Pred: Maria
                SUBJ: expletive pro
                ADJUNCT: Pred: Dopo
        SCOMP: Pred: PARLARE <SUBJ,OBLgoal>
                    SUBJ: PRO
                    OBL: Pred: Tom
23b.    ADJUNCT: Pred: Dopo
        SCOMP: Pred: PARLARE <SUBJ,OBLgoal>
                    SUBJ: PRO
                    OBL: Pred: Tom
        SUBJECT: Pred: Gino
        PRED: DIRE <SUBJ, COMP>
        SCOMP: Pred: VENIRE <OBJ> SUBJ
                OBJ: Pred: Maria
                SUBJ: expletive pro
```

In the a. example only Mary can be reached by f-command from the position of big PRO; in the b. example on the contrary, only John can be reached. The same behaviour can be predicted for little pro in tensed clauses. However, note the contrast with corresponding English complex sentences:

24) a. John beats her because he hates Mary

   b. Gino la picchia perché egli/pro odia Maria

   c. Gino la picchia perché Maria odia il gatto / John beats her because Mary hates the cat

As usual we indicate with italics purported coreference between the two items; now, whereas in the English example coreference between her in the matrix and Mary in the subordinate is possible, no such thing may apply to the corresponding Italian version, the b. example. Only the c. example allows it because the NP coreferent with the clitic pronoun is a SUBJect. Now, why the SUBJect should be privileged over the OBJect NP as possible antecedent for pronouns contained in a preposed subordinate clause? This is only explained in a theory of anaphora in discourse, and in particular by the fact that SUBJects are naturally used as topic of discourse or else some non canonical constituent order must be introduced in the sentence. For instance, in

25) a. Dopo che pro è arrivato, Maria ha sgridato Franco / After pro arrived, Mary scolded Frank

   b. Dopo che pro è arrivato, è stato sgridato Franco

   c. Dopo che pro è arrivato, Maria lo ha sgridato

coreference for little pro is only allowed in c.: the passive form with a postposed SUBJect does not permit the NP to be used as coreference, being computed as a FOCus. Being a FOCus requires a new topic of discourse to be set up and the previous references to be discarded.This is clearly shown by the specular structure in,

26) a. Dopo che è arrivato Gino, pro si è seduto. / After has arrived John, self sat down.

   b. Dopo che Gino è arrivato, pro si è seduto. / After John has arrived, self sat down.

   c. Dopo che pro è arrivato, Gino si è seduto. / After pro has arrived, John sat down.

where coreference in a. between Gino and pro is blocked because Gino is a focussed constituent and ARRIVARE has a lexical form with a focussed OBJect at lexical level(see Bresnan and Kanerva). When the OBJect/Theme is used as a SUBJect/Theme, however, coreference between the proper noun and the pro is possible, as shown by b.; the same applies to pro in the preposed adjunct clause and the proper noun as SUBJect of the main clause.

In order to cope with these facts, the algorithm must compute Obviation and from the obviative clausal structure see whether it can access another clausal structure at the same level or at a level below the one in which it is contained. This is done in our parser by a special procedure called "contains",

```
contains (index1,index2) :-
      node(node1):index:index1,
      node(node1):path(Bo):index:index2,
      node(node2):index:index2.
contains(index1,index2) :-
      node(node1):index:index1,
      node(node1):path(Bo):index2,
      node(node2):index:index2.
```

Here below we list the program predicate which takes care of little pros and possible antecedents contained in another clause:

```
refer(Net,Ind,[+ref,+pro,-ana,-me],Ante/N):-
      node(node):index:Ind,
      node(node):cat:features,
      node(node):num:number,
      find_gender(node,Gen),
      f_command(NAnte,F_ante,Ind,N),N > 0,
      f_structure(NAnte,F_ante,N_ante),
      not contains(NAnte,Ind),
      node(N_ante):F_sup:node(N2),
      node(N2):F/R:index:Ante,
      not node(N2):path(_):Ind,
      write(Ante/N),nl,
      node(N2):F/R:cat:Cat,
      features(Cat,features),
      node(N2):F/R:gen:Gen_ante,
 ((Gen_ante = Gen) ; (Gen = nil) ; (Gen_ante = nil)),
      node(N2):F/R:num:Num_Ante,
      number = Num_Ante,
      node(N2):F/R:ref_tab:List,
      poss_ante(Ind,Ante,List),
      non_referred_in(Ind,Ante).
```

6.2 Arbitrary or Generic Reading

All [+ana] marked pronouns do not possess intrinsic reference, being also marked [-ref] and two consequences ensue: they must be bound in their sentence and cannot look for antecedents in the discourse, unless there are additional conditions intervening, i.e. tense must be specific and not generic, and so on; they can be assigned ARBITRARY interpretation, when a controller is lacking, and a series of semantic conditions are met as to tense specification. Since ARBITRARY interpretation is a generic quantification on events this can be produced with untensed propositions or tensed ones, but with no deictic or definite import as shown by:

20)a. I think that [prop[+arbitrary]killing oneself is foolish]

   b. I think that [prop[+definite]killing oneself has been foolish]

Possessives pronouns are obviative according to whether they are contained in a predicative or open function. A further argument may be raised for Arbitrary PROs which in LFG are introduced each time the clause does not contain a controller because being a closed function it does not need one: we quote here Bresnan(1982,345) example, in Italian,

24) E' difficile andarsene./It is difficult to leave

where the infinitive "to leave" may be analysed as an extraposed COMP bound to the SUBJect. The PRO generated as SUBJect of the predicate "LEAVE" receives [arbitrary] interpretation. In general, reflexive pronouns lacking the ability to refer independently receive their reference from their binders: in case no binder is available reflexive pronouns are assigned arbitrary or generic reference. This may be detected both from structural cues and from properties associated with the predicate of the matrix clause. In 24 the copulative sentence is a typical case in question: the adjective "difficult" may or may not select a binder for the infinitive which should appear with

the preposition "for", thus turning the PRO from arbitrary to controlled,

24i. E' difficile per Gino andarsene/It is difficult for John to leave.

A similar case may be raised for anaphoric pronouns, whenever they are contained in a subject NP, as follows,

25) La propria$_{arb}$ libertà è una cosa importante/One's freedom is an important thing

The sentence contains a generic statement absolutely parallel to the reading of 24; the same happens whenever the anaphoric pronoun is contained in the subject position of a closed function like a sentential complement,

26) Marta$_i$ pensa che la propria$_{i/arb}$ libertà sia una cosa importante/ Martha thinks that one's freedom be an important thing

in a parallel way to the behaviour of PRO

26i) Mary thinks that [PRO to behave oneself is important.

We may note at this point the fact that English possessive pronouns behave in a different way from Italian ones: in particular "his" may be bound by a quantifier through PRO, and it may be taken to corefer to a non c-commanding NP, differently from what happens in Italian,

27) *La sua$_i$ salute preoccupa ognuno$_i$

28) PRO Knowing his$_i$ father pleases every boy$_i$

Conoscere proprio$_i$/suo$_x$ padre fa piacere a ogni$_i$ ragazzo

29) His$_i$ mother loves John$_i$ Sua$_x$ madre ama Gino$_i$

In particular, "his" seems to possess the ability to be bound by quantifiers like "proprio" does: in 28 the Italian version becomes analogous to the English one if we substitute "proprio" to "suo". In other words, Italian has two separate lexical pronouns for bound and unbound reference whereas English has only one and the conditions on binding are simply structural whereas in Italian they are both structural and lexical.The peculiarity of long-distance anaphors emerges from the dependency of binding on the presence of a feature at sentence level, the one related to the mood of the subordinate clause. In particular, as also detected in other languages (cf. Zaenen, 1983) the choice of Indicative vs. Subjunctive Mood is relevant for the binding possibilities of anaphors contained in the clause. The presence of the Indicative, in the most embedded clause, the one containing the long-distance anaphor seems to block binding from the matrix clause, as shown in:

30) Gino$_i$ pensa che tu sia convinto che la propria$_{i/*arb}$ famiglia sia la cosa più importante.

31) Gino$_i$ pensa che tu sei convinto che la propria$_{*i/arb}$ famiglia è la cosa più importante. /John thinks that you be/are convinced that self's family be/is the most important thing.

where we changed subjunctive in 30 to indicative in 31: only 30 allows binding, hence bound reference, and disallows arbitrary reference; on the contrary 31 only allows arbitrary reference i.e. no reference at all. As discussed at length in Zaenen(1983) the choice of the mood is bound by the matrix verb which permits only certain kind of referential acts to be realized by the complement clause. Being lexical, this information can be easily transmitted in features to the c-structure and percolated according to the usual LFG conventions(see

Giorgi,1984, for a lexical typology of the governing verbs).

The same applies to derived nominals like "suspicion" which can be the head of a sentential complement, inducing long-distance binding or preventing according to the presence of [+BOUND] feature,

32) Gino$_i$ ritiene che il sospetto di Carlo$_j$ che la propria$_{i/j}$ sorella sia un assassino abbia determinato la sua condanna.

33) Gino$_i$ ritiene che l'affermazione di Carlo$_j$ che la propria$_{*i/j}$ sorella è un assassino abbia determinato la sua condanna.

 / John believes that the Karl's suspicion that self's sister be/is a murdered had determined his/her trial.

6.3 Quantifiers and quantified NP's as antecedents

As a first approach to the problem of quantifiers, the algorithm takes care of precedence whenever a quantifed NP is indicated as possible antecedent for a pronoun. Quantified antecedents are individuated by the presence of the feature ±part in SPEC, as follows,

34) quantified(Ante) :- node(N):index:Ante,
          node(N):spec:part:_.

This predicate is used for quantified antecedents in a simple declarative with psychic verbs: as discussed above, binding of a possessive long distance anaphor can take place from a quantified antecedent contained at clause level.

However, when we want to deal with quantifiers and quantified NPs as possible antecedents of little pros, clitics or independent pronouns a different procedure must be called in, and is the following one,

35) a. non_quantif(Ante) :- node(N):index:Ante,
          not node(N):spec:part:_, !.
   b. non_quantif(Ante) :- node(N):index:Ante,
          node(N):spec:part:X,
          (X = '-'),
          node(N):spec:def:'+'.

This procedure is integrated into the predicate for referring clitics, in particular as follows,

36) refer(Net,Ind,[+ref,+pro,+ana,+me],Ante/N):-
          node(node):index:Ind,
          node(node):cat:features,
          node(node):num:number,
          node(node):gen:gender,
          find_gender(node,Gen),
          f_command(NAnte,F_ante,Ind,N),N > 0,
          f_structure(NAnte,F_ante,N_ante),
          not contains(NAnte,Ind),
          node(N_ante):F_sup:node(N2),
          node(N2):F/R:index:Ante,
          non_quantif(Ante),
          not node(N2):path(_):Ind,
          node(N2):F/R:cat:Cat,
          features(Cat,features),
          node(N2):F/R:gen:Gen_ante,
          node(N2):F/R:num:Num_Ante,
          number = Num_Ante,
          node(N2):F/R:ref_tab:List,
          poss_ante(Ind,Ante,List),
          non_referred_in(Ind,Ante).

In this way we can account for lack of coreference between a clitic pronoun contained in a fronted subordinate clause and a quantified NP contained in the main clause, as in the a. example

37)a. When I insulted him, every student went out of the room.

   b. When I insulted him, John went out of the room.

as opposed to the b. example, where coreference is allowed as usual. Here below we show the f-structure and the anaphoric binding processing results of the two sentences:

Net ex28
index: f1
main: index:f5
     pred:go_out
     mood:indic
     tense:past/simple
     cat:extensional
     aspect:accomplishment
     subj/agent:ref_tab:[+ref,-pro,-ana,+class]
               index:np6
               pred:student
               gen:mas
               num:sing
               pers:3rd
               spec:def:0
                    part:-
                    quant:every
     oblique/locative:ref_tab:[+ref,-pro,-ana,+class]
               index:np7
               pred:room
               gen:mas
               num:sing
               pers:third
               spec:def:+
adj:pred:when
   subordinate_clause:index:f3
     pred:insult
     mood:indic
     tense:past/simple
     cat:evaluative
     aspect:achievement
     subj/agent:ref_tab:[+ref,+pro,-ana,+me]
               index:np4
               pred:I
               gen:nonspec
               num:sing
               pers:first
               spec:def:+
     obj/theme_affect:ref_tab:[+ref,+pro,-ana,+me]
               index:np5
               pred:him
               gen:mas
               num:sing
               pers:first
               case:acc
               spec:def:+
OUPUT OF THE ANAPHORIC BINDER
Net index: ex28
TO RESOLVE: np5
CONTROLLED: nil
PRONOMINALS:[np5/[+ref,+pro,-ana,-mu]]
EXTERNAL(ex28,np4)

Possible Antecedent/s of np5: none
Net ex29
index: f1
main: index:f5
     pred:go_out
     mood:indic
     tense:past/simple
     cat:extensional
     aspect:accomplishment
     subj/agent:ref_tab:[+ref,-pro,-ana,-class]
               index:np6
               pred:John
               gen:mas
               num:sing
               pers:3rd
               spec:def:+
     oblique/locative:ref_tab:[+ref,-pro,-ana,+class]
               index:np7
               pred:room
               gen:mas
               num:sing
               pers:third
               spec:def:+
adj:pred:when
   subordinate_clause:index:f3
     pred:insult
     mood:indic
     tense:past/simple
     cat:evaluative
     aspect:achievement
     subj/agent:ref_tab:[+ref,+pro,-ana,+me]
               index:np4
               pred:I
               gen:nonspec
               num:sing
               pers:first
               spec:def:+
     obj/theme_affect:ref_tab:[+ref,+pro,-ana,+me]
               index:np5
               pred:him
               gen:mas
               num:sing
               pers:first
               case:acc
               spec:def:+
OUPUT OF THE ANAPHORIC BINDER
Net index: ex29
TO RESOLVE: np5
CONTROLLED: nil
PRONOMINALS:[np5/[+ref,+pro,-ana,-mu]]
EXTERNAL(ex29,np4)
Possible Antecedent/s of np5: [np6/131]

This notion of binding relevant for long-distance anaphors is also important for quantifiers as discussed in another work(Delmonte, 1989), in particular the fact that pronouns embedded in an Indicative or [-BOUND] clause need referential antecedents and not arbitrary or generic ones, as shown by the pair

34) A woman requires/demands that many/every men be in love with her, *and John knows her.

35) A woman believes that many men like her, and John knows her.

in 34, in English as in Italian, the indefinite "a woman" is computed as generic in the main clause and the same happens to the pronoun "her" in the complement clause introduced by "that"; but the conjoined sentence is expressed in the indicative and requires a specific woman to be picked up for referring the pronoun "her", which in this case must be computed as referential and not as generic, so the sentence is ungrammatical. The opposite happens in 35, where the indefinite is taken to refer to a specific woman in the discourse, and the two occurrence of "her" to be bound to this individual. As clearly shown, the referential capabilities of pronouns are tightly linked to the ones of their antecedent: but the opposite may happen, i.e. the referential abilities of the antecedents are bound by those of the pronouns, and these in turn are conditioned by the referential nature of the RD- referential domain - in which they are contained: an [-BOUND] domain is one containing indicative mood and reference is free, whereas a [+BOUND] domain is one containing a subjunctive mood and reference not free but locally bound, for anaphors, or lacking in referential import for lexical pronouns.

7. Chains and Binding

As we know, when at c-structure level a syntactic variable is bound to a TOPic or a FOCus a chain is created, which essentially is a couple of f-structures carrying the same index. One of the two members of the chain - the tail, is the controlled or bound element: this is an argument function and carries a theta-role; on the contrary, the head of the chain, the controller or binder is a non-argument function and has no theta-role. At f-structure level, the chain counts as a single element, in other words, the head of the chain plays no independent referential role from its tail, which is the argument function. Thus a short anaphor can be bound by the tail of a syntactic chain if contained in the same clause. On the contrary the head of the chain, which is contained in the higher domain cannot be the antecedent of anaphors or pronouns. The head of the chain, in turn, can contain a referring expression, a quantified expression, a pronoun or an anaphor: in the latter case, the tail cannot act as an antecedent, being conindexed with an element which must be itself bound in some domain. The domain is the one of the tail to which the anaphor contained in the head of the chain must be bound. We shall discuss some examples, now:

36) a. Parlando di suo suocero, Nixon ha ordinato a Bush, che lo ascoltava, di lasciarlo perdere.

/ Talking about his brother-in-law, Nixon ordered Bush, who listened to him, to let him go.

b. A se stesso Franco crede che Tom non pensa e mai. / Himself Frank believes that Tom never thinks to.

c. Parlando di se stesso, Nixon ha detto a Bush che ama la propria famiglia. / Talking about himself, Nixon told Bush that he loves his own family.

Consider a. and the status of suo/his: it is contained in an OBLique/theme and as such it can either be bound to the local SUBJect, big PRO, which in turn being contained in an untensed adjunct is bound under f-command by the SUBJects of the matrix, or be free and be bound to the coargument of the matrix SUBJect, the OBJ2 "Bush". Now consider lo/him which is contained within the non-

restrictive relative clause: being a pronoun it is obviative within its minimal clause and must look in the higher f-structure, the matrix clause. At this level, two possible antecedents seem to be available: Nixon and Bush. However, Bush is already bound to the relative pronoun which is the SUBJect of the relative clause that contains the pronoun lo. Thus, it must be eliminated from the list of the possible candidates. In example b. a short anaphor se stessa/herself has been left dislocated and is thus bound to its bindee in the embedded clause: since the anaphor requires a binder, and the interpretation of the anaphor is derived from the location of its bindee, the antecedent of the anaphor should be found in its minimal clause. Tom is thus the binder of the anaphor and not Frank.

Finally, in the c. example, the anaphor contained in the adjunct clause is bound only to big PRO and this in turn is anaphorically controlled by the SUBJect of the matrix, Nixon. Differently from the pronoun in the a. example, the anaphor cannot pick Bush as its possible antecedent. Now consider propria/his own: the reportive verb of the matrix dire/say requires the matrix SUBJect to bind the lower little pro and thus to act as antecedent for the possessive anaphor.

The main predicate which spots chain members contained in a separate f-structure from the one containing the variable and the reflexive or pronominal element is non_referred_in, which we list here below:

```
non_referred_in(index,Ante) :-
            pair_level(index,ListPair),
            maplist(find_ind,ListPair,ListInd),
            not referenced(Ante,[],ListInd).
referenced(N,Path,ListPair) :-
            member(N,ListPair), !.
referenced(Npx,Path,ListPair) :-
     (antecedent(_,Npx,Np1);antecedent(_,Np1,Npx);
      controlled(Npx,Np1);controlled(Np1,Npx)),
      not member(Np1,Path),
      riferimento(Np1,[Npx|Path],ListPair).
find_ind(node/_,Ind):- node(node):index:Ind, !.
find_ind(node/_,nil).
```

This predicate deletes from the list of possible antecedents for lexical pronouns the Np head of the chain, and takes as local binder of a reflexive the controlled variable or tail of a chain.

Let's consider now more closely the English version for 36b., with examples taken from Barrs(1988). First of all, the English version which we repeat here below, where we indicate with superscripts the syntactic index and with subscript the anaphoric index,

36b. Himself$_{j/k}$, Frank believes that Tom$_k$ never thinks to e$_j$.

has a lexical anaphor "himself" which can be bound both by Frank and by Tom. This is not allowed in Italian: in other words, Italian requires the anaphor to be "reconstructed" back into the place from which it has been extracted to produce the Topicalized structure. This is possible by considering the variable as the tail of a chain and the topicalized element as its head. Barrs's examples are very similar (his 7a,42)

37)a. Which pictures of himself did John say Bob liked e?

b. Himself, he thinks Mary loves e.

in 37a, the sentence is ambiguous - either John or Bob may be interpreted as the antecedent of the reflexive, in the b. example binding by "he" is grammatical, however in the corresponding Italian examples, no such ambiguity may arise and the b. version becomes ungrammatical,

38)a. Quali foto di se stesso Gino ha detto che Bruno
ama e?

   b. *Se stesso, egli pensa che Maria ama e.

Ungrammaticality is readily explained by the fact that "se stesso" must be locally bound and "Maria" is not an adequate SUBJect binder because of failure of agreement features. Two cases suspend ambiguity: the anaphor is contained in a predicative function, an ACOMP, or there is an accessible SUBJect, and are illustrated by the following examples, (his 7b, 17)

39)a. Whose pictures of himself did John say Bob liked
e?

   b. How proud of himself did John say Bob became e?

In 39a. the possessive pronoun "whose" provides a POSSessor or a SUBJect for the binding of the anaphor in its minimal local domain; in 39b. the head predicate "proud" is a predicative function with a functionally controlled SUBJect which is lexically bound to the available SUBJect "Bob". This happens before f-structure is accessed, so that no more binding domains may be accessed. Barrs gives a version within Chomsky's(1986) "Barriers" framework and Higginbotham's(1983) Linking Theory which accounts for the same facts in a transformation model.

8. Current Status and Comparison with Related Work

In using f-structures rather than syntactic constituency, LFG makes it more natural and direct looking for information such as being the "subject of", a notion crucial for antecedenthood.

Each referring expression receives a separate treatment by the rules for binding according to its feature matrix, grammatical function, and thematic role. For instance, little pro and clitics are included in the same class, but their grammatical function is crucial for distinguishing among them in their ability to be bound by an antecedent: little pro's can only be bound by subject antecedents, or nominative ones, whereas clitics being assigned accusative, dative or oblique can never be bound by a subject antecedent.

A set of criteria for assigning priority scores to candidates for antecedenthood and binding are used in order to define what can be bound by what: candidates receive scores according to their grammatical function, SUBject scoring the highest; and to thematic role, agent scoring the highest, and so on. Exceptions are also individuated on the basis of the interplay of grammatical functions and thematic roles: for instance one such rule says that a possessive anaphor contained in a subject f-structure can be bound to a NP in its sentence unless it is a Theme. As appears, binding is crucially performed on a structural basis, rather than on a functional basis as the approach based on Functional Uncertainty would require. The structures involved are f-structures: the parser makes reference to the SUBJect a primitive notion which is used primarily to set NP f-structure apart from clausal ones;

untensed clauses may either appear as controlled complements, or as closed adjuncts or closed functions such as SUBJect: also in this case anaphoric binding applies as long as structural conditions allow it. In this sense, anaphoric binding together with syntactic binding are structurally determined and can be opposed to lexical binding which is entirely functionally determined. Scores are also very important and are based on the superiority hierarchy of theta-roles, and on the degree of referentiality a certain NP possesses.

In particular, the difference in binding domain existing between an anaphor like "himself" and a pronoun like "him" is obtained simply by reference to the level at which these two lexical items must start out looking for their antecedent: for the former it would be equal to 0, while for the latter would be equal to 1. Rather than formulating a "Coargument Disjointness Condition" it is sufficient to individuate a viable f-structure, which looks for the accessible SUBJect in the case of nominal ones and let the feature matrix do the rest.

As we saw, reference to the particular domain in which a certain element must be bound or be disjoint, and reference to the particular grammatical function the antecedent should bear in a particular environment is not sufficient to deal with the inventory of pronominals available in Italian and other languages: reference to the thematic role is sometimes required, whenever a psychic verb is used, as well as the type of quantified NP or quantifier that can become a candidate for antecedenthood in certain environments. Our systems does this directly by means of the feature matrix associated to the referential table and by directly investigating the content of the functional node, where theta-roles are available together with the function label. Possibly, the same result could be achieved by means of Functional Uncertainty, even though we have not tried to test this hypothesis.

However, let us consider why Functional Uncertainty has been introduced: basically because syntactic restrictions could be formulated in terms of grammatical functions, and could be expressed by the introduction of equation whose right-hand side member contained regular expressions like the following,

(37) (TOPIC) = (COMP* OBJ)

which refers to the analysis of Topicalization as discussed by Kaplan & Zaenen(1989). The equation specifies an infinite disjunction of paths within f-structures, paths involving zero or more COMPs: OBJ stands for the landing site or for the bindee for the binder. Using functional attributes makes things easier and does completely away with the need to keep in memory c-structure syntactic trees once they have been used to build the corresponding f-structures. I don't intend here to comment on Kaplan & Zaenen proposal, but simply to criticize Dalrymple's idea to use this procedure with some minor modification and adaption in anaphoric binding.

It is clear to me that the regularity of syntactic phenomena has a different nature from the one belonging to anaphoric ones. An equation like the one reported in (37) states that no matter what happens within the COMP, and as long as the landing site is an OBJ, any number of COMP's may be traversed in order to adequately bind the TOPIC. This never happens with anaphoric binding: even though the difference existing

between ADJunct clauses and COMPlement ones is relevant, the depth of embedding is also a crucial factor. Structural differences like the one existing between COMP and ADJ clauses are already taken care for by f-command: however, in order to let, say, a long-distance anaphor or a clitic pierce through, inside-out, more than one relevant domain, a number of conditions on antecedenthood and distance intervening between the anaphor and the antecedent must be also accounted for.

References

Barrs A.(1988), Paths, Connectivity, and featureless empty categories, in *Annali di Ca' Foscari* XXVII, 4, 9-34.

Belletti A., Rizzi L.(1988), Psych-Verbs and Theta-theory, *Natural Language and Linguistic Theory* 6, 3, 291-352.

Bratko I.(1986), *Prolog Programming for Artificial Intelligence*, Addison-Wesley Pub.Co.

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

Bresnan J.; Per-K.Halvorsen; J.Maling(1985), Logophoricity and bound anaphora, MS, Stanford University.

Bresnan J., Mchombo J.M.(1987), Topic, Pronoun and Agreement in Chichewa, *Language* 63, 4, 741-782.

Chomsky N.(1986), *Barriers*, MIT Press, Cambridge, Mass.

Dahl V.(1981), Translating Spanish into Logic through Logic, *American Journal of Computational Linguistics* 7, 3, 149-164.

Dalrymple M.(1990), Syntactic Constraints on Anaphoric Binding, Ph.D. Dissertation, Stanford University.

Delmonte R.(1985), Parsing Difficulties & Phonological Processing in Italian, *Proceedings of the 2nd Conference of the European Chapter of ACL*, Geneva, 136-145.

Delmonte R.(1989), Grammatica e Quantificazione in LFG, MS, University of Venice.

Delmonte R.(to appear)(1990), Semantic Parsing with LFG and Conceptual Representations, *Computers & the Humanities*, 5-6,pp.30.

Enç M.(1989), Pronouns, Licensing, and Binding, Natural Language and Linguistic Theory 7, 1, 51-92.

Higginbotham J.(1983), LF, Binding and Nominals, *Linguistic Inquiry* 14, 395-420.

Ingria R., D.Stallard(1989), A Computational Mechanism for Pronominal Reference, in *Proceedings of the 27th Annual Meeting of ACL*, Vancouver, 262-271.

Kaplan R., A.Zaenen(1989), Long-distance dependencies, constituent structure, and functional uncertainty, in M.Baltin & A.Kroch(eds), *Alternative Conceptions of Phrase Structure*, Chicago University Press.

McKeown K., C.Paris(1987), Functional Unification Grammar Revisited, in *Proceedings of the 25th Annual Meeting of the ACL*, Stanford, 97-103.

Pereira F.(1981), Extraposition Grammars, *American Journal of Computational Linguistics* 7, 4, 243-256.

Pereira F.(1983), Logic for Natural Language Analysis, Technical Note 275, Artificial Intelligence Center, SRI International.

Pereira F.(1985), A Structure-Sharing Representation for Unification-Based Grammar Formalism, in *Proceedings of the 23rd Annual Meeting of the ACL,* Chicago, 137-144.

Pollard C., I.Sag(1989), Anaphors in English and the scope of the binding theory, MS, Stanford University.

Sells P.; A.Zaenen; D.Zec(1987), Reflexivazation variation: Relations between syntax, semantics, and lexical structure. In M.Lida; S.Wechsler; D.Zec(eds) Working Papers in *Grammatical Theory and Discourse Structure*, 169-238, CSLI/University of Chicago Press, Stanford University, CSLI Lecture Notes, N.11.

Zaenen A.(1983) On Syntactic Binding, *Linguistic Inquiry* 14, 3, 469-504.