Formal Representation of Clitic Ordering in Serbian

Bojana Đorđević

Faculty of Philology
3 Studentski trg, Belgrade, Serbia
bojana@lingvistika.org

Abstract

This paper presents a formal representation of clitic placement and ordering in Serbian using a tree grammar formalism and eXtensible MetaGrammar (XMG).

Keywords: clitic ordering, Serbian, metagrammar, XMG, FB-LTAG

1. Introduction

Even though Serbian is a language with an almost unconstrained word order, clitics[1] are the one segment where, at least partially, strict rules apply. The first rule is that clitics appear in the second position in a sentence. Unlike clitics in French, Spanish or Italian, clitics in Serbian do not order themselves with respect to the verb, but exclusively with respect to the position inside the sentence. The first position can be understood as the first prosodic word (1), or as the first phrase (2)[2]. In Serbian, clitics can be placed after both[3].

(1) Taj čovek *joj ga je* poklonio

That man her it AUX presented.
'That man presented her with it.'

(2) Taj *joj ga je* čovek poklonio.

That her it AUX man presented.
'That man presented her with it.'

The second rule is that clitics form an inseparable clitic cluster that cannot be interrupted by any other unit. The third rule is that inside the clitic cluster, clitics always appear in a fixed word order. We will look at that order in section 2.

All of these rules raise a question about the nature of clitics and the representation in computational models that would capture all their peculiarities. In this paper, we will propose several ways to tackle the word order between clitics in a TAG grammar for Serbian, by using the methodology of eXtensible MetaGrammar (XMG). We will only briefly touch on the word order in the sentence containing clitics. This work is part of the general effort to create a formal grammar for Serbian.

This paper is organized as follows. The second chapter gives a short overview of the behavior of clitics in Serbian, together with their typology. The third chapter presents some of the approaches to clitics in TAGs. The fourth chapter shortly presents the methodology of XMG and its use for the representation of clitics in Serbian. Chapter five gives perspectives for further work on this topic.

2. Clitics in Serbian

There are four types of clitics in Serbian, each of them occupying its own unique position (Table 1).

1st position	Question particle LI
2nd position	Clitic forms of auxiliaries BITI (*to be*) and HTETI (*to want*), except for JE (3rd person singular present tense of JESAM – *I am*)
3rd position	Clitic forms of dative pronouns
4th position	Clitic forms of accusative or genitive pronouns
5th position	JE (3rd person singular present tense of JESAM) or SE

Table 1: Order of clitics in Serbian

The four types are the weak forms of personal pronouns in accusative, genitive and dative respectively, the auxiliary verbs, question particle LI and the reflexive particle SE.

[1] Under clitics we assume enclitics – clitics that form a syntactic unit with the word preceding them.
[2] Both examples are taken from (Halpern 1995).
[3] For crosslinguistic data on second position clitics consult (Bošković 2015).

Accusative and genitive clitics share most of the forms and since the number of examples in which they appear together is insignificant, in addition to the difficulty in determining the order between them in such cases, they are typically stated as alternatives for the fourth position (Browne 1975, Schütze 1994).

	AUX		
	JESAM	BITI	HTETI
	singular		
1st pers.	sam	bih	ću
2nd pers.	si	bi	ćeš
3rd pers.	je	bi	će
	plural		
1st pers.	smo	bismo	ćemo
2nd pers.	ste	biste	ćete
3rd pers.	su	bi	će
	DAT	ACC	GEN
	singular		
1st pers.	mi	me	me
2nd pers.	ti	te	te
3rd pers.	mu/joj	ga/je/ju	ga/je
	plural		
1st pers.	nam	nas	nas
2nd pers.	vam	vas	vas
3rd pers.	im	ih	ih

Table 2: Forms of flective clitics

JE and SE are mutually exclusive inside the clitic cluster. That is why in 3rd person singular past tense of reflexive verbs only SE is preserved in declarative sentences (3), even though all other forms of the verb to be and SE appear together in all other persons (4). SE, originally being the accusative clitic form of the full reflexive form *sebe*, will not appear together with accusative forms of personal pronouns.

What can also be observed is the special behavior of the auxiliary JE. Unlike other auxiliaries which take the second position inside the clitic cluster, JE is completely stranded and takes the final, fifth, position. Although we will not be dealing with that phenomenon in this paper, it is worth mentioning that of all the clitics, only JE can serve as support for other clitics, which suggest that there are two types of JE (5, 6). Progovac (2005) attributed these special capabilities of JE to the fact that, in comparison with other auxiliaries, it does not contain the agreement part but only the root of an auxiliary (all non-clitic forms of the verb JESAM contain JE as their base – *jesam* (I am), *jesi* (you are) etc.).

(3) Ona se smejala.
She SE laughed
She laughed.

(4) Ja sam se smejala.
I AUX SE laughed
I laughed.

(5) Je li Marija stigla?
AUX LI Marija arrived?
Has Marija arrived?

(6) *Si li stigao?
AUX LI arrived?
Have you arrived?

The complete list of flective clitic forms (2nd position, 3rd position, 4th position) is given in Table 2.

2.1. A Look at the First Position

There has been a lot of debate, especially in generative grammar frameworks, on how to define the first position. Some authors have suggested a purely syntactic approach under which syntax is responsible for the specific placement of the clitic cluster (Progovac 2005), others a largely prosodic approach (Radanović-Kocić 1988) and some have proposed mixed accounts (Zec and Inkelas 1990, Halpern 1995). There has also been a lot of discussion on the ability of clitics to split phrases (1, 2), sometimes completely changing the sentence landscape (7)[4].

(7) Mlađa *joj* odlazi ćerka.

Younger to her leaves daughter.
'Her younger daughter is leaving. / It is her younger daughter who is leaving.'

Without entering the mentioned controversy, we will mention only clear-cut examples of units that cannot serve as clitic hosts. Two most obvious examples are prepositions and some conjunctions. Prepositions, regardless of them being stressed or not (Progovac 2005, 137), cannot be clitic hosts. Conjunctions such as *i* (and) and *a* (and/but) act the

[4] Example taken from (Progovac 2005)

same, although some other conjunctions such as *ali* (but) and *ili* (or) can serve as hosts.

LI is a little more restrictive when it comes to the first-position hosts. It allows for verbs in the first position, but not noun phrases or adverbial phrases. It also doesn't allow for any conjunctions or prepositions to sit there. As mentioned in the first part of this chapter, JE can be the host for LI, but no other auxiliary can do that, unless we use their full forms.

Finally, if we look at clauses, a comma represents a start of a new sentence unit and requires a host before any clitics can be added.

3. Clitics in TAGs

In this chapter, we give a brief overview of some of the approaches to clitics within TAGs. The first approach we will consider is the one of Candito (1999) for Italian and French. Following Miller (1991) she adopted the treatment of clitics as verbal affixes. In her view, morphology is the determining factor for the order of clitics and their position with respect to the host, that is the verb. Rather than including clitics in the structure overtly, she introduces them through features.

In figure 1, in features under V in schema α, clC = - represents the lack of pronoun clitics in the structure. clS stands for the subject clitic, which is also lacking in α. In β, however, clC has the value acc, signifying that it contains the accusative clitic. The subject position is still occupied by a non-clitic. In γ, both the subject and the object are presented by clitics.

This model gives an elegant solution to the issue of clitics. However, we are not familiar with the mechanism of ordering between them.

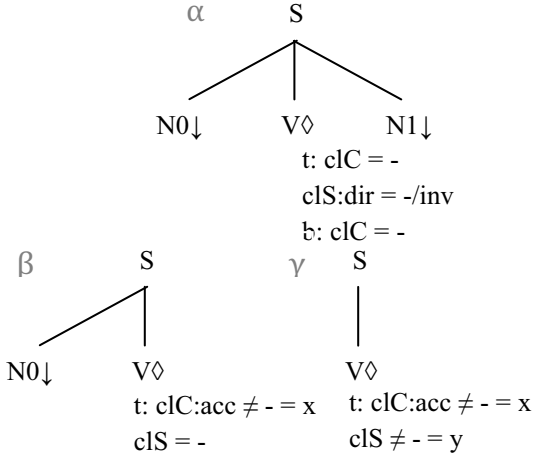

Fig. 1: Trees for the transitive sentence without clitics (α) and with them (β, γ). Taken from (Candito 1999)

Another significant approach is the one by (Abeillé 2007) for French. Under this approach, by using lexical rules, it is possible to transform a transitive sentence of type δ into type ε. The clitic position under the upper V is open for substitution in case of pronominal clitics, while auxiliaries need to be further adjoined at the lower V. This type of a structure at the same time solves the issue of the position of clitics.

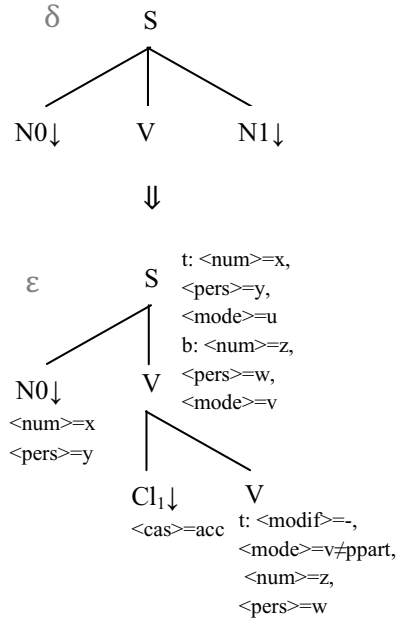

Fig. 2: Transformation into an elementary tree with a clitic complement. Taken from (Abeillé 2007)

4. Defining the order of clitics in Serbian using XMG

4.1. What is XMG

EXtensible MetaGrammar or XMG (Duchier et al. 2013) is both a metagrammar, or is a factorized model of a tree grammar, and a compiler for that grammar. Just like other metagrammars created late in the 20th century, XMG was envisaged as a practical tool for generation, maintenance and expansion of tree grammars, in particular FB-LTAG (feature-based lexicalized tree-adjoining grammar), MCTAG (multicomponent tree adjoining grammar) and IG (interaction grammar).

The main method XMG implements is abstraction – trees are searched for reoccurring fragments that are further organized into units of their own called classes. Those classes become reusable elements from which

we can create sentences by applying two operations, conjunction and disjunction. Inheritance between classes is allowed and well-constrained. Combinations of segments produce tree schemata, which then combine into families.

The following formula is an example of a transitive family:

Transitive → *(CannonicalSubject* ∨ *RelativeSubject)* ∧ *TransitiveVerb* ∧ *(CannonicalObject* ∨ *SentenceObject)*

The formula shows the basic principles of XMG – two classes of subjects (the canonical one and the relative one) are offered as alternatives for the subject position through disjunction; the same applies to the object – the offered alternative is between the canonical object and a sentence object. The three positions are then put together into a family using conjunction. The word order between them can be specified through linear precedence relations.

4.2. The Serbian Model

While deciding on how to craft the Serbian model, we have considered several options.

The first option was to treat all clitics equally, and have them adjoined to the same position, namely to V, as presented in Figure 4. In this model, the work on clitic ordering is very much simplified by a property defined in XMG – the property of rank. The property of rank allows us to define the rank for each of the clitics, which, in the final outcome produces an immaculate order between the clitics.

Word order in respect to verbal arguments can be solved through using precedence relations – between the clitics and the verb and the noun phrase and the verb.

What remains as a source of some redundance in this model is the way to distinguish between the models with full arguments and the ones with their clitic counterparts, and prevent the appearance of both in a single tree. This can be done through the use of features, much like the ones proposed by Candito.

We have experimented further and tested a model in which pronoun clitics would be treated differently from other clitics. While other clitics would be added through adjunction, pronoun clitics would be treated equally as verbal arguments with the same function. In the case of a direct object, we stipulate the existence of two types of direct objects – the clitic and the full ones. This model suggests a flat structure in both types of trees, as shown in Figure 3.

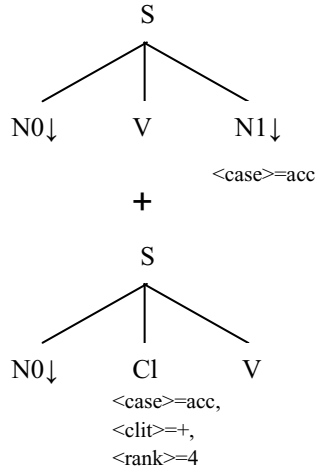

Fig. 3: A pair of trees for transitive verbs showing the same treatment of full verbal arguments and clitics

However, even though the order between the pronoun clitics and auxiliary clitics can easily be established through the property of rank, adjunction to this kind of a model gives faulty TAG trees in terms of clitic ordering, shown on Figure 5.

The third option would be to use the model proposed by Abeillé, although in the Serbian model, auxiliaries, together with LI, would have to be adjoined to the upper V and the reflexive SE to the lower V. In terms of the word order outside the clitic cluster, this model is restrictive and would not allow for all the possible word orders, such as the verb-first one.

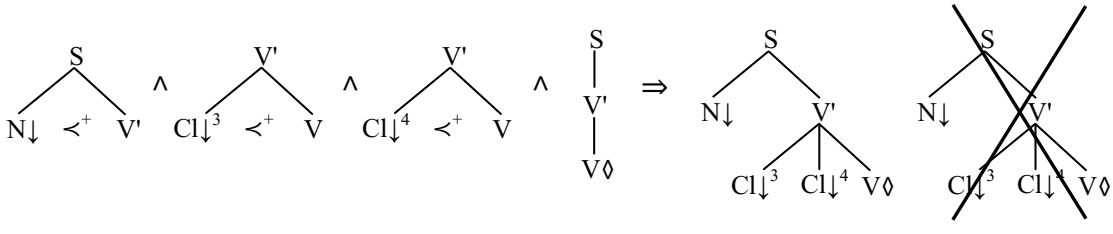

Fig. 4: Ordering of clitics using XMG. Taken from (Crabbé et al. 2013)

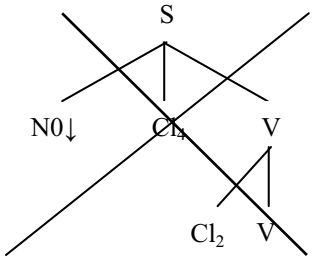

Fig. 5: Faulty clitic order by applying adjunction of auxiliaries to V

4. Conclusion and Further Work

This paper presented the initial effort on the work of clitic ordering in Serbian. Along with covering the cluster internal clitic order, we are hoping to cover the basic orders of the clitic cluster inside the sentence using the first proposed model. We have yet to tackle the nature of JE – as a copula and as the past tense auxiliary in the first position in the sentence. We also allow for the possibility to expand and change the model, especially after experimenting with different TAG formalisms such as MC-TAG.

References

Abeillé, A. (2007). *Les grammaires d'unification. Paris: Lavoisier*, Hermes Science.

Bošković, Ž. (2015). *On second position clitics crosslinguistically*. http://web.uconn.edu/boskovic/OresnikCliticsRefrences.pdf

Candito, M. (1999). *Representation hierarchique de grammaires lexicalisees: application au francais et a l'italien*, PhD Thesis, University Paris 7

Crabbé, B., Duchier, D., Gardent, C., Le Roux, J., & Parmentier, Y. (2013). *XMG: eXtensible MetaGrammar*. Computational Linguistics, *39*(3), 591-629.

Halpern, A. (1995). *On the Placement and Morphology of Clitics*. Standford: CSLI.

Miller, P. H. (1991). *Scandinavian extraction phenomena revisited: Weak and strong generative capacity*. Linguistics and Philosophy, *14*(1), 101-113.

Progovac, Lj. (2005). *A syntax of Serbian clausal architecture.* Bloomington: Slavica Publishers.

Radanović-Kocić, V. (1988). *The grammar of Serbo-Croatian clitics: A synchronic and diachronic perspective.* Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

Schütze, C. (1994). *Serbo-Croation Second Position Clitic Placement and thePhonology-Syntax Interface*, in A. Carnie and H. Harley, eds, Papers inPhonology and Morphology, MITWPL 21 373-465, Cambridge, MA

Zec, D. and Inkelas, S. (1990). *Prosodically constrained syntax.* In Sharon Inkelas & Draga Zec (Eds.), *The phonology-syntax connection*, Chicago: University of Chicago Press, 365–378