,

Italian Pronouns in a CALL Setting

Jessica Zipf
University of Konstanz

Proceedings of the LFG'22 Conference

Miriam Butt, Jamie Y. Findlay and Ida Toivonen (Editors)

2022

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper introduces an Italian LFG grammar in the context of *Computer-Assisted Language Learning* (CALL). Specifically, the grammar focuses on Italian pronouns and is capable of detecting and correcting errors in pronoun structures. In addition, we propose a new analysis into the ongoing discussion surrounding Italian pronouns and their grammatical status (e.g. Monachesi 1996, 1999, Russi 2008, Schwarze 2001, Alsina Forthcoming). We develop a straightforward analysis that makes use of LFG's projection architecture and gives a clear answer to the ongoing question of their status as affixes or independent syntactic units. We illustrate the differences between preverbal and postverbal pronouns and classify the former as true clitics and the latter as inflectional affixes and account for their properties in terms of c- and f-structure representations.

1 Introduction

In this paper we present the implementation of an Italian grammar that is situated within the field of *Computer-Assisted Language Learning* (CALL), in particular a grammar that is capable of detecting and correcting errors in structures containing Italian pronouns. As studies have shown, Italian pronouns pose a significant obstacle for foreign language learners, making them an essential part of any learning environment for Italian (e.g., Leonini and Belletti 2003). In order to base our implementation on a sound theoretical analysis, we introduce a new analysis of Italian pronouns, providing a clearer picture of their properties and behaviour. This topic has been widely discussed in literature, with a range of different approaches and analyses. An important question is their grammatical status: are Italian pronouns clitics or affixes? Previous work has drawn different conclusions to that question: For example, Monachesi (1996) argues that Italian pronouns, with the exception of strong pronouns, exhibit properties of inflectional affixes and proposes a template approach within *Head-driven Phrase Structure Grammar* (HPSG). However, Russi (2008) regards some of the arguments put forth by Monachesi as rather controversial. Beside the affixal approach to Italian pronouns, there are also proponents of the clitic analysis, seeing these elements as independent syntactic constituents (e.g. Schwarze 2001). In this work, we propose an analysis that draws a clearer distinction between the sets of pronouns present in the Italian language: In addition to the undisputed strong pronouns, we argue for two other types of Italian pronouns: preverbal clitic pronouns and postverbal affixed pronouns. In section 2, we outline LFG's beneficial architecture for a CALL environment and present existing CALL tools using LFG. Section 3 is concerned with Italian pronouns: first, we give an overview of the pronoun inventory and their behaviour, then we move to pronouns in foreign language acquisition, and finally we look at existing theoretical analyses. In section 4 we present our own analysis, arguing for three classes of Italian pronouns: strong pronouns, clitic pronouns and affixed pronouns. Section 5 illustrates our analysis within LFG's modular architecture and the computational implemen-

tation. Finally, in section 6 we present the adaption of said implementation as a CALL grammar, capable of detecting and correcting erroneous input.

2 LFG and CALL

CALL is a field of research that emerged in the 1950s and its general development has always been influenced by various fields: computer technology, language pedagogy and second language acquisition (SLA) (Davies et al. 2013). With LFG being a grammar formalism rooted in mathematics, it has strong computational power and LFG grammars can be implemented. This puts LFG grammars within the research field of *Natural Language Processing* (NLP), one of the branches most relevant to CALL. NLP can support language learning by analysing learner language, for example, in tutoring systems (Heift 2017). This is the purpose of our project: implement an LFG-grammar to analyse learner language with respect to Italian pronouns.

A crucial component of any CALL tool is its capability to not only detect ungrammatical input but to also provide the learner with the grammatical alternative and informative feedback, allowing them to learn from their mistake and understand the underlying grammatical structures (Levy 1997). LFG and its computational realization with XLE (Crouch et al. 2008) provide components that are highly beneficial for these purposes in any CALL environment.

In his dissertation, Khader (2003) evaluated the English ParGram grammar as a possible grammar checker and its suitability for CALL applications.[1] For this purpose, the exiting English grammar was adapted to deal with a Chinese learner corpus of English, introducing the special feature UNGRAMMATICAL at f-structure and incorporating OT marks. OT marks allow the statement of preferences and dis-preferences in sentence analysis and can be ordered according to their relative importance. As a result, OT marks allow the grammar to deal with ambiguous, and more importantly, ungrammatical input. In combination with the generation component of XLE, the adapted English grammar is capable of recognizing and correcting erroneous user input. Additionally, the f-structure feature UNGRAMMATICAL provides information on the error type, based on the feature's value. Looking at a specific example, common errors are coded in rules, for example, in the case of a subject-verb mismatch with the 3rd person singular in English:

(1)

$$\begin{aligned} \textsc{Verb3SG} = \{&(\uparrow \textsc{subj pers}) = 3 \\ &(\uparrow \textsc{subj num}) = \textsc{sg} \\ &|@(\textsc{otmark BadVAgr}) \\ &(\uparrow \textsc{ungrammatical}) = \textsc{sv-agr}\}. \end{aligned}$$

[1] The *Parallel Grammar Project* is an international collaboration on LFG-based grammar development.

The OT-Mark ensures that the ungrammatical analysis is only applied when a grammatical one is unavailable, while the f-structure feature UNGRAMMATICAL returns feedback on the kind of mistake (here indicated by the feature's value *sv-agr*).

A similar approach is taken by Fortmann and Forst (2004) who modified a broad-coverage LFG grammar for German to use it as a grammar checker in L2 learning. A focus was set on the grammar's capability to provide crucial information necessary for generating feedback as to what was ungrammatical in the learners' input. The CALL grammar was adapted to deal with typical errors of German L2 learners. These include word order, subject-verb agreement and the declension of determiners, adjectives and nouns within NPs.

To parse ungrammatical sentences, the grammar was extended with malrules, that is, rules that are specifically designed to map ungrammatical input. In addition to malrules, Fortmann and Forst (2004) also makes use of special f-structure features and OT-dispreference marks.

Another example of such an approach is the *Arboretum* system (Bender et al. 2004), a grammar checking tool for English. *Arboretum* is based on the English Resource Grammar (ERG) (Flickinger 2000) and the LKP parser and generator which includes a Lisp parser that interprets typed-feature structure grammars like the ERG (Copestake 2002).[2] While the previous two systems used LFG-based grammars, the ERG is written within the syntactic framework of HPSG (Pollard and Sag 1994). Relating semantic representations to surface strings, *Arboretum* is capable of parsing and generation. The aim is to produce well-formed semantic representations given ill-formed or erroneous input. Again, this is done by augmenting the ERG with malrules. The parser uses the malrules to parse erroneous input. Then, the malrules are turned off for generation. This is an important step towards the aim of grammar checking: turning off malrules for generation leaves the grammar with only grammatical rules to generate well-formed semantic representations. In a final processing step, the tool produces corrected sentences and a diagnosis of the nature of the error, based on the malrules used to parse the original ill-formed input.

As the work by Khader (2003) and Fortmann and Forst (2004) show, LFG offers components beneficial for CALL applications. While previous work was mostly done on English and German, this project develops a CALL application for Italian, focusing on Italian pronouns.

3 Italian Pronouns

In Italian, clitic pronouns can correspond to Grammatical Functions (GF), in particular to the OBJ (2), OBJ_θ (3) or an OBL (4) (Alsina Forthcoming).

[2]The LKP is a grammar and lexicon development environment that allows users to experiment with various grammars and learn the details of the formalism. Additionally, it supports the development of large-scale grammars.

(2) **Ti** ved-o.
2SG.ACC see-PRES.1SG
'I see you.'

(3) **Le** mand-o una email.
3SG.DAT.F send-PRES.1SG a email
'I send her an email.'

(4) **Ci** vad-o.
LOC go-PRES.1SG
'I go there.'

With respect to word order, the pronoun does not take the same position as its phrasal counterpart. In (2), the OBJ pronoun immediately precedes the verb, while in (5), an object NP follows the verb:

(5) Ved-o Luca.
see-PRES.1SG Luca
'I see Luca.'

The same holds for an OBJ_θ-pronoun and a phrasal expression, as indicated by examples (3) and (6):

(6) Mand-o una email a Giulia.
send-PRES.1SG a email to Giulia
'I send Giulia an e-mail.'

When both, OBJ and OBJ_θ, are expressed by a pronoun, the usual canonical word order of the whole sentence is reversed and the OBJ_θ precedes the direct pronoun ((6) & (7)):

(7) **Me** **l'**ha dato Carlo.
1SG.DAT 3SG.ACC.AUX.3SG give.PST.PTCP Carlo
'Carlo gave it to me.'

Another property is that the pronoun directly attaches to its verbal host if the verbal host is a non-finite or imperative verb:

(8) Bisogna far**lo**.
need.PRES.3SG do.INF.3SG.ACC.M
'It needs to be done.'

(9) Fam**mi** un piacere!
do.IMP.2SG.1SG.DAT a favour
'Do me a favour!'

This direct attachment becomes clearer when an auxiliary is present: In (10), the pronoun remains attached to the infinitive, whereas in a sentence without an infinite, the auxiliary is placed between the pronoun and verb (7).

(10) Hai dovuto far**lo**.
AUX.2SG must.PST.PTCP do.INF.3SG.ACC.M
'You have had to do it.'

In literature, the pronouns presented so far are often referred to as "clitics" (Monachesi 1996, 1999, Schwarze 2001, Alsina Forthcoming). However, this paper argues for a more differentiated analysis, outlined in section 4. Until then, the term "clitic" is used purely descriptively. The following two tables show the inventory of accusative and dative pronouns[3]:

NUM	Form	PERS	GEND
Singular	mi	1	–
	ti	2	–
	lo	3	m
	la	3	f
Plural	ci	1	–
	vi	2	–
	li	3	m
	le	3	f

Table 1: Accusative pronouns (OBJ)

NUM	Form	PERS	GEND
Singular	mi (me)	1	–
	ti (te)	2	–
	gli	3	m
	le	3	f
Plural	ci (ce)	1	–
	vi (ve)	2	–
	gli	3	–

Table 2: Dative pronouns (OBJ_θ)

The pronoun inventory also includes tonic/strong pronouns, given in Table 3 (Russi 2008, p. 48):

NUM	Form	PERS	GEND
Singular	me	1	–
	te	2	–
	lui	3	m
	lei	3	f
Plural	noi	1	–
	voi	2	–
	loro	3	–

Table 3: Personal tonic pronouns

In contrast to the "clitic" pronouns, tonic pronouns are the SUBJ, exhibit the same syntactic distribution as full NPs, may carry stress and may appear by themselves:

[3]The forms in parenthesis are allomorphs that occur in clitic sequences, as illustrated in (7).

(11) **Lei** mangi-a una mela.
3SG.NOM.F eat-PRES.3SG an apple
'She eats an apple.'

(12) Chi sta-nno cercando? - **Lui/Mario**.
Who stay-PRES.3PL search.PRS.PROG - 3SG.DO/Mario
'Who are they searching for?' - 'Him/Mario'

(13) Chi sta-nno cercando? - ***Lo**.
Who stay-PRES.3PL search.PRS.PROG - 3SG.ACC.M
'Who are they searching for?' - *'Him.'

3.1 Italian Pronouns in Foreign Language Acquisition

Studies have shown that Italian "clitic" pronouns constitute a rather difficult phenomenon to be acquired for L2 learners of Italian. In particular, it is a rather slow acquisition process and learners exhibit a tendency to use avoidance strategies: In 2003, Leonini and Belletti investigated the issue with adult L2 learners with different L1s, in the context of subject inversion structures, that is structures in which the subject follows the main verb. With transitive verbs,VS structures typically realize the OBJ as a "clitic" pronoun:

(14) Chi ha portato i fiori?
Who have.AUX.3PSG bring.PST.PTCP the flowers
'Who brought the flowers?'

(15) **Li** ha portati Silvia.
3PL.ACC.M have.AUX.3PSG bring.PST.PTCP Silvia
'Silvia brought them.'

The results of an elicitation task show that the use of "clitics" in such structures is not properly acquired. Instead, they are often omitted or, even more often, realized as a full NP — a typical avoidance strategy (Leonini and Belletti 2003). The study included a control group consisting of ten adult native speakers of Italian: In contrast to the learners, they exhibited a total lack of clitic omissions and use of strong pronouns. Another study (Santoro 2007) focused on L2 acquisition of both accusative and dative "clitics" with English L1 speakers. The results show a slow and difficult acquisition process. Despite being a slow process, the study also indicates that it is a steady learning trajectory, with visible improvement as learners become more proficient. The study by Schuiringa (2014) confirms that the acquisition of "clitics" is a slow yet steady process: While learners at beginner level make use of avoidance strategies, both the variety of these strategies and their frequency of use decrease over time. Learners at beginner level (Level A) make use of the following avoidance strategies, ordered in their rate of occurrence: NP (65%), incomplete answer (17.5%), tonic pronoun (7.5%), "clitic" pronoun (7.5%)

and "clitic" omission (2.5%). In contrast, intermediate learners (Level B) make use of three strategies: NP (62.5%), "clitic" (32.5%), "clitic" omission (5%). Finally, advanced learners (Level C) either use the expected "clitic" (77.5%) or a NP (22.5%). The control group of native Italian speakers produced "clitics" in 85%, used a NP in 5% and gave an incomplete answer in 10%. They never omitted a clitic and never used a strong pronoun.

It becomes clear that "clitics" pose a significant obstacle for foreign language learners, making them an essential part of any learning environment for Italian. Before we present the implementation into our LFG CALL tool, their linguistic analysis must be addressed in order to guide the computational implementation within the LFG framework.

3.2 Italian Pronouns in the Literature

A key question is their morphological status: are these pronouns independent units or affixes (e.g. Zwicky 1977, Klavans 1985, Russi 2008)? A simple definition of clitics can be formulated as follows: clitics are linguistic units, which combine with other words or phrases to create phrases, but are phonologically dependent on an adjacent word, usually called host, and cannot be stressed. According to Monachesi (1996), they are usually associated with certain properties like a specific position within a sentence, the need for a host to attach to, and the occurrence in a position other than that of its corresponding full phrase. Additionally, nothing can intervene between the clitic and its host. Important for the distinction between clitics and affixes is a set of criteria proposed by Zwicky and Pullum (1983):

A. Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.

B. Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups

C. Morphophonological idiosyncrasies are more characteristic of affixed words than of clitic groups

D. Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups

E. Syntactic rules can affect affixed words, but cannot affect clitic groups

F. Clitics can attach to material already containing clitics, but affixes cannot

While these criteria are strong indicators, they are not clear-cut and not all of these must necessarily apply at once (Bögel 2015). From an LFG perspective, the distinction between clitic and affix is visible at c-structure: if clitics are independent syntactic units, they will receive their own c-structure node, whereas affixes will

not. Previous work has drawn different conclusions from this set of tests: Monachesi (1996) argues for an affix analysis, stating that Italian pronouns exhibit properties of inflectional affixes - at least when one pronoun is present. In sequences, that is, when two or more are attached to the verb, she sees it as compounding, linking it to Prosodic Minimality (McCarthy and Prince 1990).[4] With respect to the criteria above, claimed morphological idiosyncrasies for Italian are (a) the vowel change $i \rightarrow e$ that affects the "clitic" ending when followed by another one beginning with *l-* or *-n* ((16) and (17)) and (b) the use of the masculine *gli* instead of *le* in pronoun sequences (18):

(16) **Te/*Ti lo** dir-à.
2SG.DAT 3SG.ACC.M say-3SG.FUT
'She will tell it to you.'

(17) **Me/*Mi ne** parler-à.
1SG.DAT PART talk-3SG.FUT
'She will talk to me about it.'

(18) **Glielo/*Le lo** dir-à.
3SG.DAT.M.3SG.ACC.M tell-3SG.FUT
'S/he will tell it to her.'

According to Monachesi, the "most crucial evidence for the affixal nature" of Italian pronouns comes from their rigid ordering (Monachesi 1996, p. 38). As outlined in the previous section, the dative pronoun must precede the accusative pronoun:

(19) Martina **me lo** spedisce.
Martina 1SG.DAT 3SG.ACC.M send.PRES.3.SG
'Marina sends it to me.'

(20) *Martina **lo mi** spedisce.
*Martina 3SG.ACC.M 1SG.DAT send.PRES.3.SG
'Marina sends it to me.'

Within HPSG, Monachesi (1996) considers Italian pronoun "clitics" as a special kind of affixes that attach to verbs.[5] Taking a lexical approach, she considers these as featural information that is provided in the lexicon and used in morphology and phonology for the realization of the cliticized verb. As such, "clitics" are a spell out of certain features. Lexical Rules link these features to a specific affix surface form (Monachesi 1999, p.73):

[4]According to McCarthy and Prince (1990), in some languages a clitic cannot constitute a prosodic word as its monosyllabic, thus it adjoins the host to form a prosodic word with it. If two or more clitics are present, they merge together into a bi-/trisyllabic unit that can constitute a prosodic word that combines with verb as in compounds.

[5]This is in line with existing work on Romance "clitics", e.g., Bonet (1995) and Alsina (2020) for Catalan, Miller and Sag (1997) for French, and Luís and Sadler (2003) and Luís and Spencer (2005) for Portuguese.

(21) [STEM... CLTS <dat, pers3, num, masc>] → [AFFIX [PHON <gli>]]

To account for the rigid ordering of Italian "clitics", Monachesi (1996) adopts a template approach, assigning them to different position classes:

	I	II	III	IV	V	VI
Position	mi ti gli le ci vi	ci	si (ref)	lo la li le (acc)	si (imp)	ne

Table 4: Template Account for Clitic Ordering (Monachesi 1996, p.122)

In contrast, Schwarze (2001) regards the Italian pronouns as independent syntactic constituents. Following Grimshaw (1982), Schwarze (2001) gives a c-structure analysis of Italian and French "clitics". For Italian he proposes the following c-structure rule:

(22) V'→ (CL1) (CL2) V

In this treatment, CL includes "clitic" clusters as well as single "clitics". CL1 is the negation particle *non*, which needs a verbal host, and CL2 comprises "clitics" that correspond to grammatical functions (objects, obliques and adjuncts). However, this rule only applies to sentences with a preverbal "clitic" and does not include the analysis of examples like (23):

(23) Bisogn-a far**lo**.
need-PRES.3SG do.INF.3SG.ACC.M
'It needs to be done.'

Here, Schwarze suggests that "their similarity with affixes can probably best be expressed, rather than at the c-structure level, by the representation of their functional properties" (Schwarze 2001, p.293). While this sounds in line with Monachesi, Schwarze does not propose an anlysis for these postverbal pronouns. In his analysis, Schwarze formulates lexical entries for "clitics", encoding their grammatical functions. Here is an example for the feminine object "clitic" *la*:

(24) la, CL
(↑ OBJ PRED) = 'PRO'
(↑ OBJ PERS) = 3
(↑ OBJ NUM) = SG
(↑ OBJ GEN) = FEM

To conclude, there are varying analyses of Italian pronouns. Monachesi (1996, 1999) analyses Italian pronouns as inflectional affixes. Similar approaches are proposed by Bonet (1995) for Italian, Spanish and Catalan, Luís and Spencer (2005) for European Portuguese and Alsina (Forthcoming). Schwarze (2001) in contrast, argues for an analysis as independent syntactic units, which is in line with Quaglia (2012) for Italian and Estigarribia (2005, 2013) for Spanish. As we are using LFG to develop our CALL tool, we now present an LFG analysis of Italian pronouns.

4 Proposed Analysis of Italian Pronouns

This paper presents a straightforward analysis within LFG's projection architecture, showing a clearer picture of these elements than the existing analyses. In particular, we assume three types of pronouns in Italian:

1. **Strong pronouns**
2. **Clitic pronouns**
3. **Affixal pronouns**

While this tripartition among pronouns is well known, we propose a different allocation of elements to these three classes, particularly to type 2, clitic pronouns and type 3, affixal pronouns. The first class refers to previously introduced personal strong pronouns given in Table 3. Their main properties include that they exhibit the same syntactic distribution as full NPs, they may carry stress and appear by themselves. The following presentation of our analysis focuses on type 2, clitic pronouns and type 3, affixal pronouns.

Disregarding the strong pronouns, the remaining pronouns exhibit the following syntactic distribution:

(25) **Lo** mangi-o.
3SG.ACC.M eat-PRES.1.SG
'I eat it.'

(26) Voglio mangiar-**lo**
want.PRES.1.SG eat.INF-3SG.ACC.M
'I want to eat it.'

In this paper, we argue that preverbal pronouns are true clitics (25), while postverbal pronouns are in fact inflectional affixes (26).[6] Note: when we speak of preverbal pronouns in the remaining analysis, we always refer to our type 2, clitic pronouns and not type 1, strong pronouns. Our analysis is based on their different behaviour, mainly rooted in syntax, morphology and semantics and draws on criteria by Zwicky and Pullum (1983). A similar, yet slightly different analysis for

[6]Following our analysis, we gloss all postverbal pronouns as affixes from this point on.

Italian and other Romance languages is proposed by Ordóñez and Repetti (2005), who argue for true clitics preverbally and either clitic or weak pronoun postverbally.

4.1 Syntactic Behaviour

Main evidence for the different syntactic nature of preverbal and postverbal pronouns comes from clitic climbing (27) and wide scope over coordination. Preverbal pronouns can have wide scope over coordination when corradical verbs are conjoined, as Kayne (1975) has noted for modern-day Romance languages, and when the conjoined verbal phrases share the same argument and eventive structure (Egerland and Cardinaletti 2010). In these contexts, the second pronoun may be dropped without leading to ungrammaticality ((28) & (29)).

(27) **Lo** voglio poter vedere.
3SG.ACC.M want.PRES.1.SG can.INF see.INF
'I want to be able to see it.'

(28) **Lo** legg-o e **(lo)** rilegg-o.
3SG.ACC.M read-PRES.1.SG and (3SG.ACC.M) reread-PRES.1.SG.
'I read and reread it.'

(29) **Ti** dico e **(ti)** promett-o che ...
2SG.ACC say.PRES.1.SG and (2SG.ACC) promise-PRES.1.SG that ...
'I say and promise you that..."

However, dropping the second pronoun leads to ungrammaticality with postverbal pronouns in the same contexts ((30), (31) & (32)), showing that these cannot have wide scope:

(30) *Voglio legger-**lo** e rileggere.
Want.PRES.1.SG read.INF-3SG.ACC.M and reread.INF
'I want to read and reread it.'

(31) *Voglio leggere e rilegger-**lo**
Want.PRES.1.SG read.INF and reread.INF-3SG.ACC.M
'I want to read and reread it.'

(32) Voglio legger-**lo** e rilegger-**lo**
Want.PRES.1.SG read.INF-3SG.ACC.M and reread.INF-3SG.ACC.M
'I want to read and reread it.'

With respect to this behaviour, postverbal pronouns behave like bound affixes while preverbal pronouns do not.

4.2 Morphological Idiosyncrasies

Looking at morphological idiosyncrasies, some arguments for preverbal pronouns behaving like affixes have already been discussed in examples (16), (17) and (18): The dative pronoun goes through a vowel change from $i \rightarrow e$ that effects the pronouns ending when its followed by another pronoun beginning with *l-* or *n-*. Yet, the idiosyncrasy appears to be debatable as the change in form can be explained in terms of grammaticalizaion (Russi 2008). Additionally, it can be argued that the form change avoids ambiguities in interpreting the pronouns as dative and accusative pronouns: without the change from $i \rightarrow e$, the dative pronoun is form identical with the respective accusative pronoun, leaving only word order to disambiguate. Thus, the claimed idiosyncrasy for preverbal pronouns is debatable. In contrast, morphological idiosyncrasies can be observed with postverbal pronouns:

(33) Fa-**mmi** un piacere!
do.IMP.2.SG-1.SG.DAT a favour
'Do me a favour!'

With imperatives, the consonant of the postverbal pronoun is doubled when attached to a one syllabic root (compare (33) and (34)):

(34) Porta-**mi** il libro!
bring.IMP.2.SG-1.SG.DAT the book
'Bring me the book!'

As Zwicky and Pullum (1983) point out, morphological idiosyncrasies are typical for inflectional affixes, which are capable of altering the shape of their host. Yet, this is untypical for clitics. In addition to the syntactic behaviour, the morphological behaviour of postverbal pronouns speaks for their affixal nature.

4.3 Semantic Idiosyncrasies

Monachesi (1999) and Russi (2008) agree that there is no difference in meaning between the sequence of -in their terms - "clitic" and verb, and the verb and strong pronoun alternative:

(35) Vi ved-o.
2PL.ACC see-PRES.1.SG
'I see you.'

(36) Ved-o voi.
see-PRES.1.SG you.PL
'I see you.'

Yet, with postverbal pronouns, there are examples of semantic idiosyncrasies, where a change in meaning happens once a pronoun is attached: to exemplify, the infinitive *entrare* means 'to enter', yet the verb pronoun sequence *entrarci* translates to 'to be involved in'. Russi (2008) analyses examples like this as "lexicalized clitics" that show idiosyncratic semantics typical of affixes. While she keeps the term "clitic", this is yet another argument for our referral to postverbal pronouns as affixes, dropping the term clitic.

4.4 Prosodic properties

While at this point we have established how preverbal and postverbal pronouns differ, showing that preverbal pronouns exhibit behaviour typical of true clitics and postverbal pronouns behave like inflectional affixes, we want to address one remaining question: Why are preverbal pronouns not simply full prosodic words? Prosodically, clitics are seen as deficient (Anderson 2005, Bögel et al. 2008, Bögel 2015). This is because they do not form a prosodic word of their own, but instead are phrased together with another prosodic word as part of prosodic phrasing (Bögel et al. 2008). An important characteristic of the Italian preverbal pronouns is their need for a verbal host. While they appear as syntactic independent items, they require the immediate vicinity to their verbal host with no other element allowed to appear between the pronoun and its host (e.g., Cardinaletti and Starke 1999, Monachesi 1999, Schwarze 2001). Their dependence on a verbal host also becomes evident in their incapability of appearing without it, for example as a one word answer to a question:

(37) A chi ha regalato il libro? - ***Mi** / A me.
To whom have.AUX.3.SG give.PST.PTCT the book? - *1SG.DAT / to me
To whom has he/she given the book? - To me.

Further prosodic evidence for the clitic status of the preverbal pronouns comes from stress assignment. Ordóñez and Repetti (2005) show that Italian postverbal pronouns may affect the stress pattern of the verb in some varieties of Italian, however this is not observed for preverbal pronouns:

	Verb	**DAT pronoun**	**ACC pronoun**	**Both pronouns**
Penultimate stress shift (dialect in Naples)	nárra	nárra-mi	nárra-la	narra-mí-la

Table 5: Stress assignment patterns with the verb 'narrare' (Ordóñez and Repetti 2005, p. 170)

In contrast to the postverbal pronouns, preverbal pronouns neither carry stress nor affect the stress pattern of their host - a property typical of clitics (Monachesi 1996, Halpern 1998). To summarize, we have provided morphological, syntactic,

semantic and prosodic arguments for our analysis of preverbal pronouns as true clitics and postverbal pronouns as affixes. In contrast to previous analyses, our work provides a clearer picture of these elements and their behaviour. In our work, the Italian language has three classes of pronouns: (a) strong pronouns, (b) clitic pronouns and (c) affixal pronouns.

5 An LFG Analysis

Having laid out our theory for Italian pronouns, we now embed our work into an analysis within LFG's projection architecture. Given the context of a CALL implementation, we not only present a theoretical proposal but also a computational implementation, focusing on the morphology and syntax modules.[7]

5.1 Morphological Module

We implemented a finite-state morphological analyzer (FSM) (Beesley and Karttunen 2003) as the morphological module of the Italian CALL grammar. The FSM takes a surface string as input and returns a morphological analysis:

(38) mangia → mangiare +VERB +PRES +3SG

Considering our analysis of postverbal pronouns as affixes, these are part of the FSM:

(39) mangiar**lo** → mangiare +VERB +INF +**3SG.ACC.M**

This morphological analysis is passed on to the LFG grammar and associated with f-structural information at the morphology-syntax interface. In particular, each tag coming from the FSM functions as a sublexical item that is associated with f-structure information. For example, the tag +PRES is annotated as the functional information $(\uparrow \text{TENSE} = \text{PRES})$.

Yet, the preverbal clitic pronouns are not dealt with by the FSM. Instead, the FSM simply recognizes these as independent elements and not as morphological affixes. As clitics, we encode these elements as hand-written lexical entries in the grammar:

(40)

$$\begin{array}{llll} la & \text{CL} & * & (\uparrow \text{PRED}) = \text{PRO} \\ & & & (\uparrow \text{PERS}) = 3 \\ & & & (\uparrow \text{CASE}) = \text{ACC} \\ & & & (\uparrow \text{NUM}) = \text{SG} \\ & & & (\uparrow \text{GEND}) = \text{FEM} \end{array}$$

[7] This is mainly due to the context of our CALL tool, which focuses on supporting the acquisition of morphological and syntactic properties of Italian pronouns

5.2 C- and F-Structure

The implications following from our analysis are mostly visible at c-structure: while the preverbal clitic pronouns occupy a terminal node at c-structure, postverbal pronouns are affixes, and thus simply part of the morphological make-up of the verb. With respect to the phrase-structure analysis of preverbal clitics, they are embedded under VP, with CL1 referring to the dative clitic and CL2 to the accusative clitic:[8]

(41) VP→ (CL1) (CL2) (AUX) V

It is important to note that while both types of pronouns occupy different positions, they are restricted from co-occurring because they each contribute PRED values, and therefore multiple pronouns pertaining to the same grammatical function would entail a violation of uniqueness (Dalrymple et al. 2019).

At f-structure, preverbal clitic pronouns and postverbal affixed pronouns are identical. They both express grammatical functions that the main verb subcategorizes for and both represent the same functional information. In the case of preverbal clitic pronouns, the functional information of the GF comes from the lexical entry (40). Depending on the case of the pronoun it functions either as the OBJ or OBJ_θ. With postverbal affix pronouns the functional information is provided by the sublexical entry for the respective tag coming from the FSM. For example, the pronoun affix 'lo' is associated with the tag +3SG.ACC.M in the FSM. The grammar associates this tag with f-structure information in a sublexical entry (42):

(42)

$$\begin{array}{lll} +3\text{SG.ACC.M} & \text{PRON} & (\uparrow \text{OBJ PRED}) = \text{PRO} \\ & & (\uparrow \text{OBJ PERS}) = 3 \\ & & (\uparrow \text{OBJ CASE}) = \text{ACC} \\ & & (\uparrow \text{OBJ NUM}) = \text{SG} \\ & & (\uparrow \text{OBJ GEND}) = \text{MASC} \end{array}$$

As the functional annotations show, the respective f-structure for the grammatical functions are identical for preverbal clitics and postverbal affixes: When comparing the f-structure analyses of 'Lo mangio' (I eat it) in Figure 1 and 'Maria vuole mangiarlo' (Maria wants to eat it) in Figure 2, we see that the f-structure information associated with the pronouns, that is, the information on the OBJ, are identical.

[8]The rule does not follow x-bar theory.

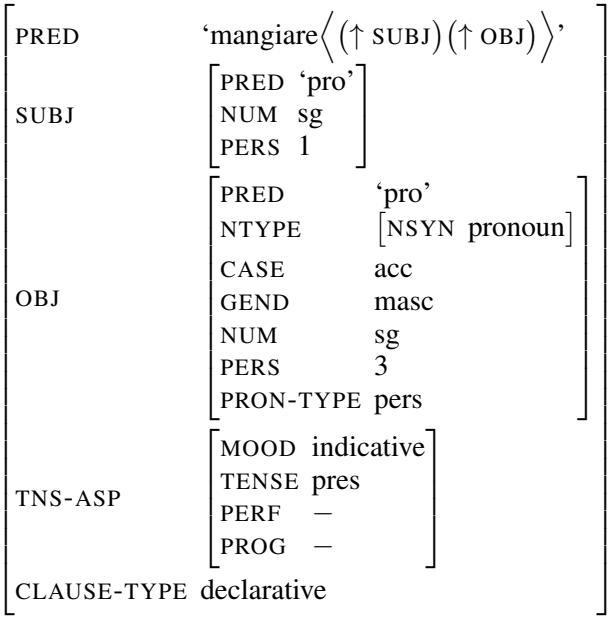

Figure 1: F-Structure for ‘Lo mangio.’ (I eat it.)

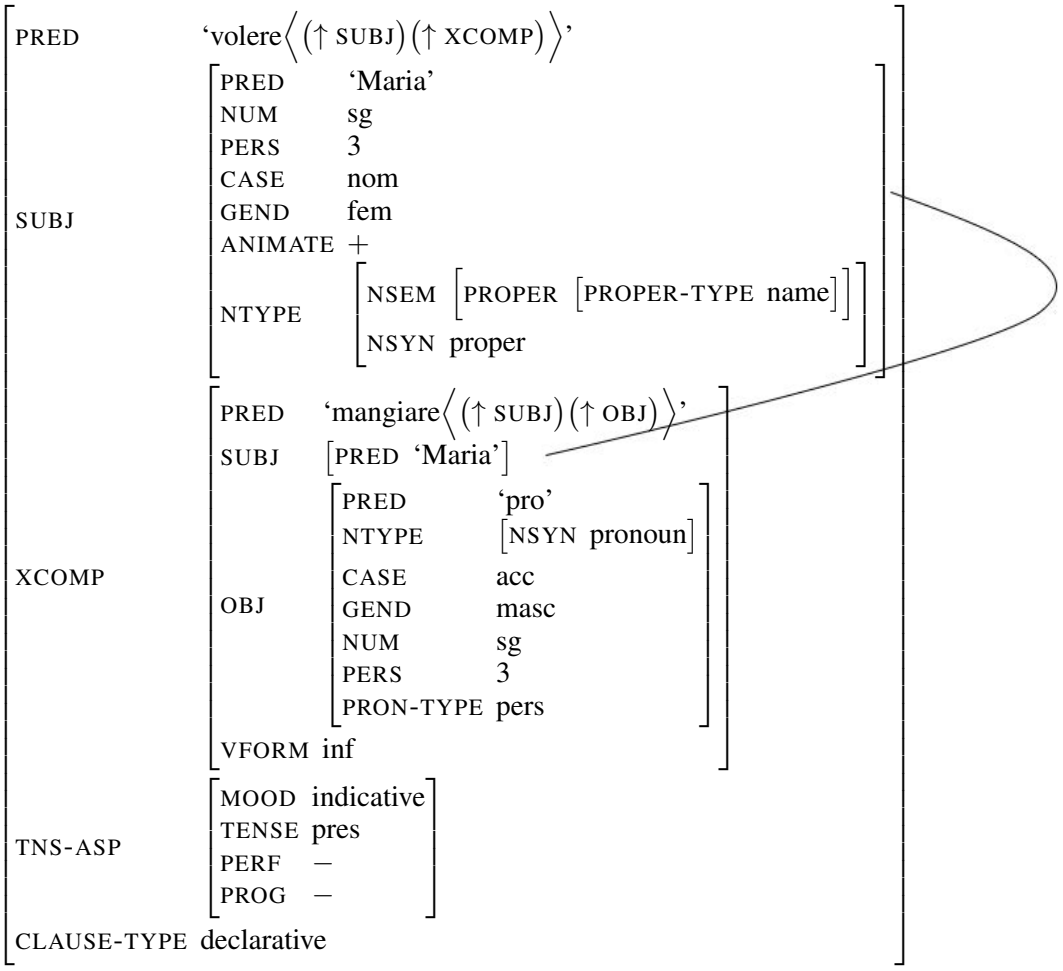

Figure 2: F-Structure for ‘Maria vuole mangiarlo.’ (Maria wants to eat it.)

6 A CALL grammar for Italian Pronouns

The previous section outlined our LFG analysis and implementation of Italian pronouns. We now turn to the process of adapting said grammar for a CALL environment, that is, enabling it to detect and correct errors with respect to pronouns in a learner's input. For that purpose, we follow Khader (2003) and Fortmann and Forst (2004) and introduce the special f-structure feature UNGRAMMATICAL and the OT-mark CLITIC-ERROR. Both are added to structures responsible for parsing erroneous input. Information on the error type is provided by the value associated with the f-feature UNGRAMMATICAL. For example, if the order of two pronouns is incorrect, UNGRAMMATICAL has the value WRONG-ORDER:

(43) Porta-**me**-**lo**!
bring.IMP.2.SG-**1.sg.dat**-**3.sg.acc.m**
'Bring it to me!'

(44) *Porta-**lo**-**me**!
bring.IMP.2.SG-**3.sg.acc.m**-**1.sg.dat**
'Bring it to me!'

The respective f-structure output for (44) is provided in Figure 3, illustrating the feature UNGRAMMATICAL with the value WRONG-ORDER:

```
[PRED           'portare<[1-SUBJ:pro], [1-OBJ2:me], [1-OBJ:lo]>'
 SUBJ           [PRED 'pro'
                 PERS 2    ]
                [PRED 'pro'
 OBJ            NTYPE [NSYN pronoun]
                CASE acc, GEND masc, NUM sg, PERS 3, PRON-TYPE pers]
                [PRED 'pro'
 OBJ2           NTYPE [NSYN pronoun]
                CASE dat, NUM sg, PERS 1, PRON-TYPE pers]
 UNGRAMMATICAL {wrong-order}
1CLAUSE-TYPE imperative, VFORM inf                                  ]
```

Figure 3: F-Structure for '*Portalome!' (Bring it to me!)

The main purpose of the special OT-mark is to allow parsing of an ungrammatical sentence, but disregarding the erroneous f-structure when generating from it. In short, the generation component of XLE uses an f-structure as input to generate the string represented by said f-structure. Yet, in a learning environment, the system needs to generate grammatical sentences. The OT-mark ensures that dispreferred — here ungrammatical — structures are disregarded in generation. That means that the system generates from the f-structure in Figure 3 while ignoring that part of the grammar that is associated with the OT-mark CLITIC-ERROR. As a result, the system returns the grammatical string in (43). In this manner we deal with typical learner errors related to Italian pronouns, for example, substitution with a

strong pronoun in an inappropriate context or incorrect placement of the pronoun within the sentence (Leonini and Belletti 2003, Santoro 2007).

7 Conclusion

In this paper, we presented an analysis of Italian pronouns that strictly differentiates between clitic and affixed pronouns - a distinction that was often blurred in previous analyses. By adopting the modular architecture of LFG, a straightforward analysis of Italian pronouns is possible, distinguishing between preverbal clitic and postverbal affixed pronouns. In our analysis, the preverbal clitic pronouns are treated as lexical items with their own lexical entry, occupying a terminal c-structure node. In contrast, the postverbal affixed pronouns are part of morphology and are attached as inflectional affixes to their verbal host. We then implement our analysis into a grammar for a CALL learning environment. The framework of LFG offers great benefits for a CALL tool, given the integration of OT-marks and XLE's generation component. In order to accommodate the grammar for a CALL purpose, we use a special f-structure feature and a dispreference OT-mark. These elements allow the parsing of an ungrammatical input, the detection of the error-type and the generation of the grammatical alternative.

References

Alsina, Alex. 2020. Obligatory clitic expression, clitic omission, and the morphology-syntax interface pages 5–25.

Alsina, Alex. Forthcoming. LFG and Romance Languages. In Mary Dalrymple (ed.), *Forthcoming: The Handbook of Lexical Functional Grammar*, pages I–1xiv, Berlin: Language Science Press.

Anderson, Stephen. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.

Beesley, Kenneth and Karttunen, Lauri. 2003. *Finite-State Morphology*. Stanford: CSLI Publications.

Bender, Emily M, Flickinger, Dan, Oepen, Stephan, Walsh, Annemarie and Baldwin, Timothy. 2004. Arboretum: Using a precision grammar for grammar checking in CALL. *Proceedings of the InSTILICALL2004 Symposium on Computer Assisted Language Learning* 17, 19.

Bögel, Tina. 2015. *The Syntax – Prosody Interface in Lexical Functional Grammar*. Ph. D.thesis, University of Konstanz.

Bögel, Tina, Butt, Miriam and Sulger, Sebastian. 2008. Urdu Ezafe And The Morphology-Syntax Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG08 Conference*, pages 129–149, Stanford: CSLI Publications.

Bonet, Eulàlia. 1995. Feature structure of Romance clitics. *Natural Language and Linguistic Theory* 13, 607–647.

Cardinaletti, Anna and Starke, Michal. 1999. The typology of structural deficiency. In Henk van Riemsdijk (ed.), *Clitics in the Languages of Europe*, pages 145–233, De Gruyter Mouton.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.

Crouch, D., Dalrymple, M., Kaplan, R. M., King, T. H., Maxwell III, J.T. and Newman, P. 2008. *XLE Documentation.*. Palo Alto Research Center.

Dalrymple, Mary, Lowe, John J and Mycock, Louise. 2019. *The Oxford Reference Guide to Lexical Functional Grammar.*. Oxford: Oxford University Press.

Davies, Graham, Otto, Sue E. K. and Rüschoff, Bernd. 2013. Historical perspectives on CALL. In Michael Thomas, Hayo Reinders and Mark Warschauer (eds.), *Contemporary Computer-Assisted Language Learning*, pages 19–38, Bloomsbury Publishing.

Egerland, Verner and Cardinaletti, Anna. 2010. I pronomi personali e riflessivi. *Grammatica dell'italiano antico* 1, 401–467.

Estigarribia, Bruno. 2005. Direct object clitic doubling in OT-LFG: A new look at Rioplatense Spanish. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG '05 conference*, pages 116–135, Stanford: CSLI Publications.

Estigarribia, Bruno. 2013. Rioplatense Spanish clitic doubling and "tripling" in Lexical-Functional Grammar. In Chad Howe, Sarah E. Blackwell and Margaret Lubbers Quesada (eds.), *Selected proceedings of the 15th Hispanic Linguistics Symposium*, pages 297–309, Somerville: Cascadilla Proceedings Project.

Flickinger, Dan. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering* 6(1), 15–28.

Fortmann, Christian and Forst, Martin. 2004. An LFG Grammar Checker for CALL. In *Proceedings of InSTIL/ICALL2004 - NLP and Speech Technologies in Advanced Language Learning Systems*, pages 2–5.

Grimshaw, Jane. 1982. On the lexical representation of Romance reflexive clitics. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, pages 87–148, Cambridge: The MIT Press.

Halpern, Aaron L. 1998. "Clitics". In Andrew Spencer and Arnold M. Zwicky (eds.), *Handbook of Morphology*, pages 101–122, Oxford: Blackwell.

Heift, Trude. 2017. *History and Key Developments in Intelligent Computer-Assisted Language Learning (ICALL)*, pages 289–300. Cham: Springer International Publishing.

Kayne, Richard S. 1975. *French Syntax: The Transformational Cycle*. Cambridge, MA and London: MIT Press.

Khader, Rafiq. 2003. *Evaluation of an English LFG-based Grammar as Error Checker.*. Msc thesis, UMIST.

Klavans, Judith L. 1985. The Independence of Syntax and Phonology in Cliticization. *Language* 61(1), 95.

Leonini, Chiara and Belletti, Adriana. 2003. Adult L2 Acquisition of Italian Clitic

Pronouns and 'Subject Inversion'/ VS Structures. *Proceedings of GALA 2003* .
Levy, Michael. 1997. *Computer-Assisted Language Learning. Context and Conceptualization*. Oxford: Oxford University Press.
Luís, Ana and Sadler, Louisa. 2003. Object clitics and marked morphology. In Claire Beyssade, Oliver Bonami, Patricia Cabredo Hofherr and Francis Corblin (eds.), *Empirical issues in syntax and semantics*, pages 133–153, Paris: Presses Universitaires de Paris-Sorbonne, fourth edition.
Luís, Ana and Spencer, Andrew. 2005. A paradigm function account of 'mesoclisis' in European Portuguese. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2004*, pages 177–228, Dordrecht: Kluwer.
McCarthy, John J. and Prince, Alan. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language & Linguistic Theory* 8(2), 209–283.
Miller, Philip H. and Sag, Ivan A. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15, 573–639.
Monachesi, Paola. 1996. *A grammar of Italian clitics*. Tilburg: ITK Dissertation Series.
Monachesi, Paola. 1999. *A Lexical Approach to Italian Cliticization*. Stanford: CSLI Publications.
Ordóñez, Francisco and Repetti, Lori. 2005. Stressed Enclitics? In Jean-Pierre Y. Montreuil (ed.), *New Perspectives on Romance Linguistics. Vol. II: Phonetics, Phonology and Dialectology*, pages 167–181, Amsterdam: John Benjamins Publishing Company.
Pollard, Carl and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
Quaglia, Stefano. 2012. On the Syntax of Some Apparent Spatial Particles in Italian. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG12 Conference*, pages 503–523, Stanford: CSLI Publications.
Russi, Cinzia. 2008. *Italian clitics. An Empirical Study*. Berlin: Mouton de Gruyter.
Santoro, Maurizio. 2007. Second Language Acquisition of Italian Accusative and Dative Clitics. *Second Language Research* 23(1), 37–50.
Schuiringa, M.H. 2014. *Pronomi Clitici e Acquisizione L2. Una Ricerca Sull'Acquisizione L2 Adulta Di Pronomi Clitici in Italiano*. Ph. D.thesis, Utrecht University.
Schwarze, Christoph. 2001. On the representation of French and Italian clitics. In *Proceedings of the LFG '01 Conference*, pages 280–304, CSLI Publications.
Zwicky, Arnold M. 1977. *On Clitics*. Bloomington: Indiana University Linguistics Club.
Zwicky, Arnold M. and Pullum, Geoffrey K. 1983. Cliticization vs. Inflection: English n't. *Language* 59(3), 502–513.