



Mario Einaudi
Center for
International
Studies

No. **13-07**

Working Paper Series

Clitics in South Slavic Languages: The View from the Interfaces

Molly Diesing and Draga Zec
December 2007

To view past working papers and guidelines for submission, please visit the Mario Einaudi Center for International Studies Working Paper Series webpage at: <http://www.einaudi.cornell.edu/initiatives/working.asp>. For more information please contact Dr. Heike Michelsen, 170 Uris Hall, Tel: 607.255.8926, hm75@cornell.edu.

Clitics in South Slavic Languages: The view from the interfaces

By

Molly Diesing and Draga Zec

Abstract

This paper analyses the placement of clitics that occupy the so-called "second" position in Serbian, in which both the first word or the first constituent can serve as host positions for clitics. In both corpus investigations and experimental research, we found that in Serbian there is more than one type of first position, both in the case of first word, and in the case of first constituent. Moreover, we found two types of cases depending on whether the sentence initial element is, or belongs to, either an argument or the predicate, yielding a four part classification. The experiments clearly establish preferred clitic placement in the two types of sentences. All four types are represented both in the investigated corpora and in the production and perception patterns, albeit in very different proportions. We attribute these differences to different discourse conditions between the first word and first phrase positions within each category.

About the Authors

Molly Diesing (Ph.D. University of Massachusetts, Amherst, 1990) is a Professor of Linguistics at Cornell University. Her main research interests are syntax and the interface between syntax and semantics. Her primary language focus has been the Germanic languages, most notably German and Yiddish. Some of the empirical issues she has concentrated on in recent syntactic work have been argument structure, clause structure, and parameters determining cross-linguistic word order variation. Her work on the syntax/semantics interface centers on the issue of determining the role that syntax plays in deriving the semantic interpretations of noun phrases (including quantifier phrases and pronominals). Among some of the specific issues she is interested in are syntax/semantics interactions in various word order and extraction phenomena, and the syntax and semantics of aspect.

Draga Zec (Ph.D. Stanford University, 1989) is a Professor of Linguistics at Cornell University. Her research focuses on phonological theory, a study of the principles that govern the patterning of sound in individual languages, as well as cross-linguistically. She has worked in several areas of phonology and its interfaces: on the moraic theory of syllable structure, the representation of pitch accent, and both the phonology-morphology and the phonology-syntax interfaces.

Contact Information

Professor Molly Diesing, Professor of Linguistics, Cornell University, 211 Morrill Hall, Ithaca NY 14853, tel. 1-607-255-8635, e-mail: md20@cornell.edu

Professor Draga Zec, Professor of Linguistics, Cornell University, 219 Morrill Hall, Ithaca NY 14853, tel. 1-607-255-0728, e-mail: dz17@cornell.edu

Clitics in South Slavic Languages: The view from the interfaces

1. Introduction

The focus of our project is the distribution of clitics, that is, weak forms of pronouns and auxiliaries, in larger linguistic units. In particular, we focus on the phenomenon of clitic placement in languages where clitics occupy the so-called “second” position. In this respect, languages fall into three classes: those in which the sentential position for clitics is after the first word, those in which clitics come after the first constituent, and those in which clitics may come either after the first word or after the first constituent. We identified the class of South Slavic languages as being of particular interest for investigating this problem, since Slovenian, Croatian and Serbian exemplify these three classes of cases. According to descriptions in traditional grammars, as well as some formal linguistic studies (such as Browne 1974), clitics are positioned after the first constituent in Slovenian, after the first word in Croatian, and in either of these positions in Serbian.

Current work on Serbian clitics has focused on whether to characterize the second position placement as being primarily a result of syntactic mechanisms (Franks and Progovac 1994, Progovac 1996), or essentially prosodic in origin (Halpern 1995, Radanovič-Kocič 1996, Bošković 2001). Within the prosodic perspective, Zec (2005), further provides a definition of the second position occupied by clitics in prosodic terms, accounting for facts unexplained within the syntactic approach. However, this work also shows that, while important, the prosodic approach alone cannot capture the distribution of clitics, which can be fully captured by further invoking the structural approach.

An initial point of failure in many of these accounts is in not recognizing that there are differences among types of sentences in terms of their “markedness”. That is, it is not sufficient to judge the grammaticality of the sentences, but their appropriateness in a given context must also be gauged. Current research has also relied heavily on native speaker judgments that have been culled primarily from previously published work, or from interrogating native speaker linguists. While these are not uncommon methods in

our field, it is worth augmenting the database with other sources; in this case searches of corpus sources and a series of experiments designed to elicit judgments from naïve native speakers.

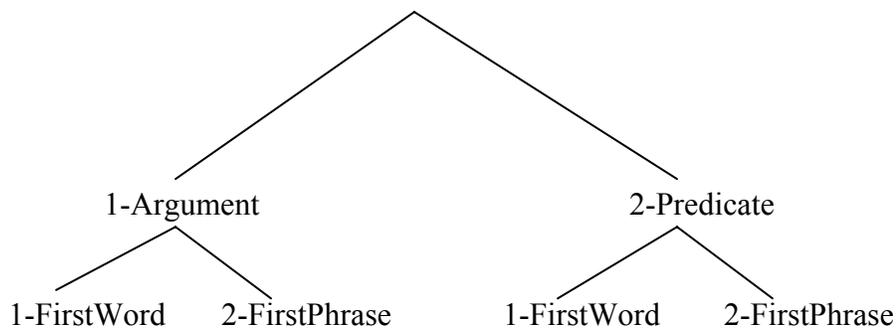
We started this project with the hypothesis that clitic placement can be properly understood as an interface phenomenon, that is, as viewed from the prosodic, structural and pragmatic vantage points. In our preliminary investigations we found that in all three cases the facts are more complex than they had been portrayed in the literature. In particular, we found that in Serbian there is more than one type of first position, both in the case of first word, and in the case of first constituent. We found four types of cases in each, yielding an eight part typology. The key factors distinguishing them are both (i) and (ii).

(i) structural: whether the element is, or belongs to, an argument (e.g., subjects or objects) or a predicate (e.g., verb phrases).

(ii) pragmatic – the role of information structure and intonation as its prosodic realization

The four types of cases are presented in (1):

(1) Four types of clitic placement in Serbian



Only two of the four cases we identified have been recognized in the literature: the two argument cases, with clitics following either the entire argument or its first word, as in (2). The predicate cases, those in which clitics may follow either the entire predicate or its first word, as in (3), have not been recognized in the literature.

- (2) a. Taj zadatak je veoma važan.
 that task is-CI very important
 “That task is very important.”
- b. Taj je zadatak veoma važan
 that is-CI task very important
 “That task is very important.”
- (3) a. Veoma je važan taj zadatak.
 very is-CI important that task
 “That task is very important.”
- b. Veoma važan je taj zadatak
 very important is-CI that task
 “That task is very important.”

We started with the hypothesis that all four cases figure in the grammar; and that, moreover, they differ in terms of pragmatics and information structure. We explored this hypothesis by two methods: data collection, and psycholinguistic experiments.

2. First phase: corpus investigation

In the first phase of our project, we collected data from two corpora of the Serbian language, one based on daily press and the other on literary prose. We got access to these corpora through collaboration with the Laboratory for Experimental Psychology, Department of Psychology, University of Belgrade. We analyzed 2993 sentences with clitics, 1323 from the corpus of daily press and 1670 from the corpus of literary prose. We placed each sentence in one of the four classes in (1). While working on this classification we decided to add another category, the adjunct, also subdivided into the first word and first constituent cases. Our results are shown in the table in (4):

(4)

| | | Daily press 1323 sentences | Fiction 1670 sentences | Daily press Proportions | Fiction Proportions |
|----|----------|----------------------------------|------------------------------|----------------------------|------------------------|
| a. | Arg-1W | 7 | 31 | 0.53 | 1.90 |
| b. | Arg-1Ph | 534 | 587 | 40.37 | 35.15 |
| c. | Pred-1W | 549 | 733 | 41.50 | 43.89 |
| d. | Pred-1Ph | 5 | 2 | 0.38 | 0.12 |
| e. | Adj-1W | 0 | 9 | 0.00 | 0.54 |
| f. | Adj-1Ph | 228 | 273 | 17.24 | 16.35 |

In the argument case, (4)a-b, we see a large proportion of the first phrase sentences, and a small proportion of the first word sentences. The situation is reversed in the predicate case, (4)c-d, where we find a large proportion of the first word sentences and a miniscule proportion of the first phrase sentences. The adjunct case, which we singled out hoping to achieve greater precision, follow the pattern of the argument case: again we find a large proportion of the first word sentences and a very small proportion of the first word sentences. It may well be that in our future work we will collapse the argument and the adjunct into a single category.

The results we got are striking in several respects. We found support for all types of cases we predicted. More importantly, we found that the two types of cases, arguments and predicates, have different default positions for clitics: the “normal” position for clitics in the argument case is after the first constituent, and in the predicate case, after the first word. In the second phase of our project, we tested this result by conducting a series of psycholinguistic experiments.

Before turning to the results we obtained from the second phase of our investigations, we need to report our corpus results in a more fine-grained form. As shown in table (5), we actually worked with a greater number of classes than the 6 classes presented in (4). Our classification of sentences included another parameter: whether the sentence has a topic constituent preceding whatever serves as first for the purposes of clitic placement or not.

(5)

| | | Daily press 1323 sentences | Literary prose 1670 sentences | Daily press Proportions | Literary prose Proportions |
|-----------|------------|----------------------------------|--|----------------------------|----------------------------------|
| 1 | Arg-1W-T | 1 | 5 | 0.08 | 0.30 |
| 2 | Arg-1W | 6 | 26 | 0.45 | 1.60 |
| 3 | Arg-1Ph-T | 130 | 122 | 9.83 | 7.31 |
| 4 | Arg-1Ph | 404 | 465 | 30.54 | 27.84 |
| 5 | Pred-1W-T | 372 | 310 | 28.12 | 18.56 |
| 6 | Pred-1W | 177 | 423 | 13.38 | 25.33 |
| 7 | Pred-1Ph-T | 1 | 1 | 0.08 | 0.06 |
| 8 | Pred-1Ph | 4 | 1 | 0.30 | 0.06 |
| 9 | Adj-1W-T | 0 | 2 | 0.00 | 0.12 |
| 10 | Adj-1W | 0 | 7 | 0.00 | 0.42 |
| 11 | Adj-1Ph-T | 80 | 41 | 6.05 | 2.46 |
| 12 | Adj-1Ph | 148 | 232 | 11.19 | 13.89 |

In the argument case, (5.1-4), a much larger proportion of sentences appear without a topic. This is certainly true of the first phrase sentences; the number of first word sentences is so small that it is hard to judge. Turning to the predicate case (5.3-8), we see a reversed situation in the daily press: two thirds of the sentences include a topic, and only one third does not. It is interesting that, in literary prose, the situation is reversed: more sentences appear without a topic. Finally, the adjunct case is again very similar to the argument case, with a greater proportion of sentences without a topic. In sum, we find differences between the argument and the predicate case in the distribution of the topic constituent. We also find a difference between the two corpora. Our decision to use more than one corpus was guided by the possibility that different registers may differ in the distribution of grammatical categories. The result we obtained regarding the distribution of the topic constituent clearly justifies this decision.

3. Second phase: psycholinguistic experiments

We conducted two psycholinguistic experiments, in which we tested the results obtained in the first phase of our project. Both experiments are based on a set of 120 sentences, specifically designed for this purpose. The sentences include two sets, 60 in each, one for the argument and the other for the predicate case. Within the set of argument sentences, there are three cases, each represented by 20 sentences, with the subject, object, and

prepositional phrase arguments. An orthogonal further division within the set of argument sentences is the presence of either the determiner or the adjective within the argument noun phrase. The set of predicate sentences is divided into three groups, with 20 sentences in each, representing three types of predicates, adjectival phrase (AP), noun phrase (NP) and verb phrase (VP). The table in (6) summarizes the types of sentences used in the experiments:

(6) Types of sentences used in the psycholinguistic experiments

| A. Argument | | | B. Predicate | |
|--------------------|------------|-----------|---------------------|----|
| | | 60 | | 60 |
| | Determiner | Adjective | | |
| Subject | 10 | 10 | AP | 20 |
| Object | 10 | 10 | NP | 20 |
| Prep Phrase | 10 | 10 | VP | 20 |

In experiment 1 we were interested in the differences between the two possible clitic positions in argument and predicate sentences manifested in language production. In experiment 2 we investigated these differences at the level of language perception, or processing. We conducted both experiment using the same set of sentences.

Our results demonstrate a clear difference between the preferred clitic position within the argument sentences on the one hand, and the predicate sentences, on the other hand. Argument sentences demonstrate a strong tendency towards clitic positioning after the first phrase. This tendency is present, both at the level of language production, and the level of perception, as well as at the levels of all the recorded measures. Argument sentences are more often completed by adding a clitic after the first phrase. At the same time, argument sentences are processed faster, and more likely to be accepted as grammatical when presented with a clitic after the first phrase. On the other hand, predicate sentences demonstrate an even stronger tendency towards clitic positioning after the first word. This tendency is also present at all of the considered levels: predicate sentences are almost always completed by placing a clitic after the first word, they are processed faster, and more likely to be accepted as grammatical when presented with a clitic positioned after the first word.

Below we present each experiment separately.

3.1 Experiment 1

Method

Participants: Thirty-eight students from The Department of Psychology, at The Faculty of Philosophy, in The University of Novi Sad participated in the experiment. All of the participants were native speakers of Serbian language, and had normal or corrected to normal vision.

Stimuli and design: One-hundred-and-twenty grammatical Serbian sentences (as described above) were presented in the experiment. The dependent variable was frequency of participants placing a clitic in one of the two possible positions, either after the first word or after the first constituent, for each of the two sentence categories, those with an argument, or with a predicate, in sentence initial position.

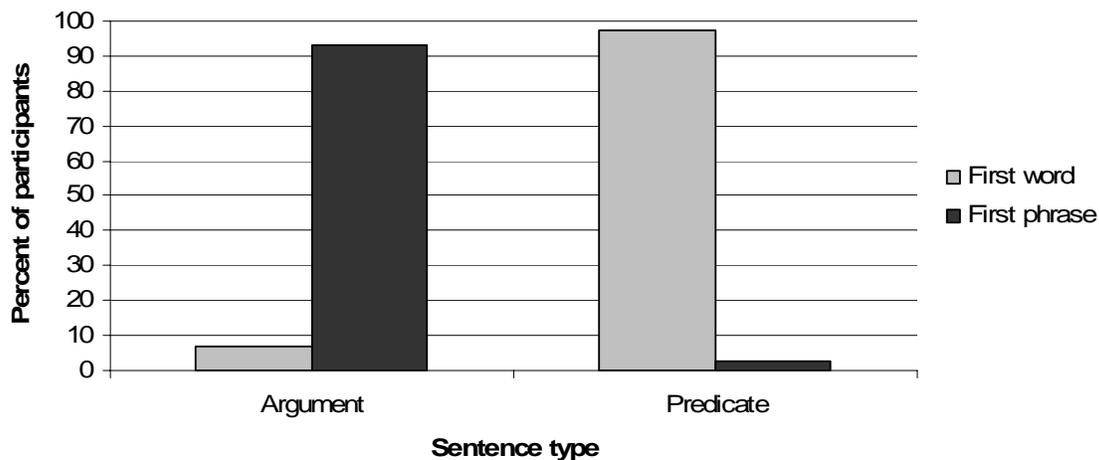
Procedure: Sentences were printed in a six-page booklet. Critical clitics were omitted from the sentences, and the two positions of clitics were replaced with a line, i.e. a blank to be filled in:

- (7) Njegov ___ auto ___ najbrži u gradu.
/His ___ car ___ fastest in the city./

There were three different random orders of sentences. Each participant was given a booklet with only one random order. Each booklet contained detailed instructions asking the participant to fill in only one of the two blanks using only one of the listed clitics. The task was to be performed in such a way as to make the sentence sound as natural as possible within the native language of the participant. Participants took approximately twenty minutes to complete the task.

Results

Analysis of responses obtained from the participants revealed a dramatic difference between clitic positions considering two sentence categories. While 92.98 % of participants placed a clitic after the first phrase in argument sentences, only 2.41% of participants placed a clitic after the first phrase in predicate sentences. The observed difference was significant: $\chi^2(1) = 1874.121, p < 0.01$ (Picture 1).

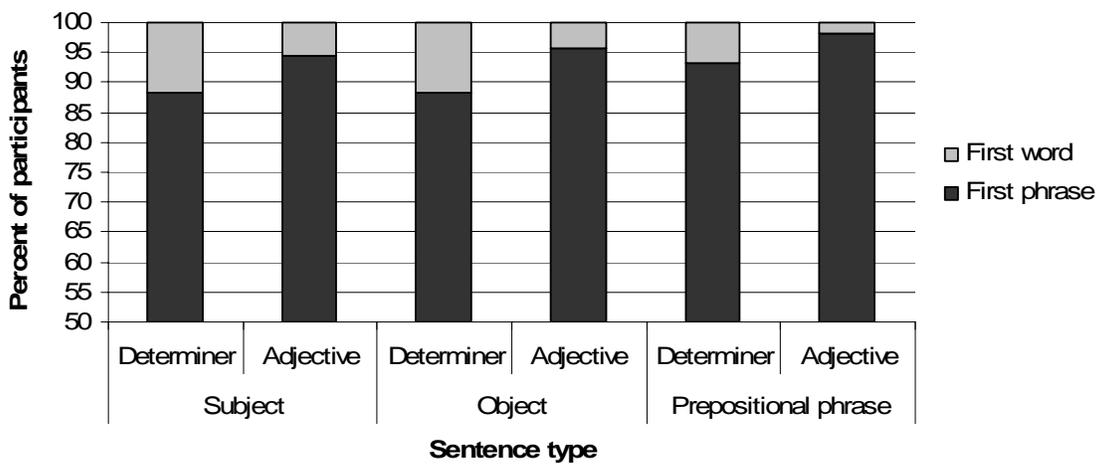


Picture 1: Percent of participants placing a clitic after the first word (light grey), and after the first phrase (dark grey) when completing argument (left), and predicate sentences (right) in experiment 1

The two sentence types were further analyzed separately in order to get more fine grained results. Analysis of the results obtained for argument sentences, apart from the fact that these sentences are predominantly completed by placing the clitic after the first phrase, also revealed that participants placed the clitic after the first phrase more often in the sentences that contain the prepositional phrase (95.66%) compared to the sentences that contain the subject, or the object (91.65%). At the same time, this tendency was present to a greater extent within sentences that contain an adjective (96.05%) compared to the sentences containing a determiner (89.91%). Sentences containing a determiner exhibited a greater tendency for the clitic to be positioned after the first word (10.09%), than adjective sentences (3.95%). These findings are summarized in Table 1.

Table 1: Percent of participants placing a clitic after the first word (left), and after the first phrase (right) when completing argument sentences containing a determiner, or an adjective in experiment 1

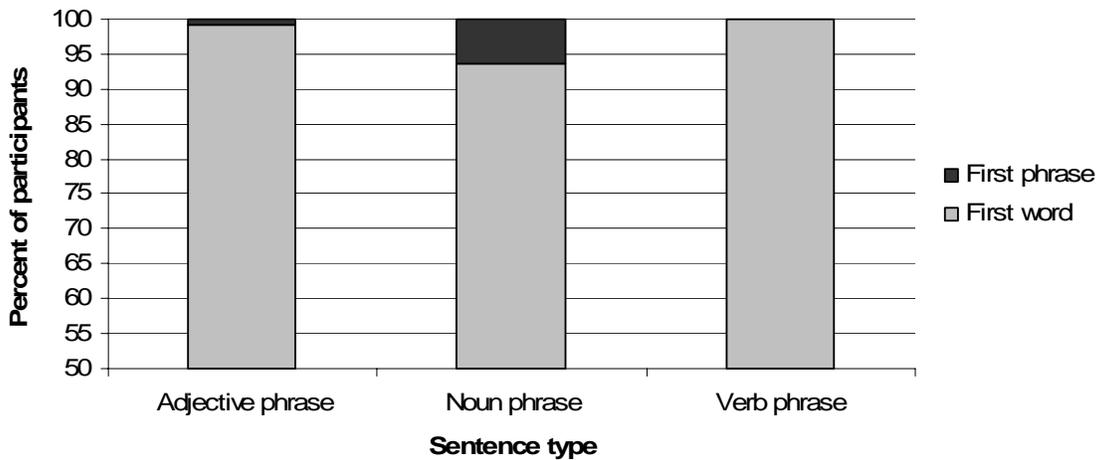
| | <i>First word</i> | | <i>First phrase</i> | |
|----------------------|-------------------|-----------|---------------------|-----------|
| | Determiner | Adjective | Determiner | Adjective |
| Subject | 11.84 | 5.53 | 88.16 | 94.47 |
| Object | 11.58 | 4.47 | 88.42 | 95.53 |
| Prepositional phrase | 6.84 | 1.84 | 93.16 | 98.16 |



A more detailed analysis of the predicate sentences confirmed a near 100% positioning of the clitic after the first word in case of the sentences that contain an adjective, or a verb phrase. The only sentence type that exhibited a slight presence of the clitic positioned after the first phrase were the sentences containing a noun phrase (6.31%). These findings are summarized in Table 2.

Table 2: Percent of participants placing a clitic after the first word (left), and after the first phrase (right) when completing predicate sentences containing an adjective phrase, a noun phrase, or a verb phrase in experiment 1

| | <i>First word</i> | <i>First phrase</i> |
|------------------|-------------------|---------------------|
| Adjective phrase | 99.08 | 0.92 |
| Noun phrase | 93.69 | 6.31 |
| Verb phrase | 99.87 | 0.13 |



Discussion

In experiment 1 participants were completing a list of sentences by placing the clitic in one of the two given positions in order to produce an adequate sentence of their native language. The two positions were marked by the blanks placed after the first word, and after the first phrase. The obtained results revealed a clear distinction considering the clitic position for the two types of sentences. The argument sentences predominantly contained a clitic placed after the first phrase, while the predicate sentences were predominantly completed by placing a clitic after the first word. The percent of participants placing a clitic after the first phrase was very high for all of the six subgroups of the argument sentences. The clitic was placed slightly more frequently after a prepositional phrase, than after a phrase marking the subject, or an object in an argument sentence. The sentences containing a determiner exhibited a greater tendency for the clitic to be placed after the first word than did the argument sentences containing an adjective. When it comes to predicate sentences, almost all of the participants placed the

clitic after the first word, except for the predicate sentences containing a noun phrase where a small number of participants placed the clitic after the first word.

3.2 Experiment 2

Method

Participants: Forty-eight students from The Department of Psychology, at The Faculty of Philosophy, in The University of Novi Sad participated in the experiment. All of the participants were native speakers of Serbian language, and had normal or corrected to normal vision. Participants were randomly assigned to only one of the two experimental blocks.

Stimuli: One-hundred-and-twenty target sentences from experiment 1 along with additional 120 ungrammatical Serbian sentences (control sentences) were presented in the experiment. One-hundred-and-twenty ungrammatical control sentences were constructed to mirror the syntactic structure of the target sentences that were presented in the experiment. Ungrammaticality was achieved by choosing a clitic that fails to agree with the verb.

Design: Sentences were constructed to fit 2x2 factorial design. Half of the sentences began with an argument and the other half began with the predicate. For each sentence, the place of the clitic alternated between two possible positions: after the first word, and after the first phrase. Clitic position was balanced in a two block latin square design. Sentences that appeared with a clitic after the first word in one block, would have a clitic positioned after the first phrase in the second block, and vice versa. This way, all of the sentences appeared with a clitic in both positions, and all of the participants were presented with all of the sentences, and both clitic positions, but none of the participants was exposed to the same sentence twice.

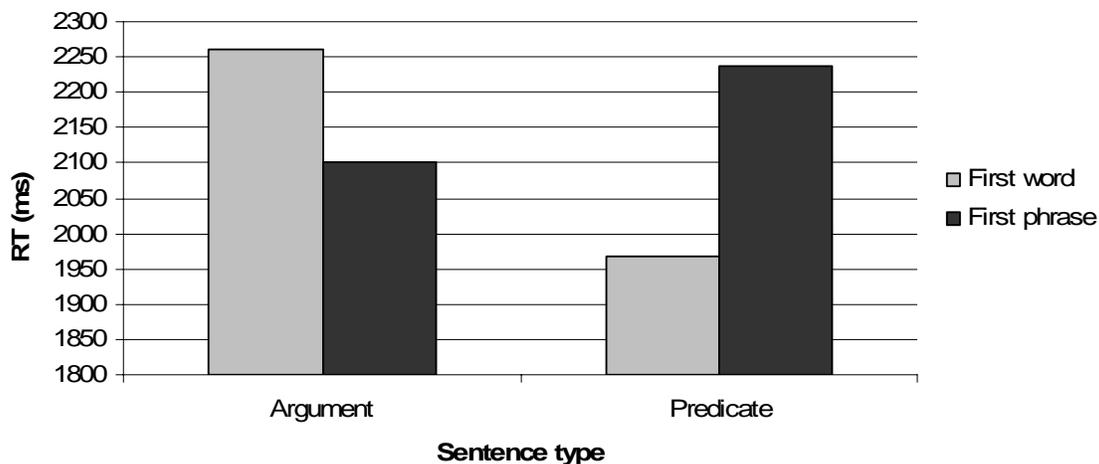
Procedure: Stimuli were presented in a sentence grammaticality judgment task. The participants were given instructions to judge whether the sentence appearing on the screen was acceptable in their language. They were told to base their answers on their intuitions as native speakers, and that there would not necessarily be right or wrong answers. Sentences were presented one-by-one, in a random order, on a computer screen. Prior to each sentence a fixation point was presented for 2000 ms. A sentence would

remain on the screen until participant's response, but its duration was limited to 8 seconds. Participants were given twelve practice trials. Sentences appearing in the practice trials were not included in the analysis.

Results

All analyses were conducted on the responses marking the acceptance of the target sentences. A by-participant analysis of variance of reaction time revealed a significant main effect of sentence type: $F(1, 41)=19.745$, $p<0.01$ ($F(1, 118)=3.200$, $p=0.08$, by item). Predicate sentences elicited shorter processing latencies. There was no main effect of clitic position (although $F(1, 118)=6.031$, $p<0.05$, by item) , but there was a significant interaction between sentence type, and clitic position: $F(1, 41)=25.644$, $p<0.01$ ($F(1, 118)=94.744$, $p<0.01$, by item).

Argument sentences with a clitic positioned after the first phrase were processed faster than argument sentences with a clitic positioned after the first word, while predicate sentences with a clitic positioned after the first word were processed faster than predicate sentences with a clitic positioned after the first phrase (Picture 2).

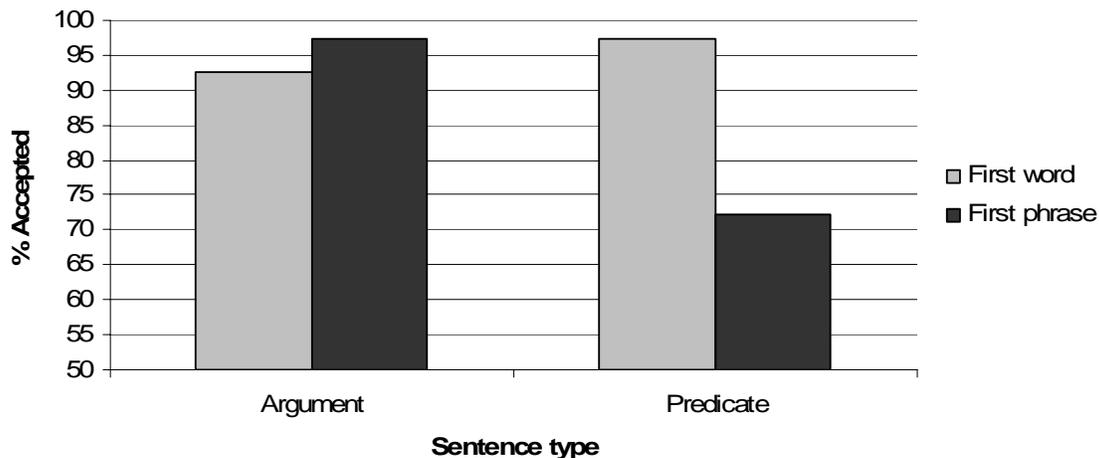


Picture 2: Mean reaction times for the argument (left), and predicate sentences (right) with a clitic positioned after the first word (light grey), and after the first phrase (dark grey) observed in experiment 2

A similar pattern of effects was observed in the analyses of acceptance rates.

Here, there was a significant main effect of both sentence type: $F(1, 41)=42.963, p<0.01$ ($F(1, 118)=35.944, p<0.01$, by item), and clitic position: $F(1, 41)=48.505, p<0.01$ ($F(1, 118)=43.371, p<0.01$, by item). Crucially, there was also a significant interaction between the two: $F(1, 41)=76.977, p<0.01$ ($F(1, 118)=92.117, p<0.01$, by item).

Argument sentences with a clitic positioned after the first phrase had higher acceptance rate than argument sentences with a clitic positioned after the first word, while predicate sentences with a clitic positioned after the first word had higher acceptance rate than predicate sentences with a clitic positioned after the first phrase (Picture 3).



Picture 3: Mean acceptance rates for the argument (left), and predicate sentences (right) with a clitic positioned after the first word (light grey), and after the first phrase (dark grey) observed in experiment 2.

4. Implications and future plans

4.1 Serbian: investigating the implications of our results

We have clearly established default clitic placement in the two types of sentences. While all four types are represented both in the investigated corpora and in the production and perception patterns in the psycholinguistic experiments we conducted, we observed a striking asymmetry between the two cases. Our next task is to address the pragmatic conditions under which non-default cases arise: clitics after the first word in the argument

case, and clitics after the first constituent in the predicate case. This will call for a new set of psycholinguistic experiments, as well as for a further investigation of corpora. In both cases we will have to focus our investigations on the contexts in which the studied sentences appear. In other words, we will have to identify the set of pragmatic and information structure factors that condition the non-default cases. We will also have to grant a more central place to the study of intonational differences between the default and non-default cases. We have already covered a lot of ground trying to understand the qualitative aspects of intonational differences. We have recorded several speakers who produced the targeted sentences in isolation. All recorded speakers exhibited a more marked intonational pattern in non-default cases: in argument sentences with clitics after the first word, in predicate sentences with clitics after the first constituent. In this case, too, we will have to introduce the context as an important factor in differentiating between default and non-default cases.

4.2 General

Once we have completed our analysis of the Serbian facts, we plan to turn to Slovenian and Croatian, and investigate clitic positioning in these languages using the same set of parameters that we used for Serbian. We may also turn to other languages with second position clitics that do not belong to the South Slavic group: Slavic language such as Czech and Polish, or to Australian Aboriginal languages such as Warlpiri or African languages such as Ngiyamba. Our plan is to submit a proposal to the National Science Foundation in January 2007. In sum, we believe that we have developed a methodology that will enable us to investigate the issue of clitic positioning from a general linguistic perspective. We also believe that the parameters we have established will enable us to capture the fine differences between individual languages, while at the same time identifying the core properties that these languages have in common.

References

- Bošković, Željko. 2001. *On the nature of the syntax-phonology*. Amsterdam: Elsevier Science.
- Browne, W. (1974). On the Problem of Enclitic Placement in Serbo-Croatian. In: Brecht, R. and C. Chvany (eds.). *Slavic Transformational Syntax*. Michigan Slavic Materials, vol. 10: 36- 52. Ann Arbor.
- Franks, S. and Progovac, L. (1994) On the placement of Serbo-Croatian clitics. In: G. Fowler, H. Cooper, and J. Ludwig (eds.) *Proceedings of the 9th Biennial Conference on Balkan and South Slavic Linguistics, Literature and Folklore* Vol. 7, pp. 69-78. Bloomington: Indiana Slavic Studies.
- Halpern, A. (1995). *On the Placement and Morphology of Clitics*. Stanford, CA: CSLI.
- Halpern, A. and Zwicky, A. (eds.) 1996. *Approaching Second: Second Position Clitics and Related Phenomena*. Stanford, CA: CSLI.
- Progovac, L. (1996). Clitics in Serbian/Croatian: Comp as the Second Position. In: Halpern & Zwicky 1996:411-428.
- Radanovič-Kocič, V. (1996). The Placement of Serbo-Croatian Clitics: A Prosodic Approach. In: Halpern & Zwicky 1996: 429-446.
- Zec, D. and S. Inkelas. 1990. "Prosodically constrained syntax." In S. Inkelas and D. Zec (eds.) *The phonology syntax connection*. Chicago: University of Chicago Press. 365-378.
- Zec, D. 1993. "Rule domains and phonological change." In S. Hargus and E. Kaisse (eds.) *Studies in Lexical Phonology*. New York: Academic Press. 365-405.
- Zec, D. 2005. "Prosodic Differences among Function Words." *Phonology* 22:77-112