IMPLEMENTING THE ROMANIAN ACCUSATIVE CLITIC PRONOUNS IN FLUID CONSTRUCTION GRAMMARS

BY

LIVIU CIORTUZ and VLAD SAVELUC*

“Al. I. Cuza” University of Iaşi,
Faculty of Informatics

Received: February 19, 2014
Accepted for publication: March 27, 2014

Abstract. The aim of this paper is to present a set of core properties of the accusative clitic pronouns in Romanian and an implementation in the framework of the Fluid Construction Grammars (FCG), as a system that parses and produces − starting from a given semantical description − sentences, based on the stated properties.

Key words: parsing, generation, Fluid Construction Grammars, clitic pronouns.

2010 Mathematics Subject Classification: 68T50.

1. Introduction

The Romance languages exhibit clitic pronouns which are quite problematic from both the linguistics perspective and the computational implementation point of view because these “weak pronouns” have an ambiguous and therefore debatable status: they share some properties of words while borrowing other properties from affixes.

The Romanian language has a clitic system viewed by some researchers as one of the most complicate among all Romance languages. An early attempt
Liviu Ciortuz and Vlad Saveluc

106

to formalize the properties of Romanian clitics in a computational grammar setup has been made by Paola Monachesi (1998; 2000) who used the HPSG theory (Pollard & Sag, 1994); she arguably opted for associating an affix status to these clitics. Two more recent presentations of the Romanian clitic system from the computational linguistics perspective are due to Udo Klein (2007) and Oana Ciucivara (2009).

As for real implementations addressing the Romanian clitics in the unification-based approach, up to our best knowledge there is none (yet). This is why we considered that it would be truly beneficial to provide such an implementation, hopefully enabling other researchers to do further work on this issue. As development platform, we took advantage of the very useful grammar development facilities offered by the Fluid Construction Grammar (FCG) framework (Steels, 2011), which was implemented by professor Luc Steels’ research teams based at Vrije Universiteit in Brussels and “Sony” Computer Science Laboratory in Paris. The FCG platform enables a bi-directional (i.e. both parsing and production) construction-based treatment of sentences that we deem suitable for modeling the rather sophisticated properties of the Romanian clitic system. The FCG design and some of its computational aspects have been presented in (de Beule, 2012). Comparison with other recent unification-based formalisms like Embodied Construction Grammars (Chang, 2008), and Sign-Based Construction Grammars (Boas & Sag, 2012) can be found in (Chang et al., 2012) and respectively (van Trijp, 2013a).

In recent years, FCG-based implementations for a number of interesting linguistic phenomena have been documented:
− for German: the definite article (van Trijp, 2013b), the case marking and agreement (van Trijp, 2011), and the phrasal constituent order (Micelli, 2012);
− for Hungarian: the agreement (Beuls, 2011);
− for Spanish: the evolution of some pronouns (van Trijp, 2010);
− for Polish: complex declension systems (Höfer, 2012);
− for Russian: the verbal aspects (Gerasymova, 2012).

We would like to add to the above list the fact that one of our previous works described some aspect-like constructions present in a dialect of the Romanian language spoken in Banat, a region in south-western Romania which has been influenced by the Serbian language (Ciortuz & Saveluc, 2012).

The organisation of this paper is as follows: Section 2 will introduce the main properties of the accusative clitic pronouns in Romanian, and Section 3 will detail an example for parsing and producing a sentence in our system. The on-line (www.info.uaic.ro/~ciortuz/PAPERS/ALL/cl.pdf) version of this document will guide the reader through the list of categories of constructions that we have implemented in FCG in order to process accusative Romanian clitics. For the constructions’ code, the reader should consult www.info.uaic.ro/~ciortuz/ ALEAR/RomanianClitics/romanian-clitics.lisp. The on-line version of this paper also presents a set of sentences that played a
testbed role while we were developing our implementation; it also shows the way in which these sentences cover the properties introduced in Section 2.

From now on, whenever we will use the term clitics in this paper it will refer to accusative clitic pronouns in the Romanian language. The case of dative clitics is considerably less sophisticated, although the combination of the dative clitics with accusative clitics is not trivial. Hopefully we will deal with these two issues elsewhere.

2. Properties of the Romanian Accusative Clitic Pronouns

Our presentation of the core/main properties of the Romanian accusative clitic pronouns was inspired by (Klein, 2007) –Udo Klein’s mother tongue is Romanian–, but it goes significantly further. Indeed, we revised and then extended the set of properties that Klein has stated as modeling the behavior of direct objects (DOs) and the associated clitics in Romanian. For the sake of conciseness, here we will list the resulting set of properties; our specific contributions will be signaled whenever appropriate.

Section 2.1 introduces the conditions for marking the DOs with the accusative case marker. Section 2.2 does the same regarding the conditions for the co-occurrence of DOs and the accusative clitics. Section 2.3 states a condition under which the clitic can render the presence of the DO optional. The eager reader could jump to Section 3 to see how parsing/production proceeds in our FCG-based system for a sample sentence that contains one clitic; to understand the grammaticality of that sentence, he will only need to know about the properties M1 and C2 presented below.

A short explanation is probably useful here, concerning the cases (N – nominative, G – genitive, D – dative, A – accusative) of Romanian nouns. Nouns in Romanian have two distinct morphological forms, one for N and A, and the other for G and D. (For instance, ‘băiatul’ in N/A and ‘băiatului’ in G/D.) As seen, the distinction between the two morphological forms is made by a suffix that characterizes the G/D case. Further more, the distinction between the G and D cases is made by a prefix (a possessive article) for G. (For instance, ‘al băiatului’.) Distinguishing between N and A cases is possible by using a preposition/marker, for instance ‘pe’. These case distinctions are needed due to the rather high degree of freedom allowed in ordering the constituents of Romanian sentences. For instance, the sentence “Mama îl vede pe Ion” (Mamma sees John) can exist – with different attached stresses and different probabilities – in five permutations: “Mama, pe Ion îl vede”; “Pe Ion îl vede mama”; “Pe Ion, mama îl vede”; “Îl vede mama pe Ion”; “Îl vede pe Ion mama”.


2.1. Accusative Case Marking

**M1.** In Romanian, DOs are obligatorily marked for the accusative case – namely using the preposition ‘pe’ – if they are expressed by either
(M1.1) proper nouns (names) referring to animate beings or
(M1.2) personal pronouns referring to animate entities.

**M2.** DOs are optionally ‘pe’-marked if
i) they refer to animate beings and
ii) they are expressed by definite (M2.1) or indefinite (M2.2) noun phrases (NPs) which are modified.

We mention that Romanian nouns can be either
− **definite:** the definite article is a suffix and it has different forms according to the noun’s gender, number and case;
− **indefinite:** the indefinite (i.e. non-specific) article is a proclitic, also varying on gender, number and case;
− **un-definite:** this situation is simply characterized by the lack of definite or indefinite articles attached to the noun.

Note that the property M2 refers to animate beings expressed as definite or un-definite modified NPs. The case of un-definite NPs referring to animate entities – not dealt with by Klein – will be treated by the next property (M3).

Before stating it, we make the following two important remarks:

**Remark 1:** Here above and also in the sequel the term noun phrase designates a noun together with its attached articles and modifiers; for convenience, here we will refer to pronouns distinctively.

**Remark 2:** The term modification (of noun phrases or pronouns) used here is limited to (simple) gerundive and relative clauses. We did not (yet) extend our grammar so to accept other modifying clauses, although in some cases (for example, simple genitive clauses) it would be straightforward. Klein did not discuss at all the modification issue in his approach.

**M3.** ‘Pe’-marking the DOs expressed as un-definite NPs is decided as follows:

(M3.1) If the DO refers to animate beings and is supposed to be identified by the hearer, then it is obligatory to ‘pe’-mark the DO.

(M3.2) Otherwise, ‘pe’-marking is ungrammatical.

Two more remarks could be added now:

**Remark 3:** In the M3.1 case, the DO can be modified by a gerundive clause, but not by a relative clause (unless it is preceded by comma).

**Remark 4 (Concerning M3.2):** Without ‘pe’-marking, it is not possible to have a DO expressed by an un-definite NP (either animate or inanimate) whose number is singular (for example: * Ana vede băiat = Anna is seeing boy.;) the plural is allowed, with the possibility of further modification, by either gerundive or relative clauses (for example: Ana vede băieți = Anna is seeing boys).
M4. It is generally ungrammatical to ‘pe’-mark DOs which are referring to either
(M4.1) inanimate entities (expressed as either noun phrases or pronouns) or
(M4.2) animate entities expressed as unmodified NPs which are either
definite or indefinite. This second case was not treated by Klein.
Klein mentions that ‘pe’-marking of DOs is “obligatory with some
quantifying expressions and optional with others” but he does not elaborate any
further (nor do we).

2.2. Co-Occurrence of Direct Objects and Clitics

Here we state the conditions under which the pronominal accusative
clitics co-occur with DOs. The co-occurrence imposes that the two agree on
number, gender and case. Otherwise said, the rules given below concern the so-
called “clitic doubling” phenomenon.
C1. As a simple, general rule, a pair clitic–DO is ungrammatical if the
DO is not ‘pe’-marked.
C2. It is almost obligatory that the DO co-occur with the corresponding
clitic if the DO is a ‘pe’-marked NP (other than a pronoun). This condition is a
more restrictive than the one given by Klein.
The above “almost obligatory” specification has the following meaning:
if (in that case) the clitic is not used, then the resulting sentence sounds “old
fashioned”; it is perfectly understandable but is nowadays rarely used by
Romanians.
C3. The co-occurrence clitic–DO is obligatory if the DO is a pronoun
(for either animate or inanimate entities).

We can add to C2 the following restriction (not mentioned by Klein): If
the DO is a personal pronoun, then it cannot be modified by a relative clause
(for example: * Ana îl vede pe el care plânge = Anna sees him who cries,);
however, a relative pronoun (cel, cea, cei, cele) must be, and a demonstrative
pronoun (acesta, aceea, aca, aceea, acaia, acalea) can be. If the DO is a
demonstrative pronoun, then it cannot be modified by a gerundive clause (For
example: * Ana îl vede pe cel plângând = Anna sees him who cries,); however,
a personal pronoun can be.

2.3. Clitics and Null Direct Objects

This issue was not dealt with by Klein. We express it from the
production point of view.
N. An entity (either animate or inanimate) which is referring to a certain
DO (better: referring to the object role for a verb.) is allowed to nullify/delete
the DO realization at surface/form level if the following condition is met:
i. it is either unmodified or
ii. modified by a gerundive clause.

Note: The deletion does not affect the modifier; it may remain in place.
(For instance: Ana îl vede plângând = Anna sees him crying.)

3. A Running Example

In this section we will consider the sentence “Ana îl vede pe Ion”, i.e. 
Anna sees John. The clitic here is ‘îl’, which is the accusative, weak form of the 
third person, singular, masculine pronoun ‘el’. The reader remembers that the 
accusative case marker is ‘pe’. Concerning the grammaticality of the above 
given sentence, and referring to Section 4, the reader can easily see that 
− ‘pe’-marking of the DO ‘Ion’ is justified by the property M1.1 (the 
DO is a proper noun referring to an animate being); 
− the introduction of the clitic ‘îl’ is supported by the property C2 
(since the DO is ‘pe’-marked and it is not a pronoun).

Section 3.1 we will comment on the parsing of the above mentioned 
sentence while Section 3.2 will detail the production of the same sentence, using 
our FCG grammar. A complete link of the FCG constructions that we 
implemented is given in the Appendix of the on-line version of this paper (see 
mention on page 2). In order to illustrate the parsing and production processes 
and thus to help the reader follow step-by-step our explanations, we displayed 
the coupled feature structures created by the FCG framework in the file 
www.info.uaic.ro/~ciortuz/ALEAR/RomanianClitics/cl.example.dvi. We hope 
that the reader who is not acquainted with the FCG formalism and its operating 
mode will get an introduction to them via this example.

3.1. Parsing

The FCG parsing process applied to the input sentence “Ana îl vede pe 
Ion”, tokenized as the list (“ana”, “îl”, “vede”, “pe”, “ion”), will start with all 
the corresponding constraints placed via the form feature of the so-called top 
unit. These constraints are encoded in FCG via the meets predicate. The parsing 
− seen as a successive application of the construction rules that constitute the 
grammar − will be considered successful if in the end all the constraints on 
strings were moved in other units linked to the top unit and there are no more 
construction rule that can be applied.

Application of lexical constructions: The lexical rules to be applied 
are those for “ana” – anna-cxn, “vede” – see-cxn, and “ion” – john-cxn. New 
units will be created accordingly, in the syntactic and semantic poles of the top 
unit. The meets constraints on the above strings will be moved to the syntactic 
poles of the newly-built units. Additional attributes will be added to the syn-cat
feature, thus marking for ana and john the fact that each one of them is an animate being \textit{(i.e.} human or animal) expressed by a noun phrase consisting of a proper noun and replaceable by a third person singular pronoun. The gender is encoded as feminine for ana and masculine for john. In the semantic pole, two constraints will be introduced: (ana ?indiv-ana ?context) and (john ?indiv-john ?context).

For the verb “vede” (to see, indicative mood, third person, singular, present tense), the same steps will be followed: first the meets constraint on this string will be moved into a newly created unit, then the following attributes will be introduced in the syn-cat feature of this unit: category – verb, number – singular, person – 3rd, gender – unbound (i.e. unknown at this stage), and also the valence of the verb, in this case indicating that it can accept only direct objects. The meaning attribute’s value as it has been produced so far is:

\[(\text{anna }\text{?indiv-anna }\text{?context}) (\text{john }\text{?indiv-john }\text{?context}) (\text{see }\text{?event }\text{?context})\].

\textbf{Application of the noun phrase building constructions:} The next step is the application of the construction rule called add-np-attributes to the units expressing nouns: ana and john. This rule will add to the syn-cat feature of these two units

\begin{itemize}
  \item five attributes called: definite-article, indefinite-article, pronoun, accusative-marker, modifier, and
  \item one attribute for each of the following 3 cases: nominative, accusative and dative.
\end{itemize}

The value of each one of these attribute will be a new unbound variable that will later take the value + or − depending on whether the corresponding element is present or not.

Next, the rule marker-acc from the construction category called marker will be applied. It will consume from the top unit the preposition “pe” that marks the accusative case. This rule will be triggered only if this preposition is followed by a noun phrase, in our case john, and it will set the value of the accusative-marker attribute to +, the accusative attribute value to +, and the other two cases (nominative and dative) to −. The string expressing the marker will be moved from the top unit to a new subunit of the john unit.

Then, one set of somehow similar rules will be applied. These rules’ names all have the prefix negatively-bound and they will bound the attributes from the noun phrases whose values have remained unbound variables until now to the value −. So, for the unit ana all the 5 attributes introduced by the add-noun-attributes rule will be bound to the value 0. Similarly, for john all these attributes will be set to 0, with the exception of accusative-marker which has been previously set to + by the marker-acc rule.

The next rule to be applied is called set-nominative-case and it will work on the unit ana. It will find that ana doesn’t have a marker, neither a
definite nor an indefinite article, and so it will infer that this noun phrase must be in the nominative case.

**Application of the verb phrase building constructions:** Next, a construction rule called move-direct-object-parsing will search for a verb situated right before a noun phrase and if so, it will designate the noun phrase as being a direct object of the verb. It will check whether the verb has the right valence and whether the noun phrase is in the accusative case and, if this is the case, it will make the noun phrase become a subunit of the verb’s unit. In our example, by applying the move-direct-object-parsing rule, the unit for john will be made a subunit of the verb’s (see) unit. In the semantic pole of the top unit, a new constraint will be introduced, (object ?event ?indiv-john), and the unit john will be bound to the verb’s object role.

In the next stage, the 5 attributes described earlier − definite-article, indefinite-article, accusative-marker, pronoun, modifier − and also the person, number and gender attributes from the syn-cat feature of the object will be moved into the syn-cat feature of the verb in a category called clitic-agreement-information. This work is achieved by applying the rule move-object-features-direct.

The above 5 attributes will be used to determine whether the verb must accept a clitic or not. This decision will be taken by a set of rules applied at the next stage. This set of rules, called determine-clitic, contains multiple patterns that will match the information which has been added by the move-object-features-direct rule. They will introduce two new attributes into the syntactic category of the verb, namely has-clitic and no-clitic. These two attributes will have the values + or − according to the properties described in Section 2. In our example, the rule determine-clitic-1 will be used. This rule says that, when the object is a proper noun, the verb must be assigned a (preceding) clitic.

At this point, only one string has remained in the top unit, namely the clitic “îl”. It will be consumed by the rule which is applied next, clitic-3sg-m-ace. Indeed, it checks whether the verb accepts a clitic and whether the clitic found in the input sentence (in this case, in front of the verb) agrees with the verb’s object with respect to the case, person, number and gender. This agreement information is found in the attribute clitic-agreement-information of the syn-cat feature of the verb.

**Application of the sentence construction:** The final rule to be applied is called sentence, and it will compose a sentence out of a subject and a predicate. In our case, anna will be the subject, and the verb see will be the predicate. It will check whether the subject is in the nominative case and whether it agrees with the predicate with respect to the number and person. In the semantic pole, the subject will be linked with to the predicate’s agent, yielding the following final meaning in the semantic pole of the top unit:

It expresses the fact that the event see has anna as agent and john as object.

3.2. Production

The production/generation process for the sentence “Ana îl vede pe Ion” goes (up to a certain extent) similarly to the parsing but in the reverse order. Starting from the next set of constraints, which are assigned to the value of the meaning feature in the semantic pole of the top unit, namely


we will illustrate how the original sentence can be produced. In general, the production process will be considered successful if, in the end, the top unit will become void of meaning constraints and there will be no more constructions that can be applied.

**Application of lexical constructions:** As in the parsing process, the lexical rules will be applied first. This time they will remove the meaning constraints for anna, john and see, placing them into the semantic pole of new units. The strings “anna”, “john” and “see” will be introduced in the syntactic poles of these new units.

**Application of the sentence construction:** Next, the sentence rule will be applied. It can be done early in the production process (much earlier than in parsing) because the agent and object roles for the event identified by the variable ?event are determined by the input meaning. The application of the sentence rule will introduce in the syntactic pole of the top unit a meets constraint on anna and see, implying that the word “ana” will be placed before the word “vede” in the final sentence.

**Application of the noun phrase building constructions:** Now the rule move-direct-object-production will be applied. Its effects are similar to those it produced in parsing. The only difference is that it will introduce a precedes constraint instead of the meets constraint, so as to place the object after the verb. The difference is that the meets constraint enforces contiguity, while precedes does not.

Afterwords, the rule add-np-attributes will be applied on the noun phrases anna and john. This rule is a syntactic one (meaning that its two poles are syntactic), having the same role in parsing as in production: it will add certain attributes to noun phrases. Further on, all the negatively-bound-... rules
will be applied, except the negatively-bound-marker. They were already explained, so we will not insist on them any more. The accusative-marker attribute will not be negatively bound because it was not expressed in the input meaning, but it will be determined later on.

**Application of the verb phrase building constructions:** Next, the rule move-object-features direct will be applied with the same effect as in parsing: it will impose the co-referenciation of certain attribute values between the verb and its direct object. Subsequently, the determine-clitic-1 rule will be applied. It will add the attribute has-clitic (valued +) to the verb’s syn-cat attribute, meaning that the verb needs a clitic. It will also bind the value of the attribute marker from clitic-agreement-information to the value +, implying that the object must be preceded by the accusative marker (“pe”).

The rule negatively-bound-marker will be applied to the unit of anna, bounding the value of accusative-marker attribute to −. Because the value of this attribute is set to + for john, the rule marker-acc will be triggered and it will introduce the marker “pe” in front of “ion”. Because it was already determined by the rule determine-clitic-1 that the verb must receive a clitic, the rule clitic-3sg-m-acc will be applied, effectively introducing the clitic “îl” in front of the verb. Now no more rules can be applied on the top unit, and putting together the values of the form feature from all sub-units in the top unit will give:

\[
\text{(string ?anna "ana") (string ?see "vede") (string ?john "ion") (precedes ?anna ?vede)
\text{(string ?clitic "îl") (meets ?clitic ?see) (precedes ?see ?john) (string ?marker "pe")
\text{(meets ?pe ?john)},}
\]

which can be uttered as: “ana îl vede pe ion”.

4. Conclusion and Further Work

This paper presents an overview of the FCG implementation that we produced for a core part of the system of accusative clitics in the Romanian language. We have shown that the tight integration of syntax and semantics throughout the parsing/production process in FCG makes it an attractive platform for implementing such a difficult-to-master phenomenon as an elaborated clitic system. Comparison with a fully declarative implementation, for instance in the spirit of SBCG (Boas & Sag, 2012), is one of our aims for the future. Also, the extension of the present system (and integration) with dative clitic pronouns remains to be done.
Acknowledgments. This work was supported by the “ALEAR” research project carried in the EU FP7 programme and also by its counterpart “ALEAR” 37EU funded by the Romanian Ministry of Education and Research.

REFERENCES


IMPLEMENTAREA PRONUMELOR CLITICE ACUZATIVE FOLOSIND GRAMATICI DE CONSTRUCȚII FLUIDE

(Rezumat)

Scopul acestei lucrări este acela de a prezenta un ansamblu de proprietăți de bază pentru pronumele clitice acuzative din limba română, precum și o implementare folosind cadrul computațional furnizat de către Gramaticile de Construcții Fluide (FCG). Această implementare constituie un sistem de parsare și generare (pornind de la o anumită descriere semantică) de propoziții, bazat pe proprietățile descrise în lucrare. Am arătat cum integrarea strânsă a sintaxei și semanticii pe parcursul procesului de parsare/producție din FCG a făcut posibilă implementarea unui astfel de fenomen dificil de descris cum este un sistem de pronume clitice foarte elaborat.