GENERATING & PARSING CLITICS WITH GETARUN

Rodolfo Delmonte
Section of Linguistics - D.S.A.O.
University Ca' Foscari - Ca' Garzoni-Moro

Abstract

In this paper we shall propose a principled solution to the problem of clitics from a computational point of view (Delmonte, 1992). In particular we shall deal with clitics in a specific Romance language, Italian, where they represent a highly ambiguous and ubiquitous morphological and syntactic problem. The solutions proposed will be in line with Monachesi, who however works only in a theoretical perspective (Monachesi, 1995). The syntactic framework that will be used is LFG theory: however, the details of the theoretical assumptions and the implementation can be transferred to other similar feature-based unification grammar formalisms like GPSG, HPSG, etc.
As to the computational implementation, we shall be referring to a system called GETARUN, which is equipped with three main modules: a lower module for parsing where sentence strategies are implemented (Delmonte, 1990; Delmonte and Bianchi and Pianta, 1992); a middle module for semantic interpretation and discourse model construction which is cast into Situation Semantics; and a higher module where reasoning and generation takes place.
We shall be referring both to the higher module and the lower one. In particular, the generation of clitics will be addressed first and then clitics parsing will be presented. We assume that from a psycholinguistic point of view, no grammar reversibility is allowed in the human language faculty, and that the language production processor is separate from the language understanding processor with which it shares a common lexicon. Language generation requires planning, a linguistic activity which is not required in the understanding process. At the same time, parsing requires setting up a number of disambiguating strategies, basically to tell arguments apart from adjuncts and reduce the effects of backtracking. The parser is a DGC depth-first highly deterministic parser equipped with a look-ahead mechanism and a system of failures recovery based on a Well Formed Substring Table (WFST) which implements a number of sentence parsing psycholinguistic strategies (Delmonte and Dolci, 1989; Delmonte and Dolci, 1997).

1 Introduction

Clitics constitute a challenge both from a theoretical and a computational point of view due to their linguistic nature, which requires the setting up of special procedures at all levels of linguistic analysis from the basic ones, the morphological and the syntactic level to the higher discourse related ones:

- in the anaphoric binding module
- in the generation module
- in the semantic interpretation module

Seen that not all languages make use of clitics, this then could be treated as a language-dependent specialization of the general language faculty both in production and understanding. In replying to general and fundamental psycholinguistic questions, we shall try to give a unified account of the way in which clitics behave starting from the generation level. However, our approach shall differ from the ones proposed within any theoretical generative framework where language understanding and production is intuitively surmised as a unified phenomenon. Thus we shall start by briefly discussing the question of clitic generation in our higher module, where reasoning, planning and generation takes place. Then we shall concentrate on the second theme of this paper, clitic parsing.

1.1 Generating Clitics

Trying to pinpoint the general question of whether it is computationally feasible and relevant to treat the generation/understanding dichotomy as if they were two facets of a unified and unification-based processor, there has been a number of attempts in the computational literature to propose "Reversible" grammars both for generation and analysis of natural language, which however we do not personally accept and find just too preliminary to present itself as a serious alternative to the widely used separation of the two processing mechanisms (Delmonte and Bianchi, 1998).
Clitics are activated at the morphological level whenever the generator requires a certain argument or adjunct to be pronominalized. In a language like Italian - and we shall be referring only to this language for lack of space - pronominal expressions can either be morphologically expressed or

unexpressed, thus resulting in an empty predicate which however is endowed with case (Delmonte, 1985; Delmonte, 1988; Delmonte, 1991a). There are two such empty pronominal expression, however only the first one can be regarded an empty clitic:
a. "little pro" (hence *little_pro*) which appears in c-structure with tensed verbal morphology hence it inherits morphological features;
b. and "big PRO" (hence *big_PRO*) which only appears in f-structure, with untensed verbal morphology.
Only *little_pro* qualifies as empty clitic due the presence of verbal morphology which is inherited from the tensed verb, and is unified to the empty SUBJect NP. As will be explained further on, this can qualify either as an argumental empty pronoun or as a non-argumental expletive, which in a language like Italian is the only means to fill a Subject NP in c(onstituent)-structure position in lack of morphologically expressed equivalent forms. On the contrary, *big_PRO* lacks morphological features and does not surface in c-structure. From a syntactic and anaphoric point of view, both pronouns can be regarded empty categories on a par with syntactic variables which arise due to the same need: that of generating a pragmatically appropriate informational structure which will allow the speakers to communicate smoothly their intentions.
It is in fact the same identical underlying principle that causes the presence of the two types of empty categories: syntactic variable are generated whenever a given sentential component has to be highlighted - i.e. focussed - or simply reasserted as relevant - i.e. topicalized. Exactly in the same way, pronoun generation is motivated by the need to establish a coherence link to current discourse structure: the topic; or to emphasize or reassert as relevant a previously asserted participant of some event which has been neglected by the interlocutor.
The difference between the two types of empty categories is not in the triggering generative component - the reasoning and the planning, but in the realization phase. Generation systems are usually equipped with a Planner and a Tactical or Realization Component where the actual generation of the output string takes place.

1.2 Tactical Component

It is generally agreed that a suitable input to the realization component must be constituted by some form of semantic representation which may include the actual lexical choice or some abstract conceptual representation of each lexical item for the final realization (Strzalkowski, 1994, Zajac, 1994, McDonald, 1994).
However, there are some differences that can be found between the approaches documented in the literature and ours. In our system, input to the realization has a general predicate-argument structure and a number of semantic and functional features associated that are used to guide the grammar to generate the most adequate structural configuration. Top-down semantic, rhetoric and pragmatic decisions are paired with bottom-up lexical requirements imposed by each predicate on the fly, while realizing each lexical item in each specific grammar rule. In particular, argument specification only reflects the order each argument has in canonical predicate argument structure. Syntactic non-canonical realizations, like for instance passive construction, expletive/pleonastic subject insertion, left-dislocation and any other possible grammatically relevant structural decision are left to the phrase structure rule component of the grammar to take. We shall discuss one simple example,

Ex.1: Maria che ieri lo cercava lo insultò / Maria who yesterday was looking for him, insulted him

```
Voice=act, Tense=past, Mood=indic, Modality=assert ,
Main_relation=insultare,
List_of_arguments=[
    First_argument=[prop, nil, sing, [maria,
    First_argument_modifier=
      [Voice=act,Tense=imperf,Mood=indic,
      Main_relation=cercare,
        Main_relation_modifier=[dtemp,ieri],
          List_of_arguments=[
            First_argument=[rel, nil, sing, maria],
            Second_argument=[top, nil, sing, mario] ]
    Second_argument=[top, nil, sing, mario] ]
```

where we show the output of the Planner, which is passed to the Tactical component of our generator. The output sentence is represented into its semantic and pragmatic elements which will be turned into the actual linguistic items in the Realization phase. The semantic and pragmatic features are organized as follows:
Input to our Tactical Component is as follows:
Voice: active/passive
Tense: any tense
Mood: any mood including imperative, interrogative etc.
Modality any modality

Main Relation: the main clause relation
Main Relation: ***Modification***
[Adverbial Phrase, Subordinate Clause, Coordinate Clause, Prepositional Phrase Predicative Adjunct]
List of Arguments: *First Argument*:
Subject argument - Sentential subject
Second Argument:
Object, Oblique, Sentential Object
Third Argument: IndirectObject or Oblique
Argument specifications 1. *Semantic Type*:
a. prop (proper name);
b. def (definite common noun);
c. ndef (indefinite common noun);
d. foc (focussed noun to be fronted by syntactic structures like left dislocation, it-cleft, topicalization, etc.);
e. top (topic noun - to be pronominalized);
f. rel (relative pronoun argument);
g. trace(controllee of syntactic or lexical controller);
i. pro(empty or lexically unexpressed noun);
Cardinality: a number/nil
Number: sing(ular)/pl(ural)
Head: lexical head *or* a concept
Argument specifications 2. *Modification*
[Adjectival Phrase, Prepositional Phrase, Predicative Adjuncts]

Consider now the need to realize one argument as clitic pronoun, as is required in Romance languages: the semantic structure would carry the information that the second argument of the predicate belongs to TOP type, as for instance in the argument representation for Mario, which is realized as the clitic pronoun "lo"/him independently by the grammar. The fact that Mario has been assigned the TOP type in the slot reserved for Definiteness does not depend on syntactic but merely on pragmatic and semantic information. Features for the choice of the adequate pronominal form are partially extracted from the lexical entry associated to Mario, which are Person=3, Gender=Masculine, Animacy=Human,

[top, nil, sing, mario] --> lo

In addition, Number is set to singular: as to Case, it is equal to Accusative owing to the fact that the argument is second in the list, and that the main predicate requires both a subject and an object to be lexically filled. The additional information that "lo" should be preposed to the verbal predicate is not encoded in the semantic structure but is independently imposed by the phrase structure rules associated to the "transitive verb" syntactic class, and the presence of a TOP referential attribute.

The actual choice of the morphological realization is dependent on other principles as also suggested by Monachesi (Monachesi, 1995; Monachesi, 1999). In particular, we would like to assume that there is a really uniform manner to choose the right realization from the list of allowable or better lexically available Italian clitics. We agree with Monachesi that clitic clustering cannot be accounted for by referring to their syntactic nature, but from their morphological nature. However rather than resorting to template morphology we assume that an approach in terms of allowable slots is much simpler to be motivated with a number of additional assumptions. Our slots originate in the morphology rather than in the syntax and are restricted in their interaction by semantic and phonological rules which may only operate in adjacency. Thus, the syntax is governed by morphology and the semantic information associated to the canonical argument position and from here passed to the lexical clitic form.

In particular, there are only two slots allowed; however bigger clusters can be built according to euphonic rules and a cyclic recursive rule which is restricted by performance factors alone. Thus, potentially, any number of clitics could be clustered seen that not only arguments but also adjuncts are cliticizable semantic linguistic units. However, it is a fact that clitics trigger rules for pronominal anaphora which are strongly dependent on performance factors in their execution (Delmonte and Bianchi, 1991). The framework we suggest is a four-ways subdivision of Italian clitics in Table 1.

In the table we indicated the lexically available clitics in bold and their secondary realization in italic due to the presence of a phonological rule of lowering. The rule states that the final /i/ is lowered into /e/ whenever a cluster with another clitic is produced and a new unstressed syllable is added. This syllable has to start with a sonorant and not with an unvoiced obstruent: in the latter case, the rule will not apply due to euphonic reasons.

The reasons why "ci, vi" must come last when clusters are built, - this is also the reason for the fourfold subdivision - is due to their semantic and referential nature as referring to non-human antecedents when preceded by another clitic in a cluster.

In sum, the internal ordering of clitics is determined by the following criteria:

a. occupying a morphological slot --> 1st and 2nd person come before 3rd person clitics, e.g. melo is allowed, but not *lome;

Tab. 1 Morphologically and Referentially Constrained Positional Slots For Clitic Clustering Rules

SLOT 1 1st & 2nd Person Singular & Plural		SLOT2 3rd Person Singular & Plural	
mi **ti**	*me* *te*	**lo** **la** **li** **le** **gli** **si**	 *glie* *se*
ci **vi**	*ce* *ve*		
		ne	

b. internal semantic ordering within a slot --> +human clitics come before, being referentially more important, e.g. glielo is allowed but not *logli;

c. apply the euphonic lowering rule for clitics occupying first position in a cluster when possible.

Performance factors are always at work: in the first slot, **vi** is never used as locative in combination with other clitics not to induce semantically ambiguous and phonetically unpleasant combinations.

1. io mi ci recherò.
   I myself there will render
   'I will render myself there.'
2. noi ci recheremo là.

≠ * noi ci ci recheremo.
* noi vi ci recheremo.
* noi ci vi recherete.
* noi ce vi recherete.
We ourselves will render there
'We will render ourselves there.'

The locative must be lexically expressed as an adverbial and may not be affixed as a clitic in front of the verb when in combination with noi/we, even though there seems to be no special restriction apart from performance related ones. Notice that the same does not apply to voi/you_plur

3. tu ti ci recherai.

= voi vi ci recherete.
You yourself there will render
'You will render yourself there.'

4. voi vi recherete là.
   You yourself will render there
   'You will render yourself there.'

Ci comes last due to semantic restrictions and no euphonic rule may apply to **vi** due to the presence of an obstruent.

These factors are not sufficient to justify the ordering rules within the second slot, where "si" can come before or after another clitic belonging to the same slot, according to whether it has an impersonal interpretation or not. Here it would seem that the use of the semantic referential feature is not sufficient to determine the position of "si", seen that both clitics refer to a +human antecedent. But the two combinations are not referentially interchangeable:

- **lo si** may only be analysed as containing an impersonal ***si***;
- ***se lo*** on the contrary may only be analysed as a case of non-impersonal ***si***, i.e. with reflexive, benefactive, pleonastic or idiomatic interpretation, which are all anaphorically bound;

We should also note the special status of "ne": it has a lexical realization which already starts out as phonologically biased towards its internal positioning - always in 1st position: in fact, "ne" may only follow other clitics in a cluster,

- ce ne, se ne, ve ne, etc.

where "ce/ve" are always referred to +human antecedents and "ne" is either partitive or locative. In this case, the phonological criterion is demoted/substituted by the referential criterion which outperforms the phonological ones and prevent the euphonic criterion to apply. In addition, clitics coming in first position are always subject to the euphonic rule. Also note that the only combination available with "si", ***se ne***, is ambiguous between the two interpretations available, the referential/argumental and the non-referential non argumental ones. We shall comment on these problems further on in the paper.

Where would this euphonic rule apply in the creation of a cluster: during the Realization phase of a generation process? This calls for a decision related to the need to keep the Phonetic component separate from the Lexical choice of the single word making up the generated string. In fact, in a system like ours which generates each single linguistic item while applying linguistic rules on the fly, a phonetic restriction like "clitic euphonic rule" may seem inappropriate for this level of computation. The first hypothesis is to apply phonological rules while generating the single linguistic items, seen that all linguistic items subject to the rule of "troncamento" share a strong

linguistic identity (Delmonte, 1981; Delmonte 1983).
It is interesting to note that some clitics - notably the accusative set - has the same form of definite articles. This might induce us to treat them all at the same level, in the generation process, seen that we need to produce "troncamento" by cutting the final middle vowel both on definite articles and on clitics. We know that phonological rules may only be activated in adjacency: this requires the two linguistic items to be lexically present in order to trigger the appropriate phonological process. However, since "troncamento" applies to the item that linearly precedes the governing head, this would have to be generated first and then followed by the generation of the specifier in a reverse order.
In the case of definite articles, the rule applies within the realization phase for the Noun Phrase: in order to activate "troncamento" the article needs to be generated after the noun, so that the phonology "knows" whether it starts with a consonant or a vowel, and in case it is a consonant what the starting sound is, what kind of consonant it is and if it is a dental fricative, whether there is a cluster: ***lo sci***mmione / the big monkey, ***lo str***umento / the instrument but ***il s***uono / the sound.
The phonology might however work at the end of the generation phase by adjusting the output string where needed. This could be done in view of the fact that clitics and articles have the same morphological realization, and that "troncamento" will produce in both cases the deletion required. However, this would require going over the output string in order to apply the phonological rule. On the contrary, by working at phrase structure level, it will be applied at the appropriate constituency level, i.e.

- NP for the article and other determiners and modifiers;
- PP for all preposition + article;
- IBAR for all clitic + tensed verb.

Thus it would seem that within a computational framework, theoretical generalizations are much more effective due to the need to take into account all the appropriate and available information at a given point of the computation. For instance, all information related to case and its associated syntactic constituency, would simply not be available within a generation module; consequently, it could simply not be used to constrain choice of realization of a given argument in the form of clitic.

2 Parsing Clitics

We shall now devote the rest of the paper to the analysis perspective within a sentence and text understanding psycholinguistic approach, using LFG as a theoretical framework (Bresnan, 2000). The parser we shall be commenting from is organized as discussed below.
It is a top-down depth-first DCG-based parser written in Prolog which uses a strong deterministic policy by means of a lookahead mechanism with a WFST to help recovery when failure is unavoidable due to strong attachment ambiguity.
It is divided up into a pipeline of sequential but independent modules which realize the subdivision of a parsing scheme as proposed in LFG theory where a c-structure is built before the f-structure can be projected by unification into a DAG. In this sense we try to apply in a given sequence phrase-structure rules as they are ordered in the grammar: whenever a syntactic constituent is successfully built, it is checked for semantic consistency, both internally for head-spec agreement, and externally, in case of a non-substantial head like a preposition dominates the lower NP constituent; other important local semantic consistency checks are performed with modifiers like attributive and predicative adjuncts. In case the governing predicate expects obligatory arguments to be lexically realized they will be searched and checked for uniqueness and coherence as LFG grammaticality principles require.
Whenever a given predicate has expectancies for a given argument to be realized either optionally or obligatorily this information will be passed below to the recursive portion of the parsing: this operation allows us to implement parsing strategies like Minimal Attachment, Argument or Thematic Precedence and other ones we don't have space to comment upon (but see Delmonte and Dolci 1989; Delmonte and Dolci 1997).
Important submodules are constituted by the Temporal Module, the Quantifier Raising Module, and the Anaphoric Binding Module: they interact partially to allow for a quantified expression to be realized as a referential or non-referential antecedent in the Anaphoric Binding Module - this will not be discussed for lack of space (but see Delmonte, 1991b; Delmonte and Bianchi, 1992).

2.1 Computational & Morphological Aspects

In most languages clitics are lexically and morphologically ambiguous: in Italian a word like "le" has the following triple interpretation:

LE-art, [fem, plur] / the – article, feminine plural
LE-clitac, [fem, plur] / her – clitic, feminine plural accusative
LE-clitdat, [fem, sing] / to her – clitic, feminine plural dative

To perform morphological analysis the parser calls a separate application IMMORTALE – Italian Morphological Tagger and Lemmatizer, which provides it with a complete lexical decomposition and interpretation of the input word into features and categories, which are then passed on to the syntactic module (Delmonte and Pianta 1996; Delmonte and Pianta 1998; Delmonte and Pianta 1999).

Word decomposition in IMMORTALE is based on morphemes: roots contained in a lexicon – they may be represented internally both as a list or in the more efficient letter-tree format; and affixes, which are not part of the lexicon and included in the main algorithm being bound morphemes: we divided up affixes into three main categories:

- inflectional morphemes which include all clitics;
- prefixes, a subclass of which are only verbal ones;
- derivational suffixes, which include all possible combinations.

Alterative suffixes are listed along with nominal inflectional suffixes.

When word decomposition takes place, each possible suffix is provided with a tag for each lexical category it can be legally affixed to. The word hypothesis is filtered out by a grammar, where rules and constraints apply in adjacency.

In the case of clitics, the main decomposition algorithm has to pass processing to a specific subroutine, in a recursive manner, in order to ascertain whether the current hypothesis can be confirmed. This would be the working of a right-to-left morphological decomposition processor which stops every time a possible affix has been individuated: clitic ambiguity calls for constraints which can be expressed at word structure level, non local and non adjacent as can be shown from the examples below:

PORTATI -->	port - a - ti Root – Infl - Clit	port - at - i Root - Theme - Infl	[+gli, *le] admis, inadmis
DALLA -->	d - a - l - la Root-Infl-Double-Clit	da – l - la Word-Double-Art	Dalla Name

Tab.2 Morphological Decomposition of ambiguous cliticized Italian words

In the first example, “portati”, two interpretations are available, the wordform being twice inherently ambiguous: thus the decomposition algorithm has to provide a first interpretation, that of a cliticized verb “portare”/bring in the imperative mood, second person singular,
(Root-port, Inflection-a, Clitic-ti),
This must then be followed by a second interpretation, in which the verb root “port”, is followed by the inflection of past participle treated as a single unit including both theme and morpheme, “-ati” which is then interpreted accordingly. The reason for compounding theme and inflection is simple: in this way the processing is reduced and more information is made available to the morphological rule component. The smaller inflection “i” for number and gender would be available as legal morpheme in case of absolute past participles, which are usually encoded with a different root in the lexicon, “portat”. It is interesting to note that “portatile” has only one meaning that of a portable computer and a single legal decomposition: “portatil-e”. In other words none of the two previous decompositions would have to be accepted: on the contrary, “portatigli” should be interpreted only as interpretation number 2, “Past participle, gli-clitic” and not as a double clitics which is only allowed in a closed set of combinations, like for instance “portatici”/bring-you-there.

The second word, “dalla” is three times ambiguous: however, both the proper name interpretation and the amalgam, preposition + article interpretation are made available directly as invariable and suppletive form. The remaining case, is represented by a special exceptional verb+clitic compound which is mediated by the presence of a doubling euphonic consonant. This phenomenon is allowed only in case the root is constituted by a single consonant: “d” for “dare”/give, “f” for “fare”/make and suppletive “v” for “andare”/go. Consonant doubling is allowed only for certain clitics and not for others: so one may have “dall-, dacc-, damm-, datt-, danne” but not *davv, *dass (but archaic yes).

As far as their phonological and lexical instantiation is concerned, in Romance languages – i.e. French, Italian, Spanish, Romanian, Portuguese – clitics are adjoined before the tensed verb or affixed to the end of verbs with indefinite mood, including past participle, gerund and infinitive (with

the exception of French, though). Clitics are adjoined after the tensed verb in Germanic languages. Even though our parser is multilingual, we shall restrict ourselves to presenting examples only from Italian.

The syntactic realization of clitics may be either NPs, PPs, depending usually on lexical case; pleonastic forms and expletives are bound to an empty function, Ø – the same as the unexpressed oblique agent of passive sentences. The clitic of there-sentences, for instance, is recovered at semantic level to reconstruct the relevant lexical form of be, "there_be", which takes on the meaning of a presentative sentence.

In LFG framework, the semantic role of a clitic must be determined by means of all its properties as derivable from the lexicon – case and form -; from lexical form or subcategorization frame of governing predicate; from language-dependent general grammatical properties (Bresnan, 2000). Generally speaking, clitics may assume all possible range of semantic properties: arguments, adjuncts, expletives, pleonastic forms. They are commented by the examples below:

• Argument

5. Gino lo conosce.
   John him knows
   'John knows him.'

• Adjunct

6. Gino ce lo mandò.
   John there him sent
   'John sent him there.'

• Adjunct

7. Ne arrivano molti.
   Of them arrive many
   'Many of them are arriving.'

• Pleonastic

8. Gino si arrabbiò.
   John "si" got angry
   'John got angry.'

• Ethic Dative

9. Gli si è rotto un braccio.
   Him "si" is broken an arm
   'He broke his arm.'

• Ergative

10. Si è rotto un bicchiere.
    "si" is broken a glass
    'A glass broke.'

• Passive-impersonal

11. Si sono costruite molte case.
    "si" are built many houses
    'Many houses have been built.'

• Reflexive-benefactive

12. Si sono costruiti la casa.
    Themselves are built the house
    'They have built their house'

• Impersonal (specific)

13. (In quell'anno) si costruì/costruirono molte case.
(In that year) one built_3rd_sing/plur many houses
'(In that year) one built many houses.'

• Middle-impersonal (generic)

14. Si dorme volentieri in aprile.
    One sleeps well in April
    'One sleeps well in April.'

• Reflexive-reciprocal

15. Due ragazzi si sono baciati.
    Two boys each other are kissed
    'Two boys kissed each other'

• Reflexive

16. Gino si rade ogni mattina.
    John himself shaves every morning
    'John shaves every morning.'

As appears from the above list, parsing "si" constitutes a real challenge, due to the variety of concurrent interpretations it may allow. We will thus concentrate on this type of clitic in the following section and only marginally comment on the remaining ones.

3 LFG Two-level Parsing

Italian is a highly structurally ambiguous or underdetermined language, so that semantic or thematic checking seems necessary at this level: in particular, long distance dependencies activate all kind of functional restrictions available, since they may be used to prevent backtracking which is time-consuming. We use Case, Gender and Person, as well as semantic categories of the bindee whenever available, to restrict the choice of the binder, as will be shown in detail later on (Delmonte, 1987).

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, 2000). In case also nouns are subcategorized for, the same requirement of coherence and completeness may be applied. Not all nouns however take arguments. In a language like Italian, at least three clause structural organizations are possible (Delmonte, 1999).

- a canonical organization, corresponding to the standard case in which constituents occupy their canonical positions; subjects come in preverbal position, objects and obliques in postverbal positions and adjuncts may alternate in preverbal or postverbal positions - although they may alternate freely also between verb and object NP and an extended number of intermediate positions;
- an inverted organization, corresponding to presentative constructions in which the subject occupies postverbal inverted position and an expletive may be present, "ci", or an oblique locative may be preposed in the subject place; or else nothing which relates to the arguments of the predicate be present in preverbal position. The latter case being allowed in Italian but not in other languages;
- a marked organization, corresponding to a complete reversal of constituents, allowed only in Italian, in which the OBJect NP comes in preverbal position and the SUBJect in postverbal position. Also in this case, the subject position may be treated as an empty category.
- Other specialized structures occur with psychic verbs which subcategorize for an open proposition, an infinitival clause as open complement; copulative constructions with a closed tensed or untensed proposition as subject which might be anaphorically controlled by an adjunct PP headed by "for". Also left dislocation constructions belong to this lot, with clitics as topic variables; topicalized impersonal structures, and other constructions.

This implies that the c-structure level rules sequence of the parser will actually be organized accordingly: adjuncts and the subject NP which constitute the high level of the grammar will be treated as universal rules, i.e. all languages will be allowed to take them at the same position. Thus the major differences encoded as parametric would typically be at sentential level rule ordering. However, some differences also apply at rule-internal level. One such wellknown case is the difference in the need to have the article lexically expressed before the possessive in Italian – be it adjectival or pronominal – apart from the exception constituted by family nouns. The nature of the complementizer which is the same word as the deictic/demonstrative in English, as opposed to the relative-clause introducer in Romance languages.

The most important case is however, the Aux-to-Comp rule, i.e. the raising of the auxiliary to the Complementizer at CP level and the main verb at IP level, with the SUBJect NP intervening in between. This case applies freely to yes-no-questions, but can only be made to apply in Italian non interrogative clauses with a gerundive adjunct, when the verb is not simple but compound. As a consequence, whenever the tensed verb has already been taken – and in German it can also be the main verb, not only the auxiliary! – it has to be copied in the appropriate variable in an IP level specific rule only accessible to German languages. VP rules will also have to take care of the fact that German may take the main verb as final element of the clause.

LFG provides the tools to build any c-structure grammar, in line with the X-bar system for the syntactic representation of constituency. As shown in Fig. 1 below, we introduced functional major constituents which identify the following sentence high levels:

- CP --> Spec, C'
  C' --> C, IP
  IP --> Spec=NP(subject), I'
  I' --> Inflected Tensed Verb Form, VP

According to this configuration, adjuncts and constituents like wh- words for questions and topicalized NPs, adjoined at sentence level, will be computed at first in a CP constituent and then passed down to the lower level of analysis. When IP is reached, the NP subject or sentential subject should be computed: at this point there are at least two possible parsing strategies to be followed, both theoretically plausible. The former is in line with LFG traditional view that no empty category should be produced unless it is strictly required by language typology. The latter is in line with the necessity to pose a basic structural configuration and goes against the tenet that assumes that the task of any grammar is that of encoding the surface structure organization of the language. In the former case no empty subject NP should arise in case the structure to be analysed is an inverted construction: this is justified by the fact that the Subject NP is actually to be found in inverted VP internal, or VP adjoined position. Since no NP movement is postulated in LFG there would be no possibility to adequately bind the empty category previously generated in

preverbal position, so little_pro is usually treated on a par with big_PRO and is generated at f-structure not at c-structure (Delmonte, 1991a).
On the contrary, we fill subject position with an empty category which can be erased (treated on a par with empty expletives) when parsing the actual lexical subject NP in postverbal position. In a sentence like "Arriva Gianni"/ Arrives John, the rule for marked presentational IP must be accessed first. In case no lexical NP is present – and this is the case of "Arriva"/Arrives, there are still two possibilities: we either have a canonical structure with an empty little pro as subject, or we have a fully inverted structure.
At first we assume that no subject is available and try to compute an inverted Subject: clearly this might fail, in case the NP computed in the VP is not interpretable as Subject but as Object of the main predicate. However, we take the marked option to be less frequent and less productive than the other way round: not every verb class may undergo subject inversion, which is not completely free (Delmonte, 1991b; Delmonte, 1987). And even if it does, there is quite a number of restrictions that may be made to apply to the inverted subject, as to its referential features (definiteness, etc.), which do not apply to the canonical object NP.
As can be gathered, there will be only one general procedure for parsing grammatical and ungrammatical sentence structure, which postulates the existence of a subject position to be filled either by lexical material or by an empty constituent. In case the sentence starts with a verb we let typologically determined parameters decide whether it is possible to build an empty subject NP or not: in case we are parsing Italian texts, this parameter would be active, but in case we are parsing a text belonging to Germanic languages, it would be deactivated - all exceptions can and must be treated locally. When we generate an empty category in subject position it remains to be decided what to do with it: in case a lexical NP in postverbal position is computed, and this is interpreted as the actual Subject function of the sentence, the trace shall be discarded as commented below. To do this, the interpretation component is left to decide how structures can be unified with the grammatical requirements imposed by lexical forms, i.e. uniqueness, completeness and coherence.
C-structure building in our parser corresponds to a partial interpretation of each constituent: in fact, when a parse is completed, we assign a structurally determined grammatical function label which could match semantic checking procedures performed when annotated c-structure is built, or it might be rejected as semantically inappropriate, due to selectional restrictions associated to that NP. Grammatical functions assignment is required in all cases in which a presentational construction has been parsed: it is just on the basis of the structural position of a given constituent, the postverbal NP, that we know what is the pragmatic import of the entire utterance. And this will be registered only in the grammatical function assigned to one of the arguments of the predicate, which is computed either as Subj_Foc, or Subj_Top according to whether it is an indefinite or definite NP respectively. The empty NP subject is not bound to the actual lexical NP found in inverted position, and it is simply discarded from the final representation. In this way, the annotated c-structure outputted by the parser is CP rewritten as VP, but the postverbal subject is associated with an adequate grammatical function. Backtracking is thus totally eliminated, and there is only one single phrase structure rule which applies to all sentential structures.
After CP has been correctly built, we activate the call to IP where subject NP and negation may be parsed; then a call to i_one_bar, will activate calls to Clitics and Infl, for all inflected verbal forms. The call to Clitics, is allowed both for German and Italian; it exceptionally applies to English "there", provided no NP subject has been analyzed. Infl is a call which is specialized for different languages and the subsequent typologically marked constructions of Italian.
Parsing the appropriate VP structure requires the instantiation of the appropriate syntactic verb class – as encoded in the lexicon - of the main predicate (Delmonte, 1989; Delmonte, 1997). in this case, it may either belong to the class of psychic or copulative verbs. Theoretically speaking, c-structure is now represented with a verbal phrase which contains no verb, which has been raised to infl, in case it is a tensed finite verb. We also apply semantic-aspectual constraints at this level that allow the call for inchoativized verb_phrase with the appropriate aspectual class, that of achievements; in addition, Subject NP should be empty, in Italian.
All subject inverted constructions are constrained by a check on the subject NP: it must be an empty category. This check also applies to impersonal-si constructions and to dislocated constructions. In this way, no backtracking will be allowed. In addition, syntactic category of the main verb should always be checked accordingly. The call to intransitive verbal phrases is subsequently further split into four syntactic classes: unaccusative,

inergative, atmospheric and impersonal. Inchoative constructions and impersonal-si constructions are also typologically marked, since they are only allowed in Romance languages; as well as fully inverted transitive constructions and intransitive reflexive structures which are only present in Romance languages. All these lexical classes share a number of common properties like for instance the possibility to use an empty expletive which will be interpreted as argumental for unaccusative and inergative and as non-argumental for the other two. Also remember that Italian verbs select auxiliaries according to syntactic class! In particular, unaccusatives only allow "essere"/be and unergatives usually require "avere"/have unless other conditions are applicable.

3.1 Unifying Annotated C-Structure

Clitic pronouns in preverbal/postverbal position are copied from IBAR to VP complement constituency level: this is done both in order for the clitic to be assigned to the appropriate syntactic constituency - NP or PP, and to receive an interpretation.
Each clitic is given the appropriate constituency, basically NP or PP, either as adjunct or as argument of the governing predicate. Note that the presence of a dative clitic requires the presence of an accusative clitic or a corresponding NP in case the governing predicate belong to the transitive syntactic class. Also reflexive interpretation depends on verb typology: not all transitive verbs can be made reflexive. In our computational lexicon, we classified verbal predicates as belonging to 9 syntactic classes to induce verb guidance and improve determinism: transitive, unaccusative, inergative, copulative, atmospheric, reflexive, inherent_reflexive, psych_transitive, psych_intransitive.
However, in order for the clitic to be assigned the appropriate constituency ambiguity must be solved

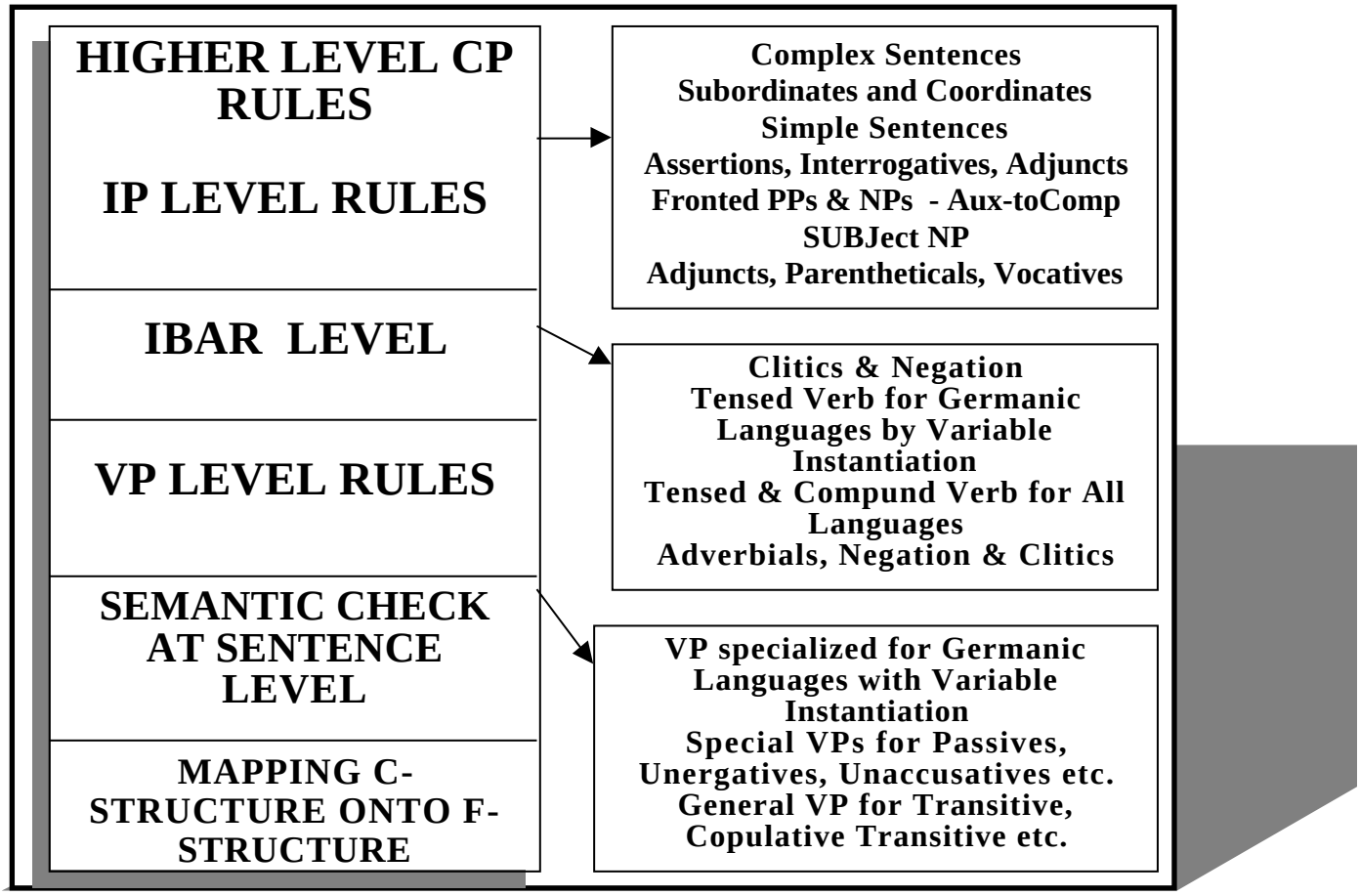

Fig. 1 Rules Sequence of the DCG Grammar Traversed by GETARUN Parser

first. All clitics are taken at the same c-structure position, i.e. there is only one rule applying in the grammar, and no difference existing between say Germanic vs. Romance languages: in Germanic

languages, tensed auxiliaries or main verbs will be analysed in CP position, before the Subject NP with inverted structures, and be copied down into a special VP constituent.
Clitics will be passed down to the VP, be it nominative, as is the case with impersonal SI or be it non nominative. As a matter of fact, no information is available as to case assignment when SI is parsed being strictly dependent on the type of governing predicate and number of arguments available. In other words, with SI case may only be assigned at the semantic interpretation level, when Grammatical Functions are actually associated to syntactic constituency (Delmonte, 1999). This can only take place when the simple sentence has been fully parsed – complex sentences are parsed recursively. No grammaticality principles can be applied while parsing each single constituent, which are however checked for internal semantic consistency – basically head – modifier, or head – complement consistency check, which in case of nouns and adjectives can never be obligatory, thus preventing the check for completeness. Thus in order for clitics to be adequately interpreted, a full parse at sentential level must be performed. Each VP rule will then typically contain a call to the interpretation level in terms of f-structure annotations which are then added to the c-structure level. The mapping from annotated c-structure to f-structure is then only a matter of variable instantiation.

3.2 Semantic Interpretation

Each clitic is assigned to a Grammatical Function and inherits default semantic features from subcategorization frames, and a semantic role. In GETARUN this is done by converting the annotated c-structure into the corresponding f-structure after grammaticality well-formedness principles have been checked. Features are usually associated with clitic interpretation and they come about partly from properties of the governing predicate and partly by structural conditions. In particular the following may be the relevant feature structures associated respectively to complement clitics like "lo", adjunct clitics like "ne", and quantifier-like clitic SI.

lo: CL, (↑ PRED) = [+PRO, +ANA]
(↑ CASE) = ACC
(↑ SPEC) = [-PART,+DEF,+REF]
(↑ NUM) = SG
(↑ PERS) = 3
(↑ GEND) = MAS

ne: CL, (↑ PRED) = [+PRO, +ANA]
(↑ CASE) = [OBL]
(↑ SPEC) = [+PART,-DEF,+REF]
(↑ PERS) = 3
(↑ NUM) = PL

si: CL, (↑ PRED) = [-PRO, -ANA, +SUBJTV]
(↑ CASE) = [NOM]
(↑ SPEC) = [+PART,-DEF,-REF]
(↑ PERS) = 3
(↑ NUM) = PL

NE and SI clitics constitute special cases. In particular, when assigned nominative case, SI clitic induces impersonal interpretation which can either be interpreted as a case of generic or arbitrary reading, or as a case of quasi-existential impersonal reading. Features associated to SI are different from other clitics not only in attribute value for CASE: PRED value is a non pronominal, non anaphoric, plus subjective clitic. All SI clitics share this property of being referred to the subject which is also assigned to long and short anaphoric pronouns in Italian, like "sé"/"se stesso", but not to possessive "proprio" being -SUBJCT. SI shares features in common with NE, i.e. the fact of quantifier-like properties testified by the presence of +PART and -DEF, two values associated also to real quantifiers.
As to the remaining SI cases, inherent reflexives and ergative SI cause the deletion of the lexical element from the f-structure. Both cases can be adequately analysed by means of lexical information, respectively by referring to the syntactic class for inherent reflexives and to the aspectual class for ergative or inchoative constructions. Reflexive and reciprocal SI interpretation may only arise on the basis of lexical information, however these cannot be regarded as unambiguous cases so they will be included in the following section where we discuss how to produce preferred SI interpretations.

3.3 Ambiguous SI Interpretation

Generally speaking, whenever lexical information is of no use to disambiguate SI interpretation, we should speak in terms of preferred interpretation. However we assume that Ethic Dative and Reflexive Benefactive interpretations are assigned every time the verb belong to the transitive class, the OBJect NP is lexically expressed in nominal or pronominal form. Relevant examples are the following:

17a. I bambini si lavano le mani.
The children themselves wash the hands
'The children are washing their hands.'
17b. I proprietari si sono costruiti la casa.
The owners themselves are built the house
'The owners have built their house.'
17c. A Gino gli si è rotto un vaso sul piede.
To Gino him 'si' is broken a vase on the foot
'To John a vase broke on his foot.'
17d. I prigionieri si sono liberati i piedi.
The prisoners themselves are freed the feet
'The prisoners untied their feet.'
17e. Gli si è rotto un braccio.
Him 'si' is broken an arm
'He broke his arm.'
17f. Si dorme di notte.
One sleeps of night
'One sleeps at night.'

where we included inergative impersonal construction, 17f. In all the other cases the clitic SI may be semantically interpreted as a POSSessive as the rough translations indicate. On the contrary whenever one of the arguments is not lexically expressed, with the exception of 17c. and 17e. which are no longer feasible, all the remaining examples are ambiguous between a reflexive and an impersonal interpretation:

18a. I bambini si lavano
The children are washing themselves
18b. Si sono costruiti/costruite le case.
(they) Themselves are built_mas_pl the houses
'They have built their houses.'
'(they) Themselves are built_fem_pl the houses.'
≠ One has built the houses.
"si" are built_fem_pl the houses
≠ 'The houses have been built.'
18c. I prigionieri si sono liberati.
The prisoners themselves are freed
'The prisoners freed themselves.'

In 18b, agreement can disambiguate the interpretation. It is interesting to note that the two other examples – 18a, 18c - are all interpretable as possible reciprocal where the expressed NP can either be the AGent or the AFFected-THeme. In other words, they belong to the semantic class of reversible predicates in which each of the argument has the same selectional restrictions. The only exception being the BUILD predicate which has two different interpretations available according to whether the past-participle agreement goes with the SUBJect/masculine or with the OBJect/feminine. In the former case, (agreement with the SUBJect) we get a BENEFactive interpretation of SI; whereas in the latter case, (agreement with the OBJect) we get an IMPersonal passive interpretation of SI. On the contrary, a fully ambiguous sentence is constituted by 18d. where agreement is no longer available and the SI can be both an IMPersonal and a BENEFactive.

18d. Poi, si costruirono le case.
Then, (they) one/themselves built the houses
'Then, one/they built the/their houses.'

As a matter of fact, it is because of the little_pro that the interpretation may vary: in case there is a discourse level antecedent who is the Main Topic to which the empty pronominal corefers, then the SI is no longer allowed to absorb one semantic role, that of the AGent which is no longer free. Impersonal interpretation would then be barred. However, this would oblige us to use discourse level information in order to assign the adequate sentence level interpretation, which in our case, also corresponds to two different f-structures - assuming that c-structure, where the SI clitic is computed does not change. It is important to note that since we use Semantic Roles to help find and disambiguate antecedents in the Anaphoric Binding module, they are already introduced in the f-structure level, i.e. before the semantic structure is built (Delmonte and Bianchi 1991).
Thus, in order to leave little_pro free to be bound at discourse level this should be associated with the SUBJ/nom assignment and the SI would be associated to the POSS modifier constructed within the OBJ le case/the houses as follows:

A. Interpretation 1: Benefactive SI

```
pred:costruire
lex_form:[np/subj/agent/[human],
np/obj/theme_affected/[object]]
mood:ind
tense:past
cat:accomplishment
subj/agent: index:sn1
            cat:[human]
            pred:little_pro
            gen:mas
            num:plur
            pers:3
            spec:def:'+'
           tab_ref:[+ref, +pro, -ana, -me]
obj/theme_affected:index:sn4
        cat:[object]
        pred:casa
        pers:3
        gen:mas
        num:plur
```

spec:def:'+'
mod/poss:indice:sn7
cat:[human]
pred:SI
spec:def:+
controller:sn1
tab_ref:[+ref, +pro, +ana, +me]
tab_ref:[+ref, -pro, -ana, +class]

B. Interpretation 2: Impersonal SI
pred:costruire
lex_form:[np/subj/agent/[human],
np/obj/theme_affected/[object]]
mood:ind
tense:past
cat:accomplishment
subj/agent: index:sn1
cat:[human]
pred:SI
gen:mas
num:plur
pers:3
spec:def:'-'
quant:exist
part:'-'
tab_ref:[+ref, -pro, -ana, +me]
obj/theme_affected:index:sn4
cat:[object]
pred:casa
pers:3
gen:mas
num:plur
spec:def:'+'
tab_ref:[+ref, -pro, -ana, +class]

Referential Tables are automatically generated for each nominal/pronominal expression to be used by the Anaphoric Binding algorithm. Whereas the quantified SI is not treated as a pronoun, the benefactive SI is treated as a CLITic.
The contribution of tense should also be considered in the disambiguation process: if we want to express generalizations, present tense should be used as the following examples show:

19a. Qui gli spaghetti si mangiano alla carbonara.
Here the spaghetti one eat “alla carbonara”
‘Here one eats spaghetti "alla carbonara".’
19b. Qui gli spaghetti si sono mangiati *(spesso).
Here the spaghetti one are eaten *(often)
‘Here one has eaten spaghetti *(often).’

where impersonal and middle SI both receive case and grammatical function SUBJ/nom but a different semantic value: middle interpretation corresponding to generic quantification and impersonal interpretation corresponding to a quasi-existential interpretation. We assume that middle interpretation may only arise whenever an appropriate frequency adverbial is lexically expressed to modify the event. Other non-ambiguous interpretations are assigned with Inverted Impersonal Constructions as the following ones:

20a. Si vede/è visto/vedono/sono visti spettacoli stupendi.
One sees/is seen/see/are seen_mas_pl shows fantastic
‘One sees/has seen fantastic shows.’
20b. Si affonda/affondano le navi.
‘One sinks/sink the ships.’
20c. Si costruì/costruirono le case.
One built the houses.

When interpreting complements of transitive structures:
- Use Selectional Restrictions to interpret Benefactives;

In case there is such a SI clitic, we build a POSS empty function which is coindexed with the clitic, then we assign BENEFactive Semantic Role;
- Use Selectional Restrictions to interpret Impersonal Inverted Constructions;

In case there is such a SI clitic, and a little “pro” as syntactic SUBJect, discard it (empty expletive), then assign a new quantified SUBJect with a SI pred and the Semantic Role of the governing predicate.

3.5 NE interpretation

We deal with this clitic as if it were a pronoun which constitutes the restriction of a partitive quantifier and binds its reference as in:
21a. "Ne_j ho conosciuti molti $Ø_j$.
Of them I have known many.
‘I knew many of them.’
21b. "Ne_j sono arrivati alcuni $Ø_j$.
Of them are arrived some
‘Some of them arrived.’
When interpreting complements of transitive structures, we check for the Object constituent and in case it is a partitive quantifier, we look for NE clitic and assign it to an open adjunct; then we build an empty adjunct in the restriction of the quantifier and bind it syntactically as shown in the example reported here below, "Gino ne conosce molti"/John of_them knows many.

[gino, ne, conosce, molti, .]
index:f2
pred:conoscere
lex_form:[np/subj/experiencer/[human],
np/obj/topic/[any]]

mood:ind
tense:pres
cat:ment_act
subj/esperiencer:
index:sn1
cat:[human]
pred:gino
gen:mas
num:sing
pers:3
spec:def:'0'
tab_ref:[+ref, -pro, -ana, -class]
obj/topic:
index:sn4
cat:[any]
pred:molto
pers:3
gen:mas
num:plur
spec:def:-
part:+
quant:molto
obj/theme:indice:sn7
cat:[any]
pred:pPro
spec:def:+
controller:sn2
tab_ref:[+ref, +pro, +ana, -me]
tab_ref:[+ref, -pro, -ana, +me]
qmark:q1
adj/theme:index:sn2
cat:[any]
pred:ne
pers:3
gen:mas
num:plur
case:[obl]
spec:part:+
tab_ref:[+ref, +pro, +ana, +me]
aspect:activity
rel1:[td(f2_free_a11)=tr(f2_free_a11)]
rel2:[included(tr(f2_free_a11), tes(f2_free_a11))]
definiteness:+
ref_int:[tr(f2_free_a11)]
qops:q1

4 Conclusions

In this paper we presented a general framework for the computation of clitic pronouns from the point of view of a linguistic theory LFG which aims at a psycholinguistic plausibility. Seen our concern with psycholinguistic modeling we justified our approach by discussing morphological features of clitics withing a "sentence generation" framework; on the contrary, syntactic and referential properties of clitics have been presented within a "parsing" scheme. This has been made possibile by referring to our system for text understanding and generation called GETARUN, which can be tested on the web, in our website, at http://byron.cgm.unive.it. There are a number of generalization which can only be arrived at from a computational perspective, which allows to simulate the actual internal psycholinguistic processing that might take place whenever a clitic has either to be generated or parsed.

5 References

Bresnan J.(2000), Lexical-Function Syntax, Manuscript.

Delmonte R.(1981), An Automatic Unrestricted Tex-to-Speech Prosodic Translator, Atti del Convegno Annuale A.I.C.A., Pavia, 1075-83.

Delmonte R.(1983), A Phonological Processor for Italian, Proceedings of the 2nd Conference of the European Chapter of ACL, Pisa, 26-34.

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.(1987), Grammatica e Ambiguità in Italiano, Annali di Ca' Foscari, XXVI, 1-2, 257-333.

[26] Delmonte R.(1999), Grammar and Structure, BULAG 24, 19-37.

Delmonte R.(1988), Focus and the Semantic Component, in Rivista di Grammatica Generativa, 81-121.

Delmonte R.(1989), From Subcategorization Frames to Thematic Roles, in R.Delmonte(ed.),Lessico, Stutture e Interpretazione - Studi Linguistici Applicati I, Unipress, Padova, Chapt.IV,167-212.

Delmonte R.(1990), Semantic Parsing with an LFG-based Lexicon and Conceptual Representations, Computers & the Humanities, 5-6, 461-488.

Delmonte R.(1991a), Empty Categories and Functional Features in LFG, Annali di Ca'Foscari, XXX, 1-2,79-140.

Delmonte R.(1991b), Grammatica e Quantificazione in LFG, Quaderni Patavini di Linguistica, 10, 3-71.

Delmonte R.(1992), Linguistic and Inferential Processing in Text Analysis by Computer, UP, Padova.

Delmonte R.(1997), Lexical Representations, Event Structure and Quantification, Quaderni Patavini di Linguistica, 15, 39-93.

Delmonte R., D.Bianchi(1991), Binding Pronominals with an LFG Parser, Proceeding of the 2nd International Workshop on Parsing Technologies, Cancun (Mexico), ACL 1991, pp. 59-72.

Delmonte R., D.Bianchi(1992),Quantifiers in Discourse, in Proc. ALLC/ACH'92, Oxford(UK), OUP, 107-114.
Delmonte R. D.Bianchi(1998), Dialogues From Texts: How to Generate Answers from a Discourse Model, Atti Convegno Nazionale AI*IA, Padova.
Delmonte R., E.Pianta(1996), IMMORTALE - Analizzatore Morfologico, Tagger e Lemmatizzatore per l'Italiano, in Atti V Convegno AI*IA "Cibernetica e Machine Learning", Napoli, 19-22
Delmonte R., E.Pianta(1998), Immortal: How to Detect Misspelled from Unknown Words, in BULAG, PCUF, Besançon, 1998, 193-218.
Delmonte R., E.Pianta(1999), Tag Disambiguation in Italian, in Proc. Treebank Workshop ATALA, Paris, pp.43-49.
Delmonte R., D.Bianchi, E.Pianta(1992), GETA_RUN - A General Text Analyzer with Reference Understanding, in Proc. 3rd Conference on Applied NLP, Systems Demonstrations, Trento, ACL, 9-10
Delmonte R., R.Dolci(1989), Parsing Italian with a Context-Free Recognizer, Annali di Ca' Foscari XXVIII, 1-2,123-161.
Delmonte R. R.Dolci(1997), Sound Parsing and Linguistic Strategies, Atti Apprendimento Automatico e Linguaggio Naturale, Torino, pp.1-4.
McDonald D. (1994), Reversible NLP by Linking the Grammar to the Knowledge Base, in Strzalkowski T., op.cit., 257-292.
Monachesi P.(1995), A Grammar of Italian clitics, ITK Dissertation series 1995-3.
Monachesi P.(1999), A Lexical analysis of Italian clitics, Proceeding of Vextal - Venezia,57-65.
Strzalkowski T. (ed.) (1994), Reversible Grammar in Natural Language Processing, Kluwer, Boston.
Zajac R. (1994), A Uniform Architecture for Parsing, Generation and Transfer, in Strzalkowski T., op.cit., 97-112.