Word Order and Clitics in Bulgarian

Dissertation zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes

vorgelegt von
TANIA AVGUSTINOVA

Saarbrücken, im April 1997

Der Dekan: Prof. Dr. Karl-Heinz Ohlig

Berichterstatter: Prof. Dr. Roland Marti
Prof. Dr. Hans Uszkoreit

Tag der letzten Prüfungsleistung: 16. Juli 1997

Kurzfassung (Summary in German)

In der vorliegenden Dissertation wird der Zusammenhang von Klitika und Wortstellung im Bulgarischen untersucht. Die Dissertation entstand aus dem Interesse an einer Sprache, die trotz ihres vereinfachten Deklinationssystems, aber aufgrund eines gut entwickelten Mechanismus für klitische Replizierung, über beträchtliche Wortstellungsvarianz verfügt. Sprachübergreifend variiert das Verhalten der Klitika von Abhängigkeit morphologischer Affixe bis hin zu Autonomie selbständiger syntaktischer Formen. In dieser Hinsicht sind die bulgarischen Klitika wegen ihrer Zwischenstellung linguistisch besonders interessant.

Die in dieser Arbeit durchgeführte linguistische Forschung ist durch den Bedarf an einer expliziten formalen Beschreibung der bulgarischen Konstituentenstruktur und Wortstellung motiviert. Die formalen Fragen wurden aber in der vorliegenden Dissertation mit Absicht zurückgestellt, um die Analyse für einen möglichst breiten Leserkreis mit slawistischem Hintergrund nachvollziehbar zu machen. Dennoch impliziert dieser Mangel von Akzenten auf Formalisierung nicht, daß die dargestellte Theorie nicht formalisierbar ist. Die Tatsache, daß sie in der Form eines Parsers erfolgreich implementiert wurde, weist eindeutig darauf hin, daß strenge Formalisierung in der Tat möglich ist. Der erwähnte Parser liegt einem experimentellen Grammar-Checker für das Bulgarische zugrunde.

Als theoretischer Rahmen für diese Dissertation wurde die Head-driven Phrase Structure Grammar (HPSG) gewählt, da diese eine multidimensionale, aber trotzdem einheitliche Darstellung linguistischer Objekte ermöglicht. Die Vielschichtigkeit der strukturellen Relationen im bulgarischen Verbkomplex stellt die Angemessenheit und Allgemeingültigkeit eines lexikalistischen Ansatzes zur Behandlung von Klitika in Frage. In dieser Arbeit wird die Ansicht vertreten, daß die Klitisierung im Bulgarischen eine morphosyntaktische Dimension aufweist, wobei verbale Klitika zur Verbkomplexkonstituente gehören. In bezug auf das lexikalische Verb und den von diesem Verb regierten Satz ist die Verbkomplexkonstituente als eine Zwischenkonstruktion zu betrachten. Dementsprechend basiert die durchgeführte Analyse auf einer Variante von HPSG, die eine zusätzliche Dimension für die Modellierung der analytischen Verbmorphologie und der Klitisierung

vorsieht. Es werden drei Typen von Objekten unterschieden: lexikalische, morphosyntaktische und syntaktische. Der in dieser Arbeit eingeführte Begriff der morphosyntaktischen Markierung spielt eine zentrale Rolle in der Behandlung der analytischen Verbformen im Bulgarischen. Aufgrund der Konstituentenstruktur und des syntaktischen Verhaltens wird zwischen zwei Verbkomplextypen unterschieden: einerseits kompakte Verbkomplexe, die durch strenge Nachbarschaft ihrer Komponenten charakterisiert sind, andererseits zusammengesetzte Verbkomplexe, die aus zwei locker zusammenhängenden, aber nicht unbedingt adjazent stehenden Teilen bestehen. Formal wird die Klitisierung als morphosyntaktisches Phänomen und nicht als Teil des Lexikons aufgefaßt. Demgemäß wird die Stellung einzelner verbaler Klitika sowie deren Sequenzen auf der Ebene der verbalen Morphosyntax interpretiert, wo auch prosodische Restriktionen mitwirken. Infolgedessen sind Pronominalklitika keine Konstituenten auf der Satzebene.

Die entworfene Theorie der Konstituentenstruktur des Bulgarischen ermöglicht eine adäquate Darstellung der "zweiseitigen" Erscheinung von Objektklitika, die sich weder als echte Morpheme noch als vollständige syntaktische Konstituenten zufriedenstellend behandeln lassen. Die Intuition hinter der hier vorgeschlagenen Alternative besagt, daß ein gewisser Parallelismus in der Beziehung der Verbflexion zur Subjekt-NP und des Pronominalklitikums zur ensprechenden Objekt-NP besteht. Die Person-, Numerus- und Genusinformation über die Subjekt-NP ist in der Verbmorphologie vorhanden. Dieselbe Indexinformation, sowie Information über den syntaktischen Kasus der jeweiligen vollen NP-Ergänzung, wird von Pronominalklitika innerhalb des morphosyntaktischen Verbkomplexes geliefert. So stellen die beide Mechanismen — der morphologische der Verbflexion und der morphosyntaktische der Objektklitisierung — gleichartige Ergebnisse auf der Syntaxebene zur Verfügung, die die syntaktische Optionalität der entsprechenden NP-Konstituenten befördern.

Mit der Zulassung der morphosyntaktischen Konstituenz gewinnt die Sprachbeschreibung deutlich an Erklärungsmächtigkeit und Transparenz hinsichtlich einer Reihe von Phänomenen, die zum vagen Schnittstellenbereich zwischen Lexikon und eigentlicher Syntax gehören. Durch das in der Grammatik definierte morphosyntaktische Modul unterscheidet sich das Bulgarische von den anderen slawischen Sprachen. Das veranschaulicht augenfällig, daß in HPSG die Parametrisierung der sprachlichen und sprachübergreifenden Varianz in der Grammatik stattfinden kann. In diesem Sinne wäre es angemessen zu erforschen, ob sich ein morphosyntaktisches Grammatikmodul auch in der Beschreibung anderer

Sprachen rechtfertigen läßt, die für den lexikalistischen HPSG-Ansatz problematische Phänomene enthalten.

Als Voraussetzung für die Diskussion über die Rolle der Replizierung durch Klitika auf Satzebene wurde eine Typologie bulgarischer Nomialphrasen entworfen, die allgemeine Kriterien für die Bestimmung des Replizierungspotentials des Nominalmaterials in dieser Sprache zurechtlegt. So kann die folgende Grundannahme formuliert werden: replizierbar durch ein Pronominalklitikum (unter den ensprechenden verblexem-spezifischen oder kommunikativen Bedingungen) ist nur das Nominalmaterial, das als identifizierende spezifische Beschreibung eines Objekts benutzt wird. Dagegen weisen die artikellosen Nomialphrasen, die kategorisierende oder nicht-spezifische Beschreibungen sind, sowie die Nomialphrasen mit einem Artikel, die aber als generische oder auch als nicht-spezifische Beschreibungen benutzt werden, kein Replizierungspotential auf.

Ein besonderer Beitrag dieser Dissertation ist die ins Detail gehende Begründung dafür, daß das Phänomen der klitischen Replizierung einen kommunikationsgesteuerten syntaktischen Aspekt aufweist, der als Faktor der Wortstellungsvarianz im bulgarischen Satz anzusehen ist. Ferner wird argumentiert, daß die Replizierung von vollständigen NP-Konstituenten durch Klitika zwei unterschiedliche Funktionen hat: nämlich, Identifikation des direkten Objekts durch Akkusativklitika und Thematisierung des durch Akkusativ- und Dativklitika replizierten Nominalmaterials. Es wird im einzelnen dargestellt, wie sich die lexem-spezifische Obliqueness-Anordnung grammatischer Relationen, die recht flexible oberflächliche Wortfolge und die unter Umständen vorhandene klitische Replizierung bei der Informationsstrukturierung des Satzes gegenseitig beeinflussen. Dabei wird besonders auf die Position des emphatischen Akzents geachtet. Im entworfenen Modell kann vorhergesagt werden, wann Replizierung durch Klitika ausgeschlossen, obligatorisch oder optional ist.

Eine wichtige Grundeigenschaft des hier dargestellten linguistischen Ansatzes ist seine Implementierbarkeit - vgl. die im Anhang beigefügte Information zu einer HPSG-basierten Computergrammatik für das Bulgarische. Das bereits implementierte Fragment deckt die Morphosyntax von Verben vollständig ab, sowie einen beträchtlichen Umfang der Replizierungsphänomene auf Satzebene.

Curriculum Vitae

Tania Avgustinova studied Slavistics at the Faculty of Slavic Studies at St. Kliment Ohridski University of Sofia, and received her M.A. with a specialisation in computational linguistics in 1987.

Immediately after university she joined the Linguistic Modelling Laboratory of Bulgarian Academy of Sciences. From 1987 to 1989 she was involved in a large project on Bulgarian and Russian computational morphology as well as in computer-oriented lexicographic work on Bulgarian and Russian. From 1989 to 1991 she was mainly concerned with grammar writing and implementation for a syntactic parser of Bulgarian. In 1991-1992 she co-operated in a Bulgarian-German word order project sponsored by the German science foundation (DFG). During her time at the Linguistic Modelling Laboratory, she also taught courses at the University of Sofia.

In 1993 Tania Avgustinova started work as a researcher at the Department of Computational Linguistics and Phonetics at the University of Saarland in Saarbrücken, being involved both in research projects and teaching. She worked on special issues of computational grammars of Bulgarian and Czech in the framework of an EU-funded project on language processing technologies for Slavic languages (LATESLAV). Her research was also concerned with the applicability of the Head-driven Phrase Structure Grammar (HPSG) to Slavic languages.

In 1997 she completed her Ph.D. in Slavic and computational linguistics at the University of Saarland. The outcome of her doctoral research is described in this book.

Посвящается моей маме

Contents

Preface

Even though the problem area addressed in this dissertation stems from Slavic linguistics, there is a growing awareness nowadays that clitics and related phenomena pose a challenge for theoretical and cognitive linguistics by questioning the assumptions on the autonomy of the subsystems of grammar.

Theoretical linguistics offers a formal multidimensional framework to represent the complex interaction of linguistic information, as well as to express universal principles and their language-specific parametrisation. It is often the case that phonological, morphological, syntactic, semantic and even pragmatic factors have to be considered in the treatment of clitics to provide the basis for adequate modelling of observable regularities. The interaction of various grammar components is then reflected in the ultimate word (or element) order of utterances.

While the interface between morphology and syntax is not clear in a number of languages, there is an emerging consensus on the existence of degrees of "compounding" and "looseness" as far as clitic incorporation into morphological forms is concerned: across the languages, clitics' behaviour varies from that of word affixes to the autonomy of independent syntactic forms; in this respect, the intermediate status of Bulgarian clitics is indeed interesting.

The general methodological approach to be applied in this study is prevalent in computational linguistics and cognitive science. There is presupposed modularity in the organisation of the linguistic knowledge. The interaction between different modules provides the basis, on the one hand for distinguishing such apparently "interface" areas as Bulgarian verbal morphosyntax, and on the other hand for modelling the remarkable "multidimensionality" of the communicative (informational) organisation of the sentence. The language-specific assumptions and general linguistic hypotheses are verified by testing their plausibility in real computer implementation.

Acknowledgements

I would like to thank the people who influenced my development as a researcher by supporting me in various ways and encouraging my work on this PhD project.

I owe gratitude to Georgi Gargov and Elena Paskaleva, who played a key role in determining my professional profile, and to my colleagues from the Linguistic Modelling Laboratory at the Bulgarian Academy of Sciences.

I am very grateful to my supervisors Roland Marti and Hans Uszkoreit, who worked through several versions of this thesis, and provided fruitful criticism, discussions and support.

I am deeply indebted to Hans Uszkoreit for the excellent conditions in which I carried out my PhD research.

Special thanks to Karel Oliva for his interest in my work, and for helpful comments on a draft of this thesis. He taught me a lot about parsing, and worked out the formalism underlying the computer implementation of the Bulgarian grammar.

I am grateful to Ruselina Nicolova for her detailed comments and helpful suggestions.

I have always appreciated and immensely benefited from the remarkable scientific atmosphere at the Department of Computational Linguistics and Phonetics in Saarbrücken. Thanks to all current and former colleagues, and in particular to Bistra Andreeva, Sergio Balari, William Barry, Ralf Benzmüller, Thorsten Brants, Gregor Erbach, Claire Gardent, Martine Grice, Corinna Johanns, Tatjana Klajic, Jacques Koreman, Brigitte Krenn, Jie Li, Drew Moshier, Klaus Netter, Christina Neubacher, Stephan Oepen, Maike Paritong, Manfred Pinkal, Manfred Pützer, Helga Riedel, Christer Samuelsson, Elina Savino, Werner Saurer, Jutta Schäfer, Beatrice Schmitt, Wojciech Skut, Jürgen Trouvain, Claudia Villiger, Ryuko Woigardt.

On various occasions, I had inspiring discussions with Igor Boguslavsky, Eva Hajičová, Leonid Iomdin, Martin Kay, Jarmila Panevová, Jordan Penčev, Ivan Sag, Anna Sågvall-Hein.

Many thanks to Chris May and Uta Waller for proof-reading.

This work had partial financial support from the Commission of the European Communities through the project LATESLAV (Language Processing Technologies for Slavic Languages—PECO 2824), and from the University of Saarland (FR 8.7 Computerlinguistik).

Saarbrücken, April 1997

Chapter 1

Introduction

This study is concerned with Bulgarian word-order phenomena involving clitics. It grew out of an interest in the way considerable word-order variance is achieved in a language exhibiting an impoverished declension system (in particular, lacking case declension with nouns, adjectives and numerals) in combination with a well-developed mechanism for clitic replication. The impressively rich conjugation system and the distribution of verbal clitics introduce challenging evidence in favour of distinguishing morphosyntactic and syntactic constituency in the organisation of Bulgarian sentences. Moreover, the factors determining the ultimate alignment of the verb-form components and of the verbal clitics on the level of morphosyntax appeared to be quite different from those influencing the linear order of the major constituents within the clause, i.e. on the level of syntax proper. Thus, two interrelated topics attracted the focus of my research:

- the morphosyntax of the analytic verb forms and the positioning of verbal clitics within the verb complex;
- the phenomenon of clitic replication in its syntactic, communicative and word-order aspects.

The linguistic research carried out in this thesis is strongly motivated by the need for an explicit formalisation of Bulgarian constituent structure and word order regularities for the purposes of computer implementation. This has put rigorous requirements on the language description. A proposed theory is to be applied strictly to linguistic material "with no attempt to avoid unacceptable conclusions by *ad hoc* adjustments or loose formulations" (*Chomsky 1957*). The implementation itself is an important aspect of the whole enterprise because it provides the best test-bed for the verification of linguistic hypotheses.

Up to now, there has been no systematic investigation of the order of elements in Bulgarian analytic verb forms. The positioning of verbal clitics

with respect to verb forms of higher complexity has not received due attention either. Therefore, one of the goals of this thesis is to give a detailed description of the structure and the word order properties of the Bulgarian verb complex. Verbal clitics are treated as part of verbal morphosyntax, which allows for adequate modelling of their commonly acknowledged "two-faced" behaviour, namely, that they can neither be satisfactorily considered verbal affixes nor do they function as full-fledged nominal constituents. In *Nicolova 1986* (p.54) short pronominal forms are assumed to be "strongly dependent" and "non-autonomous"—both syntactically and communicatively—sentential parts because of their close connection with the verb. In my model, this intuition is reflected in a natural way by treating the short pronominal forms as morphosyntactic—rather than syntactic—constituents. The decision to view Bulgarian cliticisation as a part of the verbal morphosyntax has the positive consequence that the problem with two-fold satisfaction of a single valence requirement in the case of cooccurrence of a pronominal clitic and a coreferential full-fledged NP-complement is principally avoided, due to the fact that clitics are not legitimate constituents on the sentential level.

At the same time, and precisely on the sentential level, the mechanism of clitic replication plays an important role in determining the relative order of the verb (complex) and its arguments, in particular the subject and the replicable objects. However, no systematic account of this phenomenon has been made as yet, even though certain aspects of clitic replication—e.g., its origin, its place in the Slavic and in the Balkan language context, its appurtenance to literary language and its distribution in dialects, its case-indicating function in language development towards analyticity, its thematicising effect, etc.—have received explicit attention (cf. the references in Section 4.3). In the proposed treatment of Bulgarian cliticisation, I clearly distinguish constructions with clitic replicants from those where the clitic pronoun is an obligatory lexical formant indicating the experiencer. The developed model of clitic replication in the simple Bulgarian declarative sentence takes into consideration the remarkable constituent order variation in its relationship with the communicative organisation of utterances.

All this is supported formally by structural differentiation of two verbal head domains that are autonomous with respect to the operation of linear precedence (or linearisation) constraints: the morphosyntactic domain of the verb complex, where grammatical as well as prosodic factors influence the element order, and the syntactic domain of the clause, where

communicative factors have a dominant relevance for the ultimate alignment.

1.1 Language Investigated

Due to the difficulty of postulating norms with respect to word order variance and clitic replication, no related codification in Bulgarian normative grammar can be found. In reality, one can only speak of (strongly) dominating tendencies observable in the language use. What is described in this thesis is, basically, what educated native speakers would accept as norm, as language usage complying with the grammar of modern standard Bulgarian.

Some of the phenomena addressed in this study are attested quite differently in the large variety—social, dialectal, stylistic, temporal (i.e. with respect to the maturity of speakers), etc.—of what is considered modern Bulgarian language. The poetic style, for example, remains totally out of the scope of this thesis. The requirements of rhyme and rhythm dominating there often result in a so-called "word order inversion". To a certain extent, the latter can even affect the positioning of clitics, which otherwise is strongly determined by the grammar.

To create an exhaustive picture of the modelled word order and replication phenomena, all possible combinations and permutations have to be taken into account. Some variants of the investigated constructions are used quite rarely, others require a very special context in order to be uttered. It is obvious, therefore, that even a very large text corpus would hardly contain a representative part—let alone all—of the relevant data. This expectation was basically confirmed during my preparatory studies of journal and newspaper articles, modern Bulgarian prose, live moderation of radio and TV programs, etc. So, the majority of the language examples and the contexts given to license the various alignments in Chapter 4 are constructed using mainly theoretical sources and introspective data in combination with extensive testing of native speakers.

1.2 On Clitic Typology

Referring back to Wackernagel's original understanding of the term "clitic" (*Wackernagel 1892*)[1], and following the substantial body of recent literature[2], clitics can be defined as weak, prosodically dependent forms, typically, though not necessarily, accentless. *Zwicky 1977* introduces the distinction between simple and special clitics. A simple clitic is an element of some basic word class, which appears in a position in which the rules of syntax would (or at least could) put it, i.e. it occupies the normal syntactic position for a non-clitic word of its category. Special clitics are items whose position within some phrasal unit is determined by principles other than or additional to those of non-clitic syntax, i.e. they either occupy positions which one would not expect based on the distribution of other words or phrases with a similar function, or they obey restrictions which are not imposed on other words of a similar class or function. Hence, some special principles and mechanisms must be invoked to get the correct order of these elements with respect to the rest of their syntactic domain.[3]

A contrastive look at two Slavic languages—Czech and Bulgarian—is particularly insightful with respect to the clitic typology. Czech is a classical Wackernagel-position language: certain set of predicative clitics always occurs in the "second" position within the clause, immediately following some ensemble of lexical material which prototypically (but not necessarily—as discussed in *Avgustinova and Oliva 1995a*, *Avgustinova and Oliva 1995b*) can be interpreted as a single syntactic constituent. In Bulgarian, the predicative clitics of the same set are verbal clitics with a very special property: depending on the prosodic context, they can be proclitic or enclitic with respect to the verbal element hosting them. Based on this parallel, two further types of clitics can generally be distinguished. The first can be regarded as autonomous clitics, reflecting thus the intuition that they are inherently related to a certain kind of "second" position rather than to a prosodic host of a particular (morphosyntactic) category. The second type of clitics can, in turn, be viewed as host-category-bound clitics, indicating thus that they belong to the syntactic domain of a particular (morphosyntactic) category which also hosts them prosodically. Such are the clitics to the verb (complex) and the clitics to the noun (phrase) in Bulgarian.

---

[1]Cf. also *Anderson 1993* for a discussion.
[2]Cf., e.g., *Joseph et al. 1994* for a comprehensive bibliography.
[3]Cf. *Anderson 1992*, *Anderson 1993*, *Halpern 1992*, *Halpern 1995*, *Miller 1992*, etc.

Clitics to the NP are those clitic elements whose syntactic domain is the Bulgarian nominal constituent. All prosodically weak prepositions are nominal proclitics which are to be regarded as simple clitics in the terminology of *Zwicky 1977*. The status of a simple clitic can be legitimately claimed also for the enclitic interrogative particle *li* when it specifies non-verbal categories, since it regularly follows the entire constituent which is in the scope of interrogation. On the other hand, Bulgarian "short" possessive pronouns are to be regarded, in the terminology of *Zwicky 1977*, as special clitics. They can occur only with definite NPs, whereby the definiteness is either inherent (as in kinship terms) or expressed by the definite-article morpheme. The resulting morphonological ensemble:

[definite_article_morpheme + possessive_enclitic]

can then be viewed as a nominal enclitic cluster whose positioning in the syntactic domain of the respective nominal constituent is governed by a special mechanism that is similar to the Marking Feature Principle proposed by *Halpern 1995*. As to the negation particle *ne*, this is a stress-bearing (non-clitic) item whenever it specifies (i.e. scopes over) a non-verbal constituent.

I shall not pursue the issue of nominal clitics in this thesis. My main concern will be Bulgarian verbal clitics.

1.3 On Word Order Types

The communicative organisation of the sentence is crucial in licensing the concrete surface orderings in Slavic languages. It is determined by the needs of the actual information a sentence is supposed to transmit in the particular situation in which it is uttered. In any given utterance, it can be distinguished between an informative part (the rheme or focus) and an anchoring or vehicular part (the theme, also called ground or topic). An alignment where the thematic elements linearly precede the rhematic ones is the unmarked case for declarative utterances in Bulgarian, while the unmarked case for Bulgarian interrogative, imperative and exclamative utterances is an alignment where the rhematic elements linearly precede the thematic ones.

In this thesis, I investigate how word order "freedom" in Bulgarian declarative clauses interacts with the mechanism of clitic replication. For this purpose, I adopt a hierarchical communicative articulation of utterances which is based on three primitive notions employed in the

information packaging approach (cf. Section 4.2). The core distinction is made between focus and ground, with the latter divided further into link and tail. Using this terminology, I postulate four types of word order for the simple Bulgarian declarative clause. The *communicatively unmarked word order* presupposes either an all-focus interpretation of the respective utterance or a communicative articulation where the link part of the ground precedes the focus. The *parenthetic word order* is a special case of communicatively unmarked word order, inasmuch as it is also an instance of ground-preceding-focus articulation; the difference, however, is that the ground here contains not only a link part but an intonationally "parenthesised" tail part too. The *communicatively marked word order,* in turn, generally amounts to focus-preceding-ground articulation. The most typical instance thereof is a communicative articulation where the focus precedes the tail part of the ground. The *emphatic word order* can be viewed as a case of "split ground" where the focus is interlocated between the link part and the tail part of the ground; such a pattern presupposes local emphasis on the focal segment.

1.4 Formalisation

It is necessary to state at the very beginning that the formal issues have been moved to the second plan in this thesis, with the intention of making the analysis comprehensible for a broadest possible circle of readers with a background in Slavistics, i.e., for general linguists with a special interest (or specialising) in the Slavic language family—Bulgarian is a typologically important language in this respect, for comparative linguists interested in contrasting related phenomena across Slavic languages (e.g., clitics, word order variation, syntactic structures) or studying the Balkan language community, and, of course, for specialists in Bulgarian language. Taking such a position might obviously have the effect that this work would be of less direct interest for formally inclined computational and general linguists.

The lack of accent on formalisation, however, does not imply that the theory presented has inherent formal deficiencies or that it cannot be formalised. On the contrary, the fact that it has been successfully implemented in the form of a parser underlying an experimental grammar-checker for Bulgarian shows that a rigorous formalisation is indeed possible. Computational modelling plays a rather crucial role in the thesis. Implementing a theory in a formal system imposes more rigour on the

researcher, and gives an air of validity to the achieved results. Inconvenient data cannot be ignored or even falsified to make a theory seem applicable. So it is hardly surprising that nowadays computational linguistics gains recognition as a proving ground for language theories.

Although a large number of recent grammar models, formal implementations, and concrete grammars for specific languages considerably diverge from what can be considered "classical" Head-driven Phrase Structure Grammar (HPSG)—cf. *Pollard and Sag 1987*, *Pollard and Sag 1994*, their developers still classify them as HPSG-based. Adherence to the original could hardly serve as a criterion for such labelling, since even the originators of the theory divert from their earlier writings in their latest analyses. With its growing popularity, this framework is more and more understood as a "construction kit" for linguistic modelling rather than just some kind of linguistic theory. This view is promoted by Hans Uszkoreit in his ideas on HPSG and computational linguistics:

> If one understands the built-in modular design of HPSG, one also understands why the framework can serve as the basis for theoretical and computational research even if certain language-particular or language-universal linguistic theories and analyses might not be adopted or later changed. ... HPSG should rather be viewed as a construction kit for formal linguistic theories–not just grammars–that already comes with an initial set of serious theories and analyses. The modular design of HPSG offers a large degree of flexibility for further development of the overall framework, testing new variants of the grammar model, experimenting with different implementation strategies, adding different data types for grammatical description, applying the framework to new languages, changing individual components of a grammar. On the other hand, these multiple dimensions of freedom make it difficult to state what the essential properties of HPSG grammar models, implementations of HPSG formalisms, or concrete HPSG grammars for specific languages are. How many and which properties of the original HPSG must the model, formalism, or grammar possess to be called HPSG? It is obvious that there cannot be a formalisable answer to this question.[4]

In this study HPSG is taken as theoretical framework with its essential property of offering multidimensional, but nevertheless integral, sign-based representation of linguistic objects. The constrained-based approach to the interaction of various types of linguistic information (corresponding to different levels of language description—phonology, morphology, syntax, semantics, constituent structure, communicative organisation, etc.) within

[4]The text is published in 1996 on the World Wide Web at *http://www.coli.uni-sb.de/~hansu/hpsg.html*

a single data structure—the sign—allows for adequate modelling of phenomena related to word order and clitics.

One more word is due here on the status of the implementation in this thesis. Some points of deviation can be found between the theoretically postulated constructs and their particular implementation. For example, less internal structuring with respect to the feature architecture brings processing benefits and is, therefore, preferred in the implementation. There are also linguistically insignificant details in formulating the computerised grammar, as well as some application-oriented (i.e. grammar-checking-oriented) solutions. In general, however, the technical compromises made in implementing the ideas do not affect the overall layout of the proposed theory.

1.5 Outline

In the outline of the thesis, there is a theoretical part—covering chapters two to five, and an implementation part—attached as an Appendix. Chapter 2 offers an introduction to some basic notions of information-based linguistics, as well as to the main concepts of the Head-driven Phrase Structure Grammar (HPSG) which provides the theoretical framework for representation and modelling of the linguistic phenomena in this thesis. Also here, the special type of binary branching syntactic structures to be employed in the subsequent analysis is discussed, and the main assumption concerning the constituency in Bulgarian is represented. Chapter 3 investigates in detail the structure of the major sentential constituent in a Bulgarian clause—the verb complex—with special attention to the positioning of the verbal clitics. It is the structural domain of the verb complex that allows for adequate modelling of the morphosyntactic aspect of the replication phenomena. The communicatively-driven syntactic aspect of clitic replication and the closely related issue of constituent ordering on the clausal level are then the focus of Chapter 4. Inasmuch as the problem of nominal determinedness is inherently bound to the replication potential of nominal material, a typology of Bulgarian articled and non-articled NPs is proposed, which allows for clear specification of what nominal material is potentially replicable by a pronominal clitic. Two functions of clitic replication are shown to be relevant on the clausal level: direct-object identification and direct/indirect-object thematicisation. The related discussion is based on a classification of the various possible communicative structurings of different constituent-order variants of two

sample sentences. Finally, some aspects of formalisation in an HPSG-style grammar are considered. Chapter 5 offers general conclusions and a summary of the benefits of the approach. The Appendix contains an illustration of how the relevant linguistic knowledge is organised in multiple-inheritance hierarchies, as well as information on a fragment of computerised Bulgarian grammar which covers in full range the theoretical analysis developed in Chapter 3, and to a considerable extent the replication phenomena on the clausal level.

1.6 Reading Conventions

1.6.1 Transliteration of Bulgarian Cyrillic Alphabet

А а — A a	Е е — E e	К к — K k	П п — P p	Ф ф — F f	Щ щ —Št št
Б б — B b	Ж ж — Ž ž	Л л — L l	Р р — R r	Х х — X x	Ъ ъ — Â â
В в — V v	З з — Z z	М м — M m	С с — S s	Ц ц — C c	Ь ь — '
Г г — G g	И и — I i	Н н — N n	Т т — T t	Ч ч — Č č	Ю ю —Ju ju
Д д — D d	Й й — J j	О о — O o	У у — U u	Ш ш — Š š	Я я —Ja ja

1.6.2 Abbreviations

1, 1st	first person
2, 2nd	second person
3, 3rd	third person
A	adjective
ACC	accusative pronominal clitic
act-pcp	active participle
AdvP	adverbial phrase
aor	aorist
aor-aux	aorist of an auxiliary
aor-pcp	past participle active from the aorist stem
AP	adjectival phrase
art	article
AUX	auxiliary verb

aux-fin	finite auxiliary verb
aux-pcp	participle of an auxiliary verb
AuxC	auxiliary complex
AVM	Attribute-Value Matrix
B	Bulgarian
be-AuxC	auxiliary complex headed by *sâm* or *bâda* auxiliary
BE1	non-clitic forms of the *sâm*-auxiliary
BE2	*bâda*-auxiliary
C	Catalan
CAT	feature "category"
CAT_TYPE	feature "type of category"
CL-AUX	clitic present-tense forms of the *sâm*-auxiliary
CMWO	Communicatively Marked Word Order
cond	conditional mood
CONJ-PRT	the conjunctive particle *da*
CONST_ORD	feature "constituent order"
CUWO	Communicatively Unmarked Word Order
DAT	dative pronominal clitic
def	definiteness
def.art	definite-article morpheme
E	English
exp-acc	accusative formant indicating the experiencer
exp-dat	dative formant indicating the experiencer
EXP_OBJ	experiencer object
EWO	Emphatic Word Order
f	focus
F	prescript marking a focal segment
fem	feminine gender
FIN-AUX	finite auxiliary verb
FIN-MAIN	finite main verb
fut	future
FUT-AUX	positive future auxiliary verb
fut-AuxC	auxiliary complex headed by the positive or the negative future auxiliary verb
FUT-PRT	future-tense particle
GPSG	Generalised Phrase Structure Grammar
grd	present gerund active (adverbial deverbative)
HPSG	Head-driven Phrase Structure Grammar
imp	imperfect
imp-aux	imperfect of an auxiliary verb
imp-pcp	past participle active from the imperfect stem
imper	imperative mood
impers	impersonal
ind	indicative mood
INDEF-ART	indefinite article
INFO-STRUCT	feature "information structure"
intrans	intransitive verb

l	link
LEX_FORMANT	feature "lexical formant"
LFG	Lexical Functional Grammar
lim	limitedness
LP	Linear Precedence
masc	masculine gender
med	medial verb
med-se	medial verb with the reflexive formant *se*
med-si	medial verb with the reflexive formant *si*
N	noun
na-NP	noun phrase marked by the preposition *na*
NP	noun phrase
neg-fut, NEG-FUT-AUX	negative future auxiliary verb
neg-imper, NEG-IMPER	negative imperative auxiliary verb
NEG-PRT	negative particle
neut	neuter gender
OBJ, OBJ1	direct object
OBJ2	indirect object
OWO	Objective Word Order
pass-pcp	past participle passive
pcp	participle
PCP-AUX	participle of auxiliary verb
PCP-MAIN	participle of main verb
pers	personal clitic pronoun
ph-exp	phrasal experiencer verb
pl	plural number
pos-fut	positive future auxiliary verb
PP	prepositional phrase
pres	present
pres-aux	present of an auxiliary verb
pres-pcp	present participle active
PRT	particle
Q	interrogative particle
refl	reflexive
REFL-POSS-CL	reflexive possessive clitic
REL_OBL	feature "relative obliqueness"
RNM	replicated nominal material
S-V-O	subject - verb - direct object
S-V-O1-O2	subject - verb - direct object - indirect object
sg	singular number
strict-intrans	strictly intransitive verb
strict-trans	strictly transitive verb
SUBJ	subject
SWO	Subjective Word Order
SYNSEM	feature "syntax and semantics"
t	tail
trans	transitive verb

TYPE_OF_REPL	feature "type of replication"
V	verb
v-noun	verbal noun (nominal deverbative)
VC	verb complex
VP	verb phrase (clausal level)
WH-QUE	constituent containing an interrogative wh-word

Chapter 2

Theoretical Framework

The purpose of this chapter is to present the general features of the adopted framework as well as some tools and assumptions that are central to the analysis of Bulgarian word order and phrase structure in Chapters 3 and 4.

2.1 An HPSG-Based Formalism

The *Head-Driven Phrase Structure Grammar* (henceforth HPSG) originally developed in *Pollard and Sag 1987* and *Pollard and Sag 1994* is one of the major representatives of the approach to the study of natural languages in which the notion of *information* plays a central role. Assuming that a natural language can be dealt with in terms of the information that it makes available to the members of a certain community, information-based linguistics is concerned with the development of hypotheses of how this information is structured so as to make communication possible. Consequently, linguistic objects are widely considered as bearers of information. In spite of the unsolved problem concerning the nature of linguistic information in general, the possibility of information-based linguistics is advocated by Pollard and Sag in the following way:

> ... although we aren't sure what linguistic information is, we can still theorise about it by trying to say what it is like. To be more precise, we can try to construct formal *models* that reflect certain interesting aspects of the things we are studying. *Pollard and Sag 1987* (p.27)

As the name suggests, the notion of *the head constituent of a phrase* is of central importance to the theory. HPSG falls within the general framework of *unification grammar*[5] and has its roots in several contemporary

[5]Unification grammars describe language in terms of static constraints on information, as opposed to the dynamic transformation of the expressions themselves in the earlier

approaches to the study of natural languages like Generalised Phrase Structure Grammar—*Gazdar et al. 1985*, Lexical Functional Grammar—*Bresnan 1982*, Government and Binding Theory—*Chomsky 1981a*, Functional Unification Grammar—*Kay 1985*, Categorial Grammar—*Karttunen 1986*, *Uszkoreit 1986a*, *Zeevat et al. 1987*, etc., thus integrating and further developing ideas from a wide range of research traditions. In HPSG, the conventional wisdom is accepted that linguistic theory must account for linguistic knowledge (a recursively definable system of linguistic types) but not necessarily for processes by which that knowledge is brought to bear in the case of individual linguistic tokens. In some basic methodological assumptions, the authors explicitly identify their position with the one advocated by Chomsky as early as in *Chomsky 1957*:

> Precisely constructed models for linguistic structure can play an important role, both negative and positive, in the process of discovery itself. By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalised theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory have failed to recognise the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by *ad hoc* adjustments or loose formulations.

Like most modern syntactic theories, HPSG conceives of linguistic research as the task of defining the grammatical principles that determine the well-formedness of linguistic objects; in this sense its conception of grammar as a set of parametrised principles is very close to the one expressed by Chomsky in several works—*Chomsky 1981b*, *Chomsky 1981a*, *Chomsky 1986*. What is quite different is its conception of linguistic objects and of the way information (phonological, syntactic, semantic, etc.) is organised in these objects. Thus, while in transformational models linguistic objects are represented by means of a set of phrase-structure trees related to each other by movement rule *Move* α, and where each element in this set corresponds (roughly) to a different information level, in HPSG, levels of

---

transformational descriptions. This means that unification-based descriptions of language require that the application of all linguistic constraints be monotonic, that is, the constraints merely add information, without performing structural changes—*Sag et al. 1986*.

information are parallel and coexist in the same representation without resorting to a rule responsible for deriving one from another. Quite contrary, a key architectural property of HPSG is the absence of any notion of transformation, which means that it is a non-derivational linguistic theory.

Let us consider the issue from a more general perspective. In an approach to linguistic theory that may be described as information-based rather than derivation-based, linguistic generalisations are formulated entirely in terms of constraints on *partial* information structures, to the exclusion of operations that transform one fully specified structure into another (or derive one from another). Partiality of information is used in many linguistic analyses to eliminate the need to specify a set of alternatives repeatedly (*Sag et al. 1986*). Apart from processing considerations that will not be discussed here, information-based characterisations of linguistic knowledge show considerable promise for making sense of certain basic facts about language use that are too often ignored in theoretical discussions, as pointed out in *Pollard and Sag 1988*. All such theories are compatible with the assumption that in actual language-use situations, whatever the processing task at hand (e.g., interpretation, production, making judgements of grammaticality, translation, etc.), the specification of a token linguistic object comes about in a cumulative (or *monotonic*) fashion via the interaction of constraints arising from several sources. These sources include **lexical entries** (which contain phonological, morphological, syntactic and semantic information), the **grammar rules** that combine them, language-specific and language universal **principles of well-formedness**, as well as the particular language-use situation itself, i.e. **contextual factors**. The final result of a language process is obtained by *unifying* the information from all of these various sources. Yet, information-based theories are purely *declarative* (as opposed to procedural) in the sense that they characterise ***what*** constraints are brought to bear in language use ***independently of the order*** in which the constraints are applied. In addition, the theories themselves are unbiased as to the kind of processing task at hand; more specifically, they are reversible in the sense that they are neutral with respect to interpretation (parsing) vs. production (generation). As argued in *Sag et al. 1986*, there seems to be an emerging consensus in modern linguistics that explanatory accounts of syntactic phenomena can be provided in a monotonic system of equality constraints over partial information structures. Also for computational reasons monotonic systems have some advantages over derivational ones—the existence of unification as a simple, order-independent, computationally precise method for solving any system of

equations of the sort used in unification grammars allows for building very general interpreters for unification grammars that can be used to test analyses in many of the different unification-based theories.

To a growing community of linguists nowadays, the most appealing feature of HPSG is that this framework provides a very powerful formal basis for modelling linguistic knowledge. HPSG has not committed itself to strong pragmatic issues apart from trying to pursue what kind of information is represented in linguistic objects, how the different kinds of information interact with each other, and how information is represented and structured in linguistic objects. This makes HPSG a rather eclectic and open-minded theory which offers the linguist considerable freedom to borrow, synthesise and re-elaborate insights coming from other theories. However, HPSG is not totally unconstrained; its restrictiveness comes from its formal architecture—a methodological strategy advocated already within GPSG. The principle of *ontological parsimony* is taken seriously in the design of the framework: insofar as possible, no constructs are posited that do not correspond to observables of the empirical domain—cf. *Pollard and Sag 1994* (p.7). Many of the central constructs of HPSG are motivated by its adherence to strict lexicalism, a thesis that entails that syntax cannot operate on or make reference to internal properties of lexical items. Any lexically based theory necessarily employs rich lexical representations and HPSG's universal grammar is a small set of principles that allow phrases to be projected from the particular information encoded in lexical entries. The imposed further condition of computability, according to *Pollard and Sag 1994* (p.8), is the theory's reflection of the facts that the structures of linguistic expressions are capable in principle of being computed by the resource-bounded information-processing organisms which successfully employ them in a communicative function, and that language users are able to render judgements as to the well-formedness of candidate expressions.

Fundamental objects of linguistic analysis in HPSG are *signs*—linguistic expressions, either words or phrases, or even something more extended like texts. The HPSG linguistic theory conceives of signs as structured complexes of phonological, syntactic, semantic, discourse, phrase-structural, and possibly other relevant information. In the HPSG formalism, signs are modelled by feature structures which can be described in terms of attribute-value matrices (AVMs)[6]. Intuitively, a feature structure is a description of some object, specifying some or all of the information that is asserted to be true of it. Formally, a feature structure is nothing more than a

[6]A comprehensive introduction to this formalism can be found in, e.g., *Shieber 1986*.

specification of a set of attributes, each of which is paired with a particular value, as illustrated in Figure 1 for the agreement features of the verb form *igraem* '(we) play'.

$$\begin{bmatrix} \text{AGREEMENT} \begin{bmatrix} \text{PERSON 1st} \\ \text{NUMBER pl} \end{bmatrix} \end{bmatrix}$$

Figure 1

The attributes PERSON and NUMBER have atomic values (1st and pl, respectively), while the attribute AGREEMENT has a feature structure as its value (represented as an AVM).

Modelling signs by feature structures allows for parallel representation of various dimensions, e.g., Figure 2 illustrates a possible structure of a lexical sign (a word) which has some phonological and morphological properties as well as syntactic and semantic characteristics.[7] It is further shown how syntactic constituency is integrated into the structure of the phrasal sign by means of just another attribute (DAUGHTERS) whose value is a list of signs.

$$\textit{lexical sign} \begin{bmatrix} \text{PHONOLOGY} \cdots \\ \text{MORPHOLOGY} \cdots \\ \text{SYNTAX} \cdots \\ \text{SEMANTICS} \cdots \\ \vdots \end{bmatrix} \qquad \textit{phrasal sign} \begin{bmatrix} \text{PHONOLOGY} \cdots \\ \text{SYNTAX} \cdots \\ \text{SEMANTICS} \cdots \\ \text{DAUGHTERS} \langle \text{list of } \textit{signs} \rangle \\ \vdots \end{bmatrix}$$

Figure 2

A core idea of sign-based frameworks is that all relevant linguistic aspects are represented in every linguistic unit, i.e. in words, phrases, clauses, sentences, etc., and may hence interact. In this respect HPSG differs from most grammatical frameworks that have grown out of transformational (as well as dependency) tradition. In these grammars, the different linguistic aspects of a sentence are typically factored out into different levels of representation, and certain operations are employed for relating one level to another. This makes modelling of phenomena which involve different aspects of language description (e.g., interplay of phonology and syntax, or constituent order and discourse organisation, etc.) apparently problematic.

[7]The latter two are represented jointly by the attribute SYNSEM in HPSG, which attribute is assumed to include a complex of linguistic information that has the potential of being selected (subcategorised for) by other signs and that is shared by the complement subject and the controller in HPSG theory of raising.

In this respect HPSG can be viewed as a monostratal but multidimensional linguistic framework. The use of feature structures provides the grammar writer with many different ways of organising the linguistic information and the way its pieces interact. The simplest case of such interaction is the case of *identity* of information contained in different parts of a sign, expressed formally as *structure sharing* of the values of the relevant features within this sign. In AVM-notation, this is indicated by *coreference tags* occurring as values of the respective attributes, which encodes the fact that these attributes share the same structure as their common value. This structure-sharing, meaning token identity, is one of the central explanatory mechanisms in HPSG—cf. in Figure 3 the representation of the so-called Head Feature Principle postulating that the HEAD value of a headed phrase is identical to that of the head daughter. Here the effect of such a general requirement is to guarantee that headed phrases really are "projections" of their head daughters.

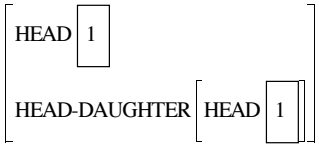


Figure 3

Token identity of values of two features is the simplest example of the general notion of a *constraint* on feature structures. More complicated cases will be exemplified below, but let us first explain the relation between *token identity* and *unification* of two values. These two terms actually represent two different views of one and the same concept—namely, the identity of information borne by two values. However, having introduced feature structures, which may be only partially specified, one needs a general method for allowing two compatible partial feature structures to amalgamate the information they contain. Two feature structures can unify only if they are consistent—that is, unless they specify conflicting types or different atomic values for the same feature. The unification of two feature structures is simply the feature structure obtained by combining all of the information from both of them, and nothing more. Hence, the fundamental operation of unification is just that of simple merging: either it amalgamates compatible partial information, as in Figure 4, or it fails if the two pieces of information are incompatible, which is illustrated in Figure 5—cf. also *Sag*

et al. 1986.[8] Intuitively, the two descriptions in the latter case cannot describe the same entity.

$$\begin{bmatrix}\text{AGREEMENT}\begin{bmatrix}\text{PERSON 1}\\ \text{NUMBER sg}\end{bmatrix}\end{bmatrix} \cup \begin{bmatrix}\text{AGREEMENT}\begin{bmatrix}\text{NUMBER sg}\\ \text{GENDER fem}\end{bmatrix}\end{bmatrix} = \begin{bmatrix}\text{AGREEMENT}\begin{bmatrix}\text{PERSON 1}\\ \text{NUMBER sg}\\ \text{GENDER fem}\end{bmatrix}\end{bmatrix}$$

Figure 4

$$\begin{bmatrix}\text{AGREEMENT}\begin{bmatrix}\text{PERSON 1}\\ \text{NUMBER pl}\end{bmatrix}\end{bmatrix} \cup \begin{bmatrix}\text{AGREEMENT}\begin{bmatrix}\text{NUMBER sg}\\ \text{GENDER fem}\end{bmatrix}\end{bmatrix} \neq \begin{bmatrix}\text{AGREEMENT}\begin{bmatrix}\text{PERSON 1}\\ \text{NUMBER ???}\\ \text{GENDER fem}\end{bmatrix}\end{bmatrix}$$

Figure 5

Viewed from the perspective of information content, two feature structures can be unified if and only if there exists a feature structure that they both subsume (a unifier).

Implicit in the use of feature structures is a commitment to developing a theory of what kinds of features go together and what values are appropriate to each particular attribute, that is, a commitment to specifying which feature structure categories are well-formed and which are not. Thus, much of the grammar development in HPSG is concerned with formulating a natural theory of linguistic generalisations in terms of the constraints that govern the feature structure categories, and hence the grammatical entities that they describe. One of the first things to do in developing a theory of grammar is to classify linguistic entities in various ways. To this end, it is particularly useful to introduce the notion of *type*. If we think of feature structures as providing descriptions of grammatical entities, then types are just a classification that allows us to develop a theory of the properties that hold of every type in the classification. Types will allow us, for instance, to specify which features are appropriate for which kinds of entities. Due to hierarchical organisation, a given feature structure contains only those features that are declared appropriate by one of its types, i.e. by its "basic" type or one of its supertypes. A feature structure also inherits any type constraints, i.e. constraints on feature

[8]However, the machinery of feature structures with constraints and unification imposes more tough structuring of the data than is sometimes needed. This is a formal issue (even though with considerable linguistic bearing) and a topic of on-going research, which will not be treated in this work.

values, that may be associated with its supertypes. The method used for combining (merging) feature structure descriptions is unification. In the *typed* feature structures that are employed in HPSG, the type actually indicates what sort of empirical object the structure is modelling, and there is one sort symbol for each basic type (ontological category) of construct. The (finite) set of all types is assumed to be partially ordered in a subsumption hierarchy, with more inclusive (less informative) types higher in the ordering, which means that they are supertypes of others which contain more information (i.e. are more informative). For example the types *phrase* and *word* are ordered below the type *sign* because signs include both phrases and words; thus it is said, e.g., that *word* is a subtype of *sign*, or that *acc* is a subtype of *case*. In general, the more specific (or explicit, or informative) a description, the less structures satisfy it.[9]

The hierarchical organisation of linguistic types in HPSG enables us to classify linguistic entities in more subtle ways which allow intermediate level categories of various sorts, so that it is possible to talk about the properties shared by two distinct types by associating a feature or a constraint with their common supertype. The use of cross-classification of types, multiple inheritance and (traditionally) lexical rules, or (alternatively) other—e.g., relational[10]—constraints, allows for a concise and highly structured representation of the lexical and other linguistic knowledge in the HPSG-based language descriptions. Along with the lexical cross-classification allowing for complex properties of words to be derived from the logic of the lexicon, there is also a hierarchical cross-classification of grammatical constructions that allows for generalisations about diverse construction types to be factored into various cross-cutting dimensions.

Summing up, with respect to this study, the specification "HPSG-based" generally means:

- use of a feature structure description formalism: under-specification, unification, constraints, etc.;
- sign-based approach: all levels of linguistic description are available in one integral representation;

---

[9] All feature structures are required to be *well-typed* (*Carpenter 1992*), i.e. their type restricts which attributes are appropriate for them and of what type the values of these attributes are. What attributes (or equivalently, features, or components of structure) an empirical object has depends on its ontological category. In a completely formalised HPSG grammar, it must be stated explicitly what the type symbols are, how the type symbols are ordered, and what the appropriate attribute labels and value types are for each type.

[10] Cf. *Oliva 1994*.

- principle-based approach: Head Feature Principle (general constraints on the sharing of properties between phrases and their heads), Valence Principle (general constraints on the realisation of the combinatoric potential of different elements), etc.;
- constraint-based grammar: rules, lexical entries and general principles all provide partial constraints on the well-formed structures of the language;
- various levels of abstraction: linguistic objects are organised in type hierarchies derived by abstraction;
- grammar rules as structural schemata specifying properties of phrases;
- lexicon with cross-classifications and multiple inheritance hierarchies providing a theory of word classes.

2.2 Binary Branching Syntactic Structures

Binary branching structures for syntactic description of languages allowing considerable constituent order variations have been proposed independently in several works—e.g., in *Uszkoreit 1986b* for the description of verb final clauses in German, in *Gunji 1987* for Japanese, in *Avgustinova and Oliva 1990* for (mainly) Slavic languages, in *Netter 1992*, *Netter 1994* again for the analysis of German, etc.

The overall shape of the syntactic structures employed in this study is illustrated in Figure 6. Motivation for adopting binary right-branching structures having a distinguished empty non-terminal as rightmost element of the branching, as well as a discussion of their adequacy for description of free word order languages, can be found in *Avgustinova and Oliva 1990* and *Oliva 1992a*. The main advantage of such a structural representation is that it potentially allows for free intermixing of heads, subjects, complements and adjuncts, which is a desired option in the description of order variation. Provided such a representation, however, no linear precedence constraints in their classical form of *Gazdar et al. 1985*, i.e. ordering sibling nodes in a local tree, can be formulated. The formal aspect of preserving—or rather re-introducing—locality in this particular type of binary branching structures is investigated in *Oliva 1992b* and *Engelkamp et al. 1992*.

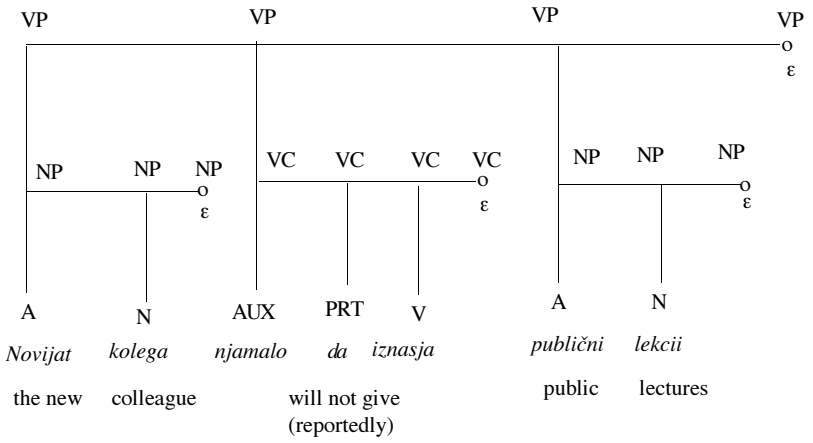


Figure 6

If we observe more closely the adopted structure (on a concrete example as in Figure 6), we can see that it can be viewed as divided into levels—contiguous sequences of nodes with identical marking and with a distinguished phonologically empty element at the end of each. This element is assumed to be the **functional head** of the phrase (or "level") it terminates. Moreover, each headed phrasal level has one central element (in a non-final position) called the **nucleus** (or nucleus head) in the approach adopted in this study—e.g., the V (the main verb) for the VC (the verb complex), and the latter for the VP (the verb phrase, or the clause), as well as the N (the noun) for the NP (the nominal phrase). The final element has the category of the immediately dominating node and is in fact an empty projection of the respective nucleus. The structural levels, representing the respective (syntactic or morphosyntactic) constituents, are autonomous *head domains* with respect to word order. In particular, the linear precedence (LP) constraints relevant for the morphosyntactic complex[11] differ in nature from those applicable within the clause. This is one of the important generalisations captured by the adopted type of binary branching structures. The LP constraints are supposed to apply to non-sibling constituents within the respective level of the structure.

Let us mention in passing that considering the distinct structural levels as separate word order domains is compatible with the proposal of *Engelkamp et al. 1992*, where the LP constraints, encoded by LP relevant

[11]Cf. Chapter 3.

features, apply to non-sibling constituents "in linguistically motivated binary branching head domains". The authors assume that a head domain consists of the lexical head of a phrase and its complements and adjuncts, and that LP constraints must be respected within a head domain. From a formal point of view, the LP constraints are encoded in such a way that violation of an LP constraint results in unification failure, and LP constraints, which operate on head domains, can be enforced in local trees by checking sibling nodes. The last condition can be ensured if every node in a projection carries information about which constituents are contained in its head domain. Since in the adopted type of syntactic structure each local tree has the general form given in Figure 7 with the mother node (*constituent*) having the same category as the phrasal head daughter (*rest constituent*), it is actually required that the phrasal head daughter carries information about the constituents already expanded from the respective structural level, i.e. about the constituents overtly realised to the right of the currently expanded element in the respective head domain. The LP constraints rely on the local interaction of this information with the constituent that is currently being expanded.

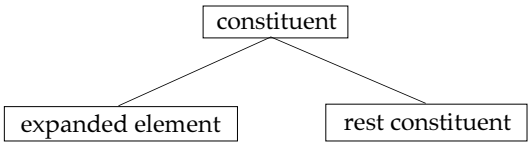


Figure 7

Encoding of LP relevant information by features occurring on the (phrasal) categories is obviously not the only possible solution. As argued in *Oliva 1992b*, order is a property of the syntactic structure (made up of categories) rather than of the categories themselves. An alternative approach put forth in this paper views the data types employed in binary-branching structures as lists, which allows for stating LP constraints within a hierarchy of lists types. The empirical question of what should be encoded by features and what by types remains open, though. Furthermore, using structural representation which is different from what is commonly assumed in HPSG certainly enforces a different form not only of the grammar rules (i.e. of the phrase structure schemata), but of the principles too. However, the intuitive linguistic background remains the same. *Oliva 1992b* gives a sample reformulation of the standard HPSG Head Feature Principle and Subcategorisation Principle which takes the peculiarities of the discussed binary branching structures into account.

2.3 Morphosyntactic vs. Syntactic Constituency

The distinction made in the HPSG hierarchy of linguistic objects between lexical and phrasal signs, i.e. between word and phrase, is a very intuitive one. The lexicon provides a theory of words, while grammar rules all specify the properties of phrases.

There are, however, phenomena based on syntagmatic relations of "intermediate" nature which—even in the strongly lexicalist perspective of HPSG—cannot satisfactorily be represented as exclusively lexical matter. Interpreting them as belonging to the area of syntax proper also seems unsatisfactory. Their adequate modelling presupposes a more radical step leading to a revision of the notion of constituency.

With the idea of syntagmatic relation remaining common to both morphosyntax and syntax, I shall distinguish between morphosyntactic and syntactic constituency. To this effect, I introduce an additional autonomous object, *morphosyntactic complex*, which intuitively is an intermediate construct between the word (understood as a synthetic word-form) and the (traditional) syntactic phrase. For Bulgarian, morphosyntactic constituency incarnates syntagmatic regularities associated with analytic (periphrastic) verb forms and verbal clitics. Such a morphosyntactic "dimension" in the hierarchy would, presumably, be distinctive also for other languages exhibiting a higher degree of "analyticity". At the same time, no such module can be claimed relevant for the majority of modern Slavic languages, which, in turn, is an adequate reflection of the fact that Bulgarian exhibits a set of properties and phenomena that are exotic in the Slavic language family, though fairly common in the Balkan languages.

Going so far as to admit the existence of morphosyntactic constituency provides us with increased explanatory potential especially in modelling phenomena related to word order and clitics in Bulgarian.

- The seemingly isomorphic mechanism of clitic replication gains in transparency if it is interpreted as having a morphosyntactic and a syntactic aspect.
- Different factors interplay in regulating the element order in morphosyntactic and syntactic formations, which is adequately modelled by distinguishing structurally the two respective word order domains.

- The operation of grammar principles, such as the HPSG Head Feature Principle and Valence Principle, in the morphosyntax differs from that in the syntax proper. Intuitively, head features have to be "collected" from various (possibly non-head) verb form components within the verb complex, and interpreted in a more elaborate way rather than just set identical with those of the head-daughter. Also intuitively, no "ultimate" satisfaction of valence requirements should take place on the level of morphosyntax. In particular, although certain valence requirements of a given verb can be modified by pronominal clitics with respect to the number, gender and person of the corresponding arguments, they can still be satisfied by full-fledged nominal constituents.

An HPSG-style type hierarchy for Bulgarian can, for instance, have a top-level partitioning of linguistic objects as illustrated in Figure 8, with *word*, *morphosyntactic complex* and *syntactic phrase* being subtypes of the type *sign*.

There are features associated with each type, i.e. the types are defined with the top-level attributes that are considered appropriate for them. For instance, the type *sign* is defined with the features PHONOLOGY and SYNSEM; for the type *word* a further feature MORPHOLOGY is appropriate; the type *morphosyntactic complex* introduces the features MS-DTRS (morphosyntactic daughters) encoding the morphosyntactic constituency and VCE (verb complex elements list)[12]; for the type *syntactic phrase*, the features DTRS (daughters) traditionally encoding the phrase structure and INFO-STRUCT (information structure, i.e. communicative organisation) are admitted. On this approach, principles of grammar are stated separately for morphosyntactic and syntactic objects.

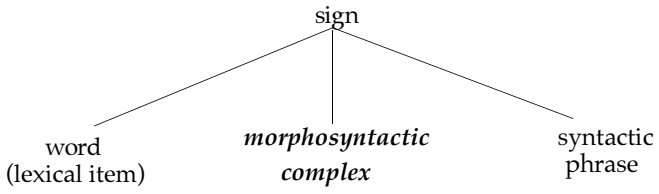


Figure 8

Alternatively, the morphosyntactic complex can be viewed as a subtype of *phrasal sign*, as illustrated in Figure 9, preserving thus the fairly standard top-most division of signs into lexical and phrasal.

[12]The constituent structure in morphosyntax is in the main focus of Section 3.3.

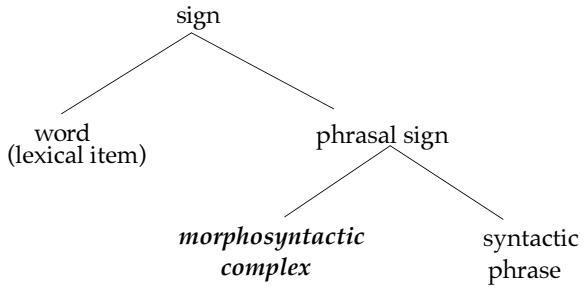


Figure 9

Such a solution would better reflect the fact that morphosyntactic and syntactic objects have an important property in common, namely, constituent structure, but the principles of grammar—e.g., Head Feature Principle, Valence Principle, etc.—must be parametrised accordingly.

In the end effect, a formal grammar of Bulgarian has to distinguish three types of objects in terms of constituents: lexical, morphosyntactic and syntactic. Lexical constituents are the words themselves as the lexicon (with an integrated morphological analysis) supplies them. Morphosyntactic constituents can be headed either by an autosemantic verb or by a functional auxiliary verb. In the former case I shall speak of a (morphosyntactic) verb complex (VC), and in the latter—of a (morphosyntactic) auxiliary complex (AuxC). A Bulgarian sentence—viewed in the present approach as a syntactic verbal constituent (or phrase, i.e. VP)—can be headed either by a lexical (autosemantic) verb or by a verb complex. Even though auxiliary verbs can form morphosyntactic constituents, they would never head a clause in Bulgarian, i.e. no auxiliary can form a syntactic phrase. In the proposed model this adequately reflects the purely functional nature of Bulgarian auxiliary verbs. At the same time, all major non-verbal categories—nouns, adjectives, adverbs, etc.—are assumed to form only syntactic constituents (or phrases, e.g., NP, AP, AdvP, etc.). In other words, the morphosyntactic level of description is considered irrelevant for Bulgarian non-verbal categories, inasmuch as these do not exhibit complex analytic behaviour.

The structural representation adopted in this study—cf. the previous section—requires some technical adjustments in the multiple-inheritance organisation of the type hierarchy. For the sake of an example, let us consider the variant in which words, morphosyntactic complexes and syntactic phrases are assumed to be subtypes of *sign*. These objects have to

be additionally classified with respect to their overt realisation into *phonologically nonempty* ("normal") and *phonologically empty* ("functional"). The former type is to be further divided into *head expansion* and *non-head expansion* depending on the nature of the expanded constituent, i.e. whether this is the nucleus or some type of modification. The particular phrase-structural schemata (corresponding to types of grammar rules) are then obtained as a result of a cross-classification illustrated in Figure 10.

The distinction between morphosyntactic and syntactic constituency is an important base for the linguistic analysis developed in Chapters 3 and 4, as well as for the computer implementation. As we shall see, the complexity of syntagmatic relations within the Bulgarian verb complex questions the adequacy and universal validity of lexicalised HPSG approaches to the treatment of clitics which have been proposed for mainly Romance languages (e.g., *Miller and Sag 1995*, *Monachesi 1995*).

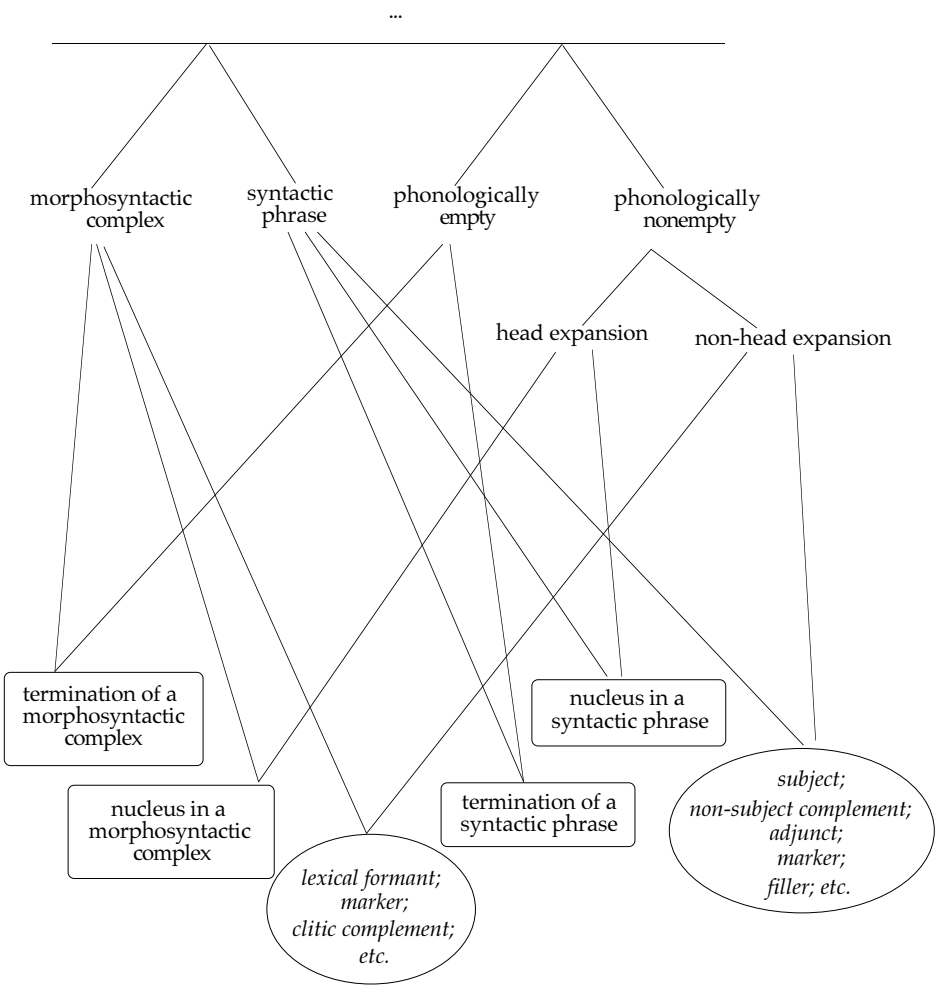

...
morphosyntactic complex
syntactic phrase
phonologically empty
phonologically nonempty
head expansion
non-head expansion
termination of a morphosyntactic complex
nucleus in a morphosyntactic complex
lexical formant; marker; clitic complement; etc.
termination of a syntactic phrase
nucleus in a syntactic phrase
subject; non-subject complement; adjunct; marker; filler; etc.

Figure 10

Chapter 3

Verb Complex and Verbal Clitics

A morphosyntactic object headed by a verbal category is called **verb complex** in this study. The verb complex is a major sentential part, and thus an immediate constituent of the Bulgarian clause. It covers the verb form, simple or compound, as well as reflexive particles, short personal or reflexive pronominal forms[13] and the negative and the interrogative particles. If, however, there is only the main verb alone occurring in the sentence (e.g., with no auxiliary or clitic components), no verb complex is assumed.

3.1 Verb Forms and Morphosyntactic Marking

A statement made once very early in the morning by my daughter when she was one year old:
— *Аз съм се наспах.*

A dialogue with my daughter when she was two years old:
— *А тебе страх ли те е от Баба Яга?*
— *Тя не съществува, ма Саше.*
— *Ако ще беше съществувала, щеше ли да те е страх?*

3.1.1 Basic Verbal Categories

In comparison to Old Bulgarian (Old Church Slavonic)[14], Modern Bulgarian is commonly considered to have developed towards an analytic type of language, since it is primarily the loss of nominal inflection, not verb inflection, that defines the development from a synthetic system to an

[13]The remarkable differences in the distribution of full and short personal pronouns in Bulgarian makes certain authors assume that the short (clitic) forms are not complements but rather verbal affixes—cf. *Walter 1965*. In my opinion, they are more adequately treated as belonging to the morphosyntax of the verb complex.
[14]Cf., e.g., *Mirčev 1972*.

analytic one. Bulgarian is exotic in the Slavic family inasmuch as it exhibits a fairly rich conjugation system and simultaneously a rather impoverished declension system, all this in combination with a considerable degree of what is usually called "word order freedom". While Bulgarian nouns, adjectives, and numerals have lost their case declension, the verbal conjugation contains more forms than that of any other living Slavic language. In comparison with the archaic Slavic system—as attested by Old Bulgarian (Old Church Slavonic)—Bulgarian has lost the dual forms and only some non-finite verbal forms, most notably the infinitive. An entirely new grammatical category called *preizkazvane* (reporting or renarrating) has enriched the paradigm[15]. Further forms—e.g. conclusive, admirative, etc.—are in a process of establishing themselves in the verbal system of modern Bulgarian. As in all Slavic languages, verbs in Bulgarian have grammatical aspect, which can be either perfective or imperfective, with some verbs being biaspectual. The systematic process of *secondary imperfectivisation* (derivation of imperfective forms from perfective ones via suffixation) in Bulgarian leads to the existence of aspectual pairs with the same lexical meaning and morphosyntactic features. In this case, the perfective and the respective secondary (derived) imperfective forms are in purely grammatical opposition, quite unlike the situation with aspectual pairs derived via prefixation. The latter consist of a primary (non-derived) imperfective verb and a perfective counterpart derived from it with a prefix or a suffix *-n-*. Here slight changes of the lexical meaning can be observed mainly due to the prefixation involved.

Not all forms of the verbal paradigm can be formed from both the perfective and the imperfective stem. The following require an imperfective stem *Maslov 1963*: negative imperative, present participle active, gerund, verbal noun. Notice that this does not necessarily mean that all of these forms are semantically imperfective. At least the verbal noun semantically represents both aspects: *razkâsvane* 'tearing' corresponds to the Polish *rozerwanie* (perfective) and *rozrywanie* (imperfective)—see *Andrejčin 1938*. Since both aspects can appear in all tenses, the **aspectual** opposition *perfective* vs. *imperfective* must be distinguished from the **temporal** opposition *aorist* ("past simple") vs. *imperfect* ("present in the past"). Also *perfect* as a **tense** must be distinguished from the *perfective* **aspect**. In the

[15]Although it is traditionally dubbed "mood", in fact it differs in nature from the prototypical moods, such as indicative, imperative, and conditional. There are reported (or renarrative) forms not only in the indicative but also in the imperative mood. Therefore, this category must be granted an independent status in the verbal system, which is reflected in my analysis by introducing the distinction between a **reported** and a **non-reported mode**.

chapter on the problems of verbal aspect, *Lindstedt 1985* (p.40) summarises certain syntactic restrictions on the use of aspect forms. The syntactic environments where the imperfective would be ungrammatical are much more specific than those excluding the perfective. The most important restriction which seems to hold in all Slavic languages is the ungrammaticality of a perfective form in the complement of a phasal verb, as in (1). Phasal verbs denote the beginning, continuation, or completion of the situation or activity expressed by their verbal complement.

(1)
Započvam / prodâlžavam / spiram da stavam $_{\text{imperfective}}$ */*stana* $_{\text{perfective}}$ *rano.*
'I begin / continue / stop to get up early.'.

Further, perfective verbs cannot be modified by frequency adverbials like 'often', 'seldom', 'sometimes', etc., as (2) illustrates.

(2)
Te se viždaxa $_{\text{imperfective}}$ */*vidjaxa* $_{\text{perfective}}$ *često / rjadko / ponjakoga.*
'They met each other often / seldom / sometimes.'

The latter restriction must always be observed in Russian and Bulgarian, although it does not necessarily hold in all Slavic languages—cf. *Ivančev 1971*.

An interesting case of "aspect agreement" can be observed with the Bulgarian perfective verbs *uspeja* and *udade mi se* (both meaning 'to succeed in'): only a perfective verb form can be used in the complement of these verbs, as (3) shows.

(3)
Uspjax $_{\text{perfective}}$ *da kupja* $_{\text{perfective}}$ */*kupuvam* $_{\text{imperfective}}$ *presen xljab.*
'I succeeded in buying fresh bread.'

Lindstedt 1985 (p.41) observes that *uspeja* (perfective) and similar verbs differ from phasal verbs in that the latter determine the aspect of their complements irrespective of their own aspect (cf. also (4)), so he regards the case of phasal verbs as "aspect government" rather than agreement. [16]

(4)
Uspjavax $_{\text{imperfective}}$ *da kupja* $_{\text{perfective}}$ */*kupuvam* $_{\text{imperfective}}$ *presen xljab.*
'I succeeded in buying fresh bread.'

---

[16]It should be noted, however, that the distribution of secondary imperfectivised verbs in these contexts differs from the distribution of primary imperfective verbs, e.g.,
Uspjavax $_{\text{imperfective}}$ *vinagi da izkupuvam* $_{\text{imperfective}}$ *navreme akciite.*
'I always managed to buy all stocks in time.'

Bulgarian perception verbs are further candidates for a treatment in terms of aspect government, inasmuch as only an imperfective aspect is allowed in their controlled verbal complements independently of the aspect of the matrix perception verb. This is illustrated in (5).

(5)

*Vižd**ax* $_{\text{imperfective}}$ / *vidj**ax* $_{\text{perfective}}$ *go da kupuv**a* $_{\text{imperfective}}$ / **kup**i $_{\text{perfective}}$ *knigi*

'I saw him buy books.'

Due to the loss of the infinitive in Bulgarian, and since some of the preserved non-finite forms have a rather bookish character, Bulgarian syntax is characterised by a high frequency of **finite verb forms**. The grammatical categories of the finite verb are as follows:

- **voice** (active and passive—with the latter being participial and reflexive[17])
- **tense** (usually nine tenses are distinguished: *segašno vreme* = present tense, *minalo nesvâršeno vreme* = imperfect, *minalo svâršeno vreme* = aorist, *minalo neopredeleno vreme* = perfect, *minalo predvaritelno vreme* = pluperfect, *bâdešte vreme* = future, *bâdešte predvaritelno vreme* = future perfect, *bâdešte vreme v minaloto* = past future, *bâdešte predvaritelno vreme v minaloto* = past future perfect).[18]
- **mood** (indicative, imperative, conditional[19])
- **person** and **number** (a system of three persons in two numbers)
- **gender** (masculine, feminine, and neuter)—for singular verb forms containing a participle
- **mode** (reported, non-reported)

  - *reported* : **reported type** (neutral, emphatic / dubitative)

---

[17] As noted in *Lindstedt 1985*, the two passive constructions, i.e. the participial passive and the reflexive passive, do not always have the same meaning and could be considered different voices. Contrary to, e.g., Russian, reflexive passive can also be formed from perfective verbs.

[18] Two further tenses are relative newcomers to the Bulgarian verbal system, and are usually not mentioned in school grammars: the **perfect progressive** resembles the perfect, except that the imperfect-stem active participle is used instead of the aorist-stem active participle (e.g., *xodel sâm/si/e, xodeli sme/ste/sa* are the respective forms of 'to walk'); the **inferred pluperfect** resembles the reported perfect or pluperfect, except that it has a clitic form of the auxiliary *sâm* also in the third person (e.g., *bil sâm/si/e kazal, bili sme/ste/sa kazali* are the respective forms of 'to say'). Both of these tenses are mostly used for inferred actions.

[19] Normally, the forms of conditional are analytic, but there are also rarer synthetic forms indicating readiness to perform the action of the verb. These, however, have almost gone out of modern usage.

- *non-reported* : **non-reported status** (vouched-for, neutral)

The neutral reported forms indicate that the information that is being conveyed to the hearer has been reported to the speaker by a third person, while the emphatic reported forms, which can also be called dubitative forms, imply that the speaker doubts the veracity of the message. The non-reported forms may have vouched-for[20] or neutral informational status. The vouched-for forms indicate that the speaker personally vouches for the veracity of the information, while the neutral non-reported forms do not carry any implication of vouching for or reporting.

The **non-finite verb forms** are: *segašno dejatelno pričastie* = present participle active, *minalo svâršeno dejatelno pričastie* = past participle active from the aorist stem, *minalo nesvâršeno dejatelno pričastie* = past participle active from the imperfect stem, *minalo stradatelno pričastie* = past participle passive, *deepričastie* = present gerund (active), i.e. adverbial deverbative, *otglagolno sâštestvitelno* = verbal noun, i.e. nominal deverbative.

The present participle active has attributive (adjectival) use only, while the past participle active from the imperfect stem can be used only in analytic verb forms. The nominal deverbative with the suffix *-ne* (unlike the obviously more lexicalised one with the suffix *-nie*) is included in the verbal paradigm because it can be formed from all imperfective verbs and is semantically regular denoting the mere activity. When used "verbally", e.g., without the definite article, nominal deverbatives in *-ne* can receive a direct object[21] and differ from all remaining verb forms in that they cannot express the subject—*pisane statija* 'writing an article', *predstavjane rabotata na kolektiva* 'presenting the work of the team', *kanene / pokanvane gosti za zabavata* 'inviting guests for the party'; when used "nominally", these nouns enter into the opposition definite / non-definite, and then the linking preposition *na* (meaning 'of' and combining functions borne by the genitive and the dative in other Slavic languages) is used—*pisane(to) na statija* '(the) writing of an article', *predstavjane(to) na rabotata na kolektiva* '(the) presenting of the work of the team', *kanene(to) / pokanvane(to) na gosti za zabavata* '(the) inviting of guests for the party'.

[20]This terminology is borrowed from *Hauge 1995*.
[21] Cf. also *Aronson 1967*, *Dyer and Fowler 1988*.

3.1.2 The Reflexive *si* and *se*

The lexical items *si* and *se* are ambiguous clitic elements: depending on the verbal lexeme and on the particular verb form, they can be interpreted either as reflexive particles or as reflexive pronouns. In the former case they are lexical or morphological formants that are part of the verb form, while in the latter case they are anaphoric pronouns which satisfy a dative (*si*) or an accusative (*se*) valence requirement of the head verb[22].

Bulgarian medial verbs[23] can be considered **lexically reflexive**, or reflexiva tantum, and the obligatorily occurring *se* or *si* is a lexical formant. The so-called "reflexive passive" can be regarded as a case of **morphological reflexivity**, with *se* being a morphological formant in an analytic verb form. Thus, only the anaphoric and reciprocal *si* and *se*, being truly referential, are instances of **syntactic (and also semantic) reflexivity**, and as complements are within the scope of the Binding Theory.[24]

3.1.3 "Analytic" Lexemes

Two cases of "analytic"[25] lexemes are to be distinguished in Bulgarian: on the one hand, the **medial verbs**, i.e. reflexiva tantum, and on the other hand, certain experience verbal predicates which I shall tentatively call **phrasal experiencer verbs**.

In the former case, the verb and the reflexive particle form a compound verb lexeme. The presence of the lexical formant *se* generally excludes any possibility of an accusative (i.e. direct object) NP complement with medial *se*-verbs, while the lexical formant *si* blocks only the dative (i.e. indirect object) *na*-NP complement. In general, the medial verbs are not recognisable

---

[22]The ethical *si*, being (diachronically) an instance of a free dative and regarded in the current approach rather as a modal particle, is not subcategorised by the head verb. Neither is the possessive *si* integrated into the verb complex as a result of a "possessor raising" ("climbing" of the possessive clitic from the direct object NP into the verb complex), e.g.,

(i) *Zagubix v navalicata novata si čanta.*
lost-1sg in crowd_def.art new_def.art REFL-POSS-CL bag
'I lost my new bag in the crowd.'

(ii) *Zagubix si v navalicata novata čanta.*
lost-1sg REFL-POSS-CL in crowd_def.art new_def.art bag
'I lost my new bag in the crowd.'

[23]See, e.g., *Penčev 1972*

[24]Only the anaphoric *se* and *si* have full-pronoun counterparts—*sebe si*, *na sebe si*—which are shaped in Bulgarian as composita.

[25] "Verbs with analitically shaped structure" is the term used, e.g., by *Čolakova 1973*.

as a derivation of a reflexive, passive or reciprocal meaning. Some medial *se*-verbs, like *boja se* 'to be afraid', *griža se* 'to care, worry about', *nadjavam se* 'to hope', *pojavjavam se* 'to appear', *staraja se* 'to strive', *sâglasjavam se* 'to agree', *usmixvam se* 'to smile', as well as some medial *si*-verbs, like *vâobrazjavam si* 'to imagine', *spomnjam si* 'to remember, recollect', have no parallel non-medial counterparts, i.e. such verbs cannot be used without the lexical formants even with different semantics. On the other hand, a few medial verbs are identical in meaning to their non-medial counterparts, e.g., *skitam se* / *skitam* 'to wander', *mrâkva se* / *mrâkva* 'it is getting dark', but it is more often the case that there is an additional semantic difference between the medial verb and the non-medial counterpart, e.g., *kazvam se* 'to be called' vs. *kazvam* 'to say', *karam se* 'to quarrel' vs. *karam* 'to drive', *namiram se* 'to be located' vs. *namiram* 'to find', *ljagam si* 'to go to bed' vs. *ljagam* 'to lie down', *služa si (s nešto)* 'to make use of s.t.' vs. *služa* 'to serve'.

In the lexical entry of a verb it must be specified if this particular verb is medial or not, and if it is, what lexical formant it combines with, e.g., as in Figure 11 for *smeja se* 'to laugh' and in Figure 12 for *spomnjam si* 'to recall, to recollect'

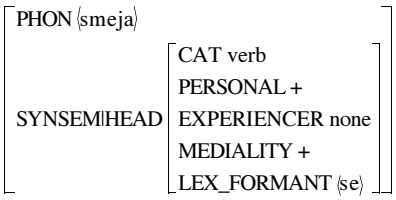


Figure 11

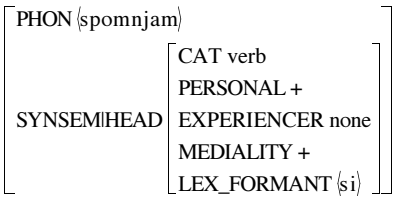


Figure 12

Most phrasal experiencer verbs are impersonal predicates of the type *domâčnjava mi* 'to become nostalgic', *trese me* 'to shiver'. The conjugation in person and number is realised by the obligatory pronominal clitic which is coreferential with the respective accusative or dative NP-complement

indicating the experiencer when this is overtly realised[26]. The result is formally quite similar to clitic replication—(6a) and (7a)—but this is a basically different phenomenon: the pronominal clitic is obligatory and can be viewed as a part of the compound verbal lexeme. Consequently, it is the full NP indicating the experiencer, and not the pronominal clitic, that is optional in such constructions—cf. (6b) and (7b) vs. the ungrammatical (6c) and (7c). And inasmuch as only determined nominal material can be referred to by a clitic, the corresponding dative or accusative experiencer NP must be determined[27].

(6)

a. *Mene ot sutrinta me trese.*

me-EXP_OBJ from morning_def.art ACC-1sg shiver-impers

'I shiver from the morning.'

b. *Ot sutrinta me trese.*

from morning_def.art ACC-1sg shiver-impers

c. **Mene ot sutrinta trese.*

me-EXP_OBJ from morning_def.art shiver-impers

(7)

a. *Na decata mnogo skoro im domâčnja za moreto.*

to children_def.art-EXP_OBJ very soon DAT-3pl become-nostalgic-impers about sea_def.art

'Very soon the children felt nostalgic about the sea.'

b. *Mnogo skoro im domâčnja za moreto.*

very soon DAT-3pl become-nostalgic-impers about sea_def.art

c.* *Na decata mnogo skoro domâčnja za moreto.*

to children_def.art-EXP_OBJ very soon become-nostalgic-impers about sea_def.art

There are also phrasal experiencer verbs which are personal, as illustrated by *lipsvam njakomu* ('s.o. misses me'—literally: 'I miss to s.o.') and by *boli / boljat me* ('to ache'—literally: 's.t. aches me'). The full-NP realisation of the experiencer is as a dative *na*-NP complement in (8) or as an accusative NP complement in (9), while the "actual" object functions as subject.

(8)

a. *Na vas sigurno vi lipsvat pet knigi.*

to you-EXP_OBJ probably DAT-2pl miss-3pl five books-SUBJ

'You miss probably five books.'

---

[26]The fact that the object clitic is an obligatory structural element in these constructions inasmuch as it plays the "semantic role experiencer" is mentioned in, e.g., *Cyxun 1968, Nicolova 1986, Popov 1962*.

[27]Cf. Section 4.1 for a discussion of this concept.

b. *Sigurno vi lipsvat pet knigi.*
probably DAT-2pl miss-3pl five books-SUBJ

c. **Na vas sigurno lipsvat pet knigi.*
to you-EXP_OBJ probably miss-3pl five books-SUBJ

(9)

a. *Vlado go boljat krakata.*
Vlado-EXP_OBJ ACC-3sg,masc ache-3pl legs_def.art-SUBJ
'Vlado feels a pain in his legs.'

b. *Boljat go krakata.*
ache-3pl ACC-3sg,masc legs_def.art-SUBJ
'He feels a pain in his legs.'

c. **Vlado boljat krakata.*
Vlado-EXP_OBJ ache-3pl legs_def.art-SUBJ

In the lexical entry of a verb, the information on the obligatory experiencer formant must be explicitly specified. The appropriate index (i.e. person, number and gender) and case agreement with the respective subcategorised experiencer object has to be established, whereby the latter is supposed to be a determined NP—an indispensable requirement for any nominal material that is referred to by a pronominal clitic[28]. This is illustrated in Figure 13 for, e.g., *trese me* and in Figure 14 for, e.g., *lipsvam njakomu*.

Of course, there might be **medial experiencer verbs** which have both a reflexive and an experiencer formant, e.g., as the verb *privižd*am se njakomu* 'to come into vision of s.o.' given in (10). The corresponding lexical entry is sketched in Figure 15.

(10)

a. *Na posetitelite im se privižd*at neverojatni nešta.*
to visitors_def.art-EXP_OBJ DAT-3pl SE come-into-vision incredible things-SUBJ
'The visitors have visions of incredible things.'

b. *Privižd*at im se neverojatni nešta.*
come-into-vision DAT-3pl SE incredible things-SUBJ
'They have visions of incredible things.'

c.* *Na posetitelite se privižd*at neverojatni nešta.*
to visitors_def.art-EXP_OBJ SE come-into-vision unusual things-SUBJ

[28]Cf. Section 4.1.

As can be expected, both the medial verbs and the experience verbal predicates discussed always create a morphosyntactic verb-complex constituent.

The obligatoriness of the respective formant with phrasal experiencer verbs makes it function as an agreement inflection which realises a certain type of experiencer-verb concord. A full NP in accusative or dative syntactic case indicating the experiencer thus always has its index specified within the verb-complex constituent, and in addition is always determined.

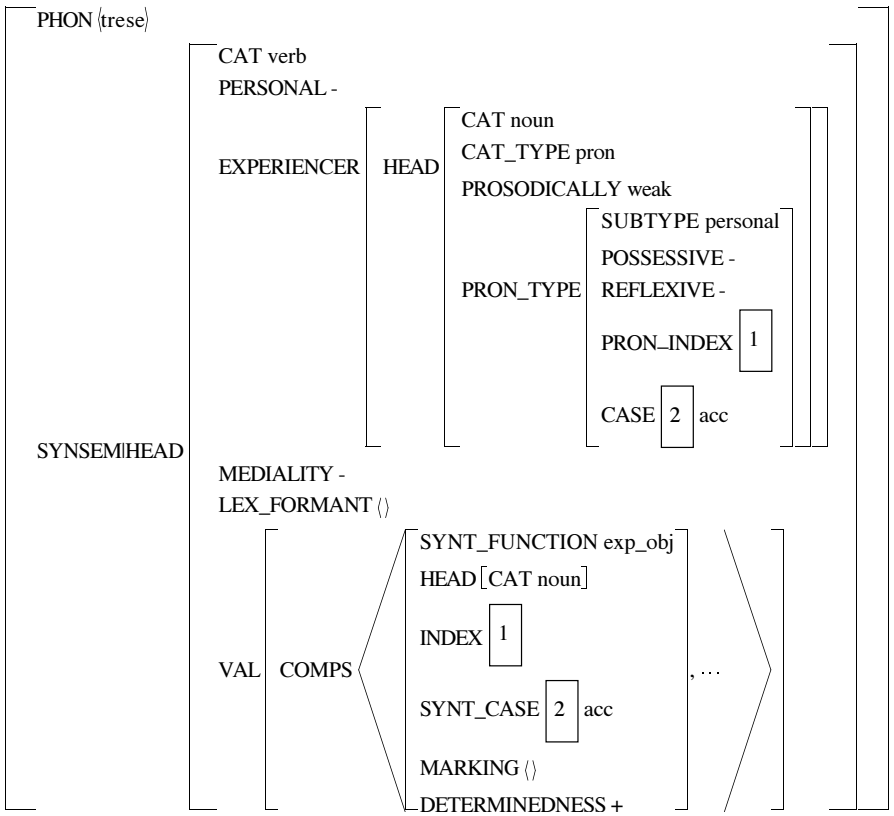


Figure 13

There is a clear tendency for the experiencer (dative or accusative) NP to precede the verb complex and thus to occupy the default neutral position of syntactic subjects in clauses headed by personal (non-phrasal-

experiencer) verbs. This is true even in constructions with *personal* phrasal experiencer verbs, where a formal subject is, in fact, available. In the latter case, however, the subject tends to follow the verb complex, i.e. to occur in the default neutral object position.

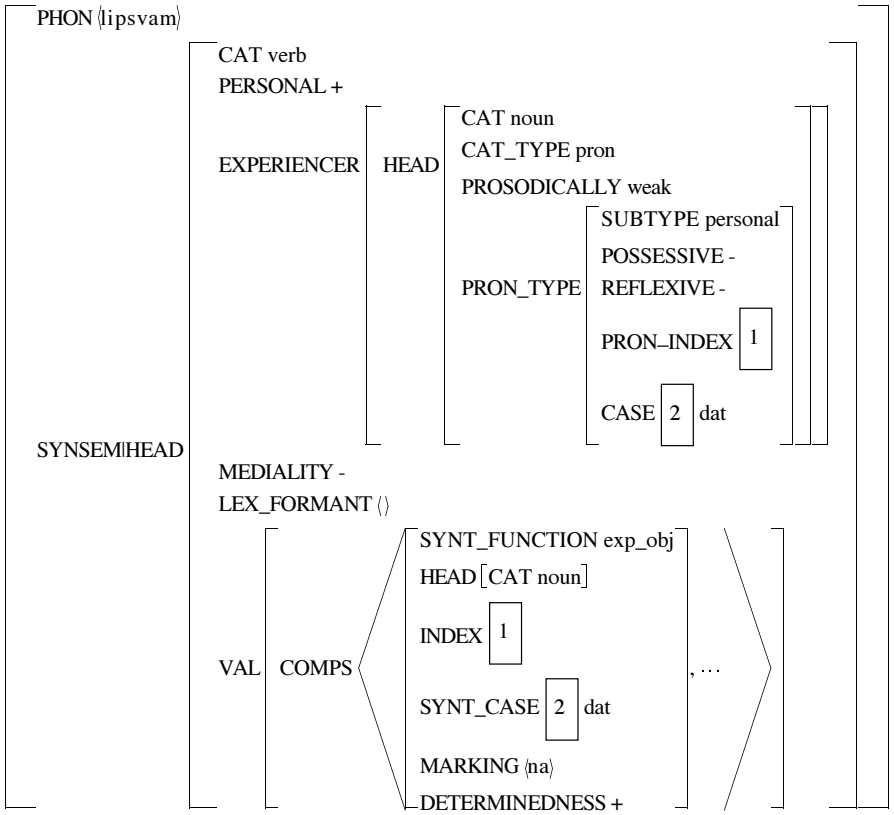


Figure 14

Since the pronominal clitic formant can never be dropped, while the coreferent full NP may be missing and this would not actually affect the syntactic well-formedness of the construction, it would be inadequate to view the formal cooccurrence of a clitic pronoun and a full NP in such constructions as "clitic doubling" or, in the terminology adopted here, as replication of nominal material. Indeed, considering it an agreement or concord phenomenon proves to be linguistically better motivated.

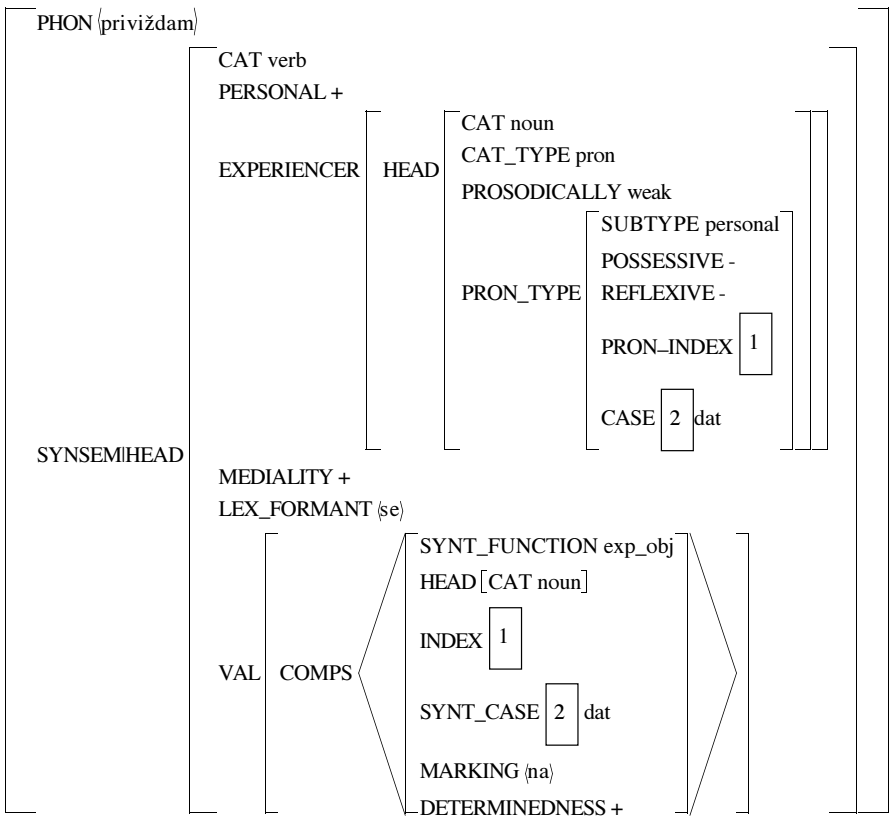


Figure 15

3.1.4 Analytic Morphological Forms

Most generally, Bulgarian verb forms are either **synthetic**, i.e. consisting of exclusively the main verb (e.g., (11a)), or **analytic / periphrastic**, i.e. employing morphosyntactic marking by auxiliary verbs and / or particles, and possibly involving lexical formants (e.g., (11b)).

(11)

a. *gledam*

look - imperfective aspect, present tense, 1st person, singular, active voice, indicative mood, non-reported

b. *štjal sâm da se usmixna*

AUX AUX PRT LEX-FORMANT smile - perfective aspect, future / past_future tense, 1st person, singular, masculine, active voice, indicative mood, reported

The features involved in the subject-verb agreement are **person**, **number**, and **gender**, and it is often the case that these are expressed by different components of a given analytic verb form. On the other hand, since the same agreement feature can be available in more than one component of the analytic verb form, high redundancy can be observed. Finite auxiliary verbs and finite main verbs are marked for person and number, while the respective auxiliary-verb and main-verb participles carry information on number and gender. For the sake of illustration, let us consider in (12) some indicative active-voice forms of the biaspectual verb *xodja* 'to go', focusing on the agreement information only.

(12)

a. past future (1st person, singular):

štjax-1sg *da* *xodja*-1sg
FIN-AUX PRT FIN-MAIN

b. past future perfect (1st person, singular, feminine):

štjax-1sg *da* *sâm*-1sg *xodila*-sg,fem
FIN-AUX PRT FIN-AUX PCP-MAIN

c. reported past future (1st person, singular, feminine):

štjala-sg,fem *sâm*-1sg *da* *xodja*-1sg
PCP-AUX FIN-AUX PRT FIN-MAIN

d. reported past future perfect (1st person, singular, feminine):

štjala-sg,fem *sâm*-1sg *da* *sâm*-1sg *xodila*-sg,fem
PCP-AUX FIN-AUX PRT FIN-AUX PCP-MAIN

The syntactic (and semantic) *head* of any verb complex is the main verb, while the contingent auxiliary verbs and particles actually realise the analytic verbal inflection. The treatment of the (possibly multiple) auxiliary verbs occurring in a given analytic verb form I am proposing here is based on regarding them as means of morphosyntactic marking rather than raising verbs. The category "marker" is fairly extensive in my understanding. In particular, I assume that all components of analytic verb forms that have a form-building function—auxiliary verbs, the conjunctive, the future-tense and the reflexive-passive particles—are *mophosyntactic markers*.

It is important to keep in mind that Bulgarian auxiliary verbs are primarily involved in analytic (periphrastic) form-building morphology

rather than in syntactic relations like subcategorization or valence. A crucial property of the auxiliary markers is to contribute to the verbal HEAD features in two aspects: first, being components of the respective verb form, they participate in specifying the tense, voice, mood and mode values of the actual verb form, and second, their agreement morphology is relevant for determining the agreement features of the entire verbal category, since information on person, number and gender may come from different components of the analytic verb form, as already illustrated in (12).

Even with a status of morphosyntactic markers, Bulgarian auxiliary verbs can head a morphosyntactic grouping which I shall call *auxiliary complex* (AuxC)[29]. The basic motivation for distinguishing an AuxC as a morphosyntactic constituent of the verb complex can be found when observing examples like those in (13) and (14). Within the verb complex, the non-contact position of the auxiliary and the pronominal clitic in (13d-e) and of the two auxiliaries in (14d-e) leads to ungrammaticality. It is obvious that such regularities have to be accounted for in the constituent structure.

(13)

a. *Az mu bjax dal knigata.*
I DAT-3sg,masc AUX give-pcp book_def.art
'I had given him the book.'

b. *Az bjax mu dal knigata.*
I AUX DAT-3sg,masc give-pcp book_def.art

c. *(Az) dal mu bjax knigata.*
(I) give-pcp DAT-3sg,masc AUX book_def.art

d.* *Az mu dal bjax knigata.*
I DAT-3sg,masc give-pcp AUX book_def.art

e.* *Az bjax dal mu knigata.*
I AUX give-pcp DAT-3sg,masc book_def.art

(14)

a. *Štjal bil da dojde.*
AUX AUX CONJ-PRT come.
'He would come.' - reported

b. *Bil štjal da dojde.*
AUX AUX CONJ-PRT come.

[29]In *Avgustinova and Oliva 1991* this morphosyntactic constituent is called auxiliary phrase (AuxP).

c. *Da dojde štjal bil.*
CONJ-PRT come AUX AUX.
d.* *Štjal da dojde bil.*
AUX CONJ-PRT come AUX.
e.* *Bil da dojde štjal.*
AUX CONJ-PRT come AUX.

A further important feature of the proposed analysis is that the introduction of verbal **complement** clitics is considered not a matter of lexicon but a morphosyntactic phenomenon to be handled on the level of morphosyntactic constituency—Figure 16. At the same time, there seems to be no empirical motivation for introducing an intermediate morphosyntactic level with Bulgarian non-verbal categories, i.e. nouns, adjectives, adverbs, etc. Therefore, they are assumed to form syntactic phrases only.

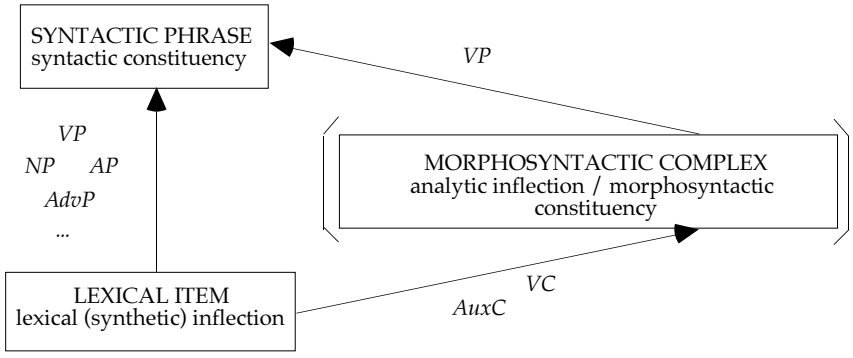


Figure 16

3.2 Clitic Elements in the Bulgarian Verb Complex

The placement of clitics belongs to the interface area between (morpho)syntax and (prosodic) phonology. An analysis which does not incorporate this intuition in some form would hardly be adequate. Usually clitics are related to a certain syntactically determined domain. This allows us to speak of sentential, verbal, nominal, etc. clitics, i.e. of clitics belonging to the clausal, verbal, nominal, and other constituents of the particular language. Within the syntactic domain, clitics are subject to general as well as to individual constraints on their prosodic environment. This makes the

problems of clitic syntax appear rather messy if prosodic factors are ignored.

The articulation of the syntactic domains into *accentual units* provides their prosodic structure. As we shall see, certain clitics-related phenomena are apparently good candidates for treatment in terms of such a structure.

The term *verbal clitics* used in this study is intended to refer to those clitic elements whose syntactic domain is the verb complex. With respect to their phonological behaviour, they are attached to the verb complex in all circumstances.

This section will focus on the positioning of verbal clitic sequences. In conformity with more general constraints that are commonly assumed to regulate the surface distribution of verbal clitics in Bulgarian, the proposed analysis takes advantage of a classification which allows us to attribute particular clitic placements to a specific clitic interaction.

3.2.1 Background

It is characteristic of verbal clitics that they are prosodically hosted by a stressed component of the verb form, and all phonologically / prosodically weak verb-complex components exhibit this property. No sentential clitics in the traditional sense are distinguishable in Bulgarian; hence no special Wackernagel position is available in the syntactic domain of the Bulgarian clause. Clitics that are commonly classified as sentential occur in the syntactic domain of the verb complex which exhibits a fair degree of freedom with respect to its positioning in the sentence. A slight trace of the second-position phenomenon, however, can be detected as a verb-complex internal regularity which I shall call the **quasi-second-position condition:** no more than one phonologically strong element is allowed to precede the clitics within this constituent, as illustrated by the variants in (15).

(15)

a. *bjax ja vidjal*

AUX-imp,1sg ACC-3sg,fem see-pcp,sg,masc

'I had seen her.'[30]

[30]The ungrammaticality of the variant below is obviously not predicted by the quasi-second-position condition.

* ... *ja vidjal bjax*

... ACC3sg.fem seepcp,sg,masc AUXimp,1sg

Instead, it is due to the discontinuity of the AuxC (see Section 3.1.4).

b. *vidjal ja bjax*
see-pcp,sg,masc ACC-3sg,fem AUX-imp,1sg

c. ... *ja bjax vidjal*
... ACC-3sg,fem AUX-imp,1sg see-pcp,sg,masc

d. *?? vidjal bjax ja*
see-pcp,sg,masc AUX-imp,1sg ACC-3sg,fem

e. * *bjax vidjal ja*
AUX-imp,1sg see-pcp,sg,masc ACC-3sg,fem

The term *clitic* in Bulgarian traditionally refers to a lexical item which is phonologically weak and usually unstressed[31]. This word "leans" onto some phonologically strong (stress-bearing) lexical item which hosts it in a larger accentual unit (also referred to as "prosodic word"). For the purposes of the present analysis, clitics are considered to be **primarily unstressed**, as opposed to stress-bearing non-clitic lexical items, which are considered to be primarily stressed. This assumption does not really exclude cases where, as a rule, certain verb clitics bear an accent, since these are **secondarily-stressed**, phonologically weak elements. The main support for this claim is provided by the fact that these elements still preserve clitic properties with respect to the word order[32].

As their name suggests, verbal proclitics and enclitics have predetermined positions in the accentual unit. On the other hand, the existence of verbal clitics with no predetermined positioning, commonly called "movable", or "endoclitics", is a well-known specific property of Bulgarian. The verb clitics of this special type either precede (like proclitics—(16b)) or follow (like enclitics—(16a)) their prosodic host. It is important to note, however, that this behaviour does not affect the prosodic organisation (in the sense of articulation into accentual units) because the host element remains constant. Together with the enclitics, the "movable" clitics are subject to a word-order constraint that prevents them from occurring in clause-initial position or after an intonational pause. This general constraint on clitic distribution in modern Bulgarian (henceforth, the **clause-initial restriction**) undoubtedly induces the observed "movability".

[31]See, e.g., *Tilkov 1980*.
[32]Cf. also *Penčev 1980, Penčev 1984*.

(16)

a. *Vidjax___ja.*

see-aor,1sg ACC-3sg,fem

'I saw her.'[33]

b. *Otnovo ja___vidjax.*

again ACC-3sg,fem see-aor,1sg

'I saw her again.'

c. **Ja___vidjax.*

ACC-3sg,fem see-aor,1sg

Inasmuch as all phonologically weak elements leaning onto the same host tend to occur adjacent to one another, this results in the creation of a strictly ordered sequence—the **clitic cluster**[34]. The mutual order of the components of this prosodically set grouping is not a matter of analysis here but is rather assumed—following *Ewen 1975* and *Hauge 1976*, and as suggested in *Avgustinova and Oliva 1991*—to conform to the linear precedence scheme (henceforth, **clitic LP-scheme**) presented in Figure 17. Though the interrogative particle must be recognised as verb enclitic when it modifies the verbal constituent, it cannot be included in the clitic LP-scheme. The position of this element in general, and its incorporation into a clitic cluster in particular, are determined by the only requirement that, as a verbal clitic, it immediately follows the leftmost stressed element in the verb complex.

CONJ-PRT	NEG-PRT	FUT-PRT	AUX	pronoun pers./refl.[35]	pronoun pers./refl.[36]	AUX
da <	ne <	šte <	sâm-1/2 <	dative(s) <	accusative <	sâm-3sg[37]

Figure 17

Of course, not all of these elements can be combined into a real clitic cluster: certain combinations of short pronouns are excluded, namely, dative non-reflexive short pronouns of any person cannot cooccur with a

[33]The prosodic units are marked off by linking their components with underscores.

[34]See *Ewen 1975*, *Hauge 1976*, etc.

[35]Note: the reflexive particle *si* is homonymous with the reflexive dative clitic; it occurs in the position reserved for the dative(s).

[36]Note: the reflexive particle *se* is homonymous with the reflexive accusative clitic; it occurs in the position reserved for the accusative.

[37]Note: the 3rd person plural form of *sâm* can occur in either of the two positions reserved for the clitic present-tense forms of this auxiliary verb

1st- or 2nd-person short accusative pronoun; furthermore, no clitic cluster can simultaneously contain the conjunctive and the future-tense particles, since such a combination does not occur in any analytic verb form; moreover, any cooccurrence in one clitic cluster of two clitic forms of the verb *sâm* is ruled out.

Since the positioning of verb-clitic sequences is of primary concern in this analysis, let us consider the range of material offered by the language. The most extended type of verb complex allows for two clitic clusters hosted by different verb-form components. These constructions involve analytic verb forms containing the primarily stressed future auxiliaries *štjax* or *njama*[38], as illustrated in (17-19)[39].

(17)

a. ***Štjal___[sâm]*** *spored nego* ***[da sâm vi ja]___pokazal***.[40]

FUT-AUX-pcp,sg,masc_[CL-AUX-1sg] according to him [CONJ-PRT CL-AUX-1sg DAT2-pl ACC-3sg,fem]_show-pcp,sg,masc

'According to him, I will have shown her to you.'- reported

b. *Spored nego* ***[sâm]___štjal*** ***[da sâm vi ja]___pokazal***.

according to him [CL-AUX-1sg]_FUT-AUX-pcp,sg,masc [CONJ-PRT CL-AUX-1sg DAT-2pl ACC-3sg,fem]_show-pcp,sg,masc

(18)

a. ***Njamaše___[li]*** ***[da ni ja]___pokažeš****?*

NEG-FUT-AUX-imp,impers_[Q] [CONJ-PRT DAT-1pl ACC-3sg,fem]_show-pres,2sg

'Won't you have shown her to us?'

b. ***Njamalo___[li]*** ***[da ni ja]___pokažeš****?*

NEG-FUT-AUX-pcp,sg,neut,impers_[Q] [CONJ-PRT DAT-1pl ACC3sg,fem]_show-pres,2sg

'Won't you show her to us?' - reported

(19)

a. *Ti* ***štjal___[li si]*** *predvaritelno* ***[da si ni ja]___pokazal****?*

you FUT-AUX-pcp,sg,masc_[Q CL-AUX-2sg] in advance [CONJ-PRT CL-AUX-2sg DAT-1pl ACC-3sg,fem]_show-pcp,sg,masc

'Will you have shown her to us in advance?' - reported

---

[38]The latter is the negative variant of the former.

[39]In the glosses below, the following abbreviations are used: FUT-AUX for the positive future auxiliary verb, NEG-FUT-AUX for the negative future auxiliary verb, CL-AUX for the clitic present-tense forms of the *sâm*-auxiliary, BE1 for the non-clitic forms of the *sâm*-auxiliary, BE2 for the *bâda*-auxiliary, Q for the interrogative particle, NEG-PRT for the negative particle, CONJ-PRT for the conjunctive particle, FUT-PRT for the future-tense particle, DAT for dative clitics, ACC for accusative clitics.

[40]The boundaries of clitic clusters are delimited by square brackets; underscores conjoin accentual units.

b. *Utre* ***štjala___[li si]*** ***[da mi gi]___bâdeš*** ***pokazala****?*

tomorrow FUT-AUX-pcp,sg,fem_[Q CL-AUX-2sg] [CONJ-PRT DAT-1sg ACC-3pl]_BE2-2sg show-pcp,sg,fem

'Will you have shown them to me tomorrow?' - reported

c. ***Šteli___[li sme]*** *spored tjax* ***[da sme im]___bili predstaveni*** *utre?*

FUT-AUX-pcp,pl_[Q CL-AUX-1pl] according to them [CONJ-PRT CL-AUX-1pl DAT-3pl]_BE1-pcp,pl introduce-pass-pcp,pl tomorrow

'According to them, shall we have been introduced to them tomorrow?' - reported

d. *Utre* ***šteli___[li ste]*** *navreme* ***[da im]___bâdete predstaveni****?*

tomorrow FUT-AUX-pcp,pl_[Q CL-AUX-2pl] in time [CONJ-PRT DAT-3pl]_BE2-2sg introduce-pass-pcp,pl

'Will you be introduced to them in time tomorrow?' - reported

The remarkable situation in which both clusters consist of more than one element can be realised only in the 1st or 2nd person of the reported complex *šta*-tenses, as in (19). In all cases distinct from these constructions, only one clitic cluster is formed, as in (20).

(20)

a. *Dnes* ***[sâm mu ja]___pokazval****.*

today [CL-AUX-1sg DAT-3sg,masc ACC-3sg,fem]_show-pcp,sg,masc

'I have shown her to him today.'

b. ***Pokazval___[li si mu ja]*** *dnes?*

show-pcp,sg,masc_[Q CL-AUX-2sg DAT-3sg,masc ACC-3sg,fem] today

'Have you shown her to him today?'

c. ***[Ne sâm mu ja]___pokazval*** *dnes*.

[NEG-PRT CL-AUX-1sg DAT-3sg,masc ACC-3sg,fem]_show-pcp,sg,masc today

'I haven't shown her to him today?'

d. ***[Ne si li mu ja]___pokazval*** *dnes?*

[NEG-PRT CL-AUX-2sg Q DAT-3sg,masc ACC-3sg,fem]_show-pcp,sg,masc today

'Haven't you shown her to him today?'

Interestingly, the tendency of the unstressed elements participating in the same accentual unit to group into a cluster is not always obeyed. There are remarkable cases where, due to the individual prosodic requirements of the clitics involved, a "discontinuous" clitic sequence is the only possible variant, as in (21a-c), or a grammatical alternative, as in (21d).

(21)

a. ***[Šte si ni ja]___pokazal___[li]*** *navreme?*

[FUT-PRT CL-AUX-2sg DAT-1pl ACC-3sg,fem]_show-pcp,sg,masc_[Q] in time

'Will you have shown her to us in time?'

b. ***[Šte]___zaminavate___[li]*** *za Sofia?*
[FUT-PRT]_leave-pres,2pl_[Q] for Sofia
'Are you leaving for Sofia?'

c. ***[Ne]___igraete___[li]*** *v tozi otbor?*
[NEG-PRT]_play-pres,2pl_[Q] in this team
'Don't you play on this team?'

d. ***[Ne]___beše___[li mu]*** ***predstavena*** *na zabavata?*
[NEG-PRT]_BE1-imp,2sg_[Q DAT3sg,masc] introduce-pass-pcp,sg,fem at the party
'Weren't you introduced to him at the party?'

variant of: ***[Ne mu li]___beše*** ***predstavena*** *na zabavata?*

3.2.2 An Approach to Clitic Syntax

The primary goal of the present section is to propose an analysis of clitic cluster placement within the verb-complex constituent. The key idea is to incorporate the generalisations from the previous section into a classification on the basis of further criteria related to cluster formation and the positioning of clitic sequences. I propose to divide Bulgarian verb clitics into **core** and **peripheral**, from the point of view of clitic-cluster formation. It is possible to attribute all legitimate clitic cluster placements to the interaction of these two types, together with general constraints on clitic order (e.g., the clause-initial restriction, the clitic LP-scheme, etc.).

Let us refrain from discussing in detail the rather simple single-clitic cases. They can certainly be viewed as a special case of one-element clitic clusters. As such they will only trivially obey the general rules of clitic placement, since single clitics realise their inherent properties without any complex interaction with other clitics. Thus, the data on which the present analysis will concentrate successively reduces to verbal clitic sequences in which at least one of the clitics is a core clitic. This will provide a better illustration of the core-periphery opposition which, as we shall see, offers considerable explanatory potential.

The clitics in the Bulgarian verb complex have an independently observed tendency to attach accentually to, and therefore combine syntactically with, an auxiliary verb if one is available in the respective verb form; this has justified the introduction of a phrase of the auxiliary verb in *Avgustinova and Oliva 1991*, a concept which is presented in the current study under the term *auxiliary morphosyntactic complex*, or **AuxC** (cf. Section 3.1.4). Syntactically, the AuxC is a sub-constituent of the verb

complex. It is headed by an auxiliary verb and can never be discontinuous. The ungrammaticality of (22e) results from an AuxC discontinuity, while in (22d) only the quasi-second-position condition is violated; in (22f) both of these problems arise, rendering the wording unacceptable.

(22)

a. *(...) bjax vi ja pokazvala*[41]

BE1-imp,1sg DAT-2pl ACC-3sg,fem show-pcp,sg,fem

'I had shown her to you.'

b. ... *vi ja bjax pokazvala*

DAT-2pl ACC-3sg,fem BE1-imp,1sg show-pcp,sg,fem

c *(...) pokazvala vi ja bjax*

show-pcp,sg,fem DAT-2pl ACC-3sg,fem BE1-imp,1sg

d. *?? (...) pokazvala bjax vi ja*

show-pcp,sg,fem BE1-imp,1sg DAT-2pl ACC3sg,fem

e. *... *vi ja pokazvala bjax*

DAT-2pl ACC-3sg,fem show-pcp,sg,fem BE1-imp,1sg

f. **(...) *bjax pokazvala vi ja*

BE1-imp,1sg show-pcp,sg,fem DAT-2pl ACC-3sg,fem

With respect to the prosodic organisation, two cases must be distinguished. In the first, the syntactic constituent corresponds to (coincides with) the prosodic word. This is possible when AuxC is headed by a non-clitic auxiliary verb which hosts the verbal clitics. In the second case, the AuxC does not form an autonomous accentual unit, since it is headed by a clitic auxiliary verb. Such a phonologically weak AuxC is prosodically hosted either by the main verb or by some other non-clitic auxiliary verb. The discrepancy between the syntactic and prosodic structure is obvious here. As will be seen in the discussion of peripheral clitics, only a prosodically autonomous AuxC allows for optional splitting of the clitic sequence. In particular, the host (the head auxiliary) may be located on the periphery-core border, instead of following the (pro)clitic cluster. A clitic sequence incorporating a clitic auxiliary verb cannot be split by an intervening host because this would lead to an AuxC discontinuity. The conclusion to be drawn is that, at the very least, the categorial specification of the elements contained in the clitic cluster must be accessible in some way at the prosodic level of language description.

[41] AuxCs are marked with double underlines.

3.2.2.1 Core Clitic Cluster

I consider the movable (loosely positioned) verb clitics—such as present tense forms of the auxiliary / copula *sâm*, the short dative and accusative forms of the personal and reflexive pronouns, and the reflexive particles—to be **basic** for the process of cluster formation. From this standpoint, they constitute the **core** of any clitic cluster.

The **core clitic cluster** consists of only core (basic) clitics and exhibits the same movability that is assumed to be a distinctive feature of each of its contingent elements. The rigid internal word order of the core clitic cluster is, of course, predetermined by the clitic LP-scheme. The ungrammaticality of (23a) is now predictable as a consequence of the clause-initial restriction. It is further anticipated that the core clitic cluster exhibits enclitic behaviour exactly when there is no prosodically autonomous lexical material to precede the accentual unit (23b). The sentence/clause-initial position in this case is occupied by the host itself, which can in turn take the core clitic cluster only as enclitics. Otherwise, the behaviour of the core clitic cluster is apparently proclitic, as illustrated in (23c-d).

(23)

a. ***(Sâm ti ja)___***pokazvala na snimkata sigurno.*[42]*

(CL-AUX-1sg DAT-2sg ACC-3sg,fem)_show-pcp,sg,fem on picture certainly

'Maybe I have shown her to you on the picture.'

b. *Pokazvala___**(sâm ti ja)** na snimkata sigurno.*

show-pcp,sg,fem_(CL-AUX-1sg DAT-2sg ACC-3sg,fem) on picture certainly

c. *Na snimkata sigurno **(sâm ti ja)___**pokazvala.*

on picture certainly (CL-AUX-1sg DAT-2sg ACC-3sg,fem)_show-pcp,sg,fem

d. **Na snimkata sigurno pokazvala___**(sâm ti ja)**.*

on picture certainly show-pcp,sg,fem_(CL-AUX-1sg DAT-2sg ACC-3sg,fem)

3.2.2.2 Periphery of the Clitic Cluster

The periphery of the clitic cluster is constituted by strictly proclitic and strictly enclitic elements which are integrated into the verb complex by virtue of being verbal clitics. Unlike core clitics, they cannot be viewed as obligatory components of a clitic cluster, since they need not always join it, as illustrated in (21).

---

[42] Core clitic clusters are enclosed within parentheses.

My main claim is that attachment of a peripheral clitic to the core neutralises the movability of the latter so that the position of the **entire** clitic cluster with respect to the host is fixed. The ultimate result complies with the prosodic characteristic of the **leftmost peripheral clitic**. In particular, the clitic cluster attains the property of being *proclitic* if **any** of the proclitics proper is involved in it, as in (24a), and *enclitic* if **only** the enclitic proper *li* has joined the core clitic(s), as in (24b).

(24)

a. *Sestrata **[šte** (ni go e)]___pokazala ošte predi operacijata.*

nurse [FUT-PRT (DAT-1pl ACC-3sg,masc AUX-3sg)]_show-pcp,sg,fem still before surgery

'The nurse will have already before the surgery shown him to us.'

b. *Sestrata pokazvala___[**li** (ni go e)]?*

nurse show-pcp,sg,fem_[Q (DAT-1pl ACC-3sg,masc CL-AUX-3sg)]

'Has the nurse shown him to us?'

A cluster containing a peripheral clitic is, for obvious reasons, out of the scope of the clause-initial constraint. What really proves to be relevant for its positioning is the core—periphery interaction just presented. In the next two sections, the essential contribution of each peripheral verbal clitic will be analysed from this perspective.

3.2.2.3 Contribution of Proclitics Proper

The strictly proclitic elements, or **proclitics proper**, always precede the core clitic(s) in a clitic cluster, obeying the word-order and cooccurrence restrictions introduced by the clitic LP-scheme—cf. (25).

(25)

a. *[**Šte** (sâm mu ja)]___pokazvala.*

[FUT-PRT (CL-AUX-1sg DAT-3sg,masc ACC-3sg,fem)]_show-pcp,sg,fem

'I will have shown her to him.'

b. *[**Ne** (sâm mu ja)]___pokazvala.*

[NEG-PRT (CL-AUX-1sg DAT-3sg,masc ACC-3sg,fem)]_show-pcp,sg,fem

'I haven't shown her to him.'

c. *[**Ne šte** (sâm mu ja)]___pokazvala.*

[NEG-PRT FUT-PRT (CL-AUX-1sg DAT-3sg,masc ACC-3sg,fem)]_show-pcp,sg,fem

'I will not have shown her to him.'

d. *Štjax ... [**da** (sâm mu ja)]___pokazvala.*

FUT-AUX-imp,1sg ... [CONJ-PRT (CL-AUX-1sg DAT-3sg,masc ACC-3sg,fem)]_show-pcp,sg,fem

'I would have shown her to him.'

*e. Štjax ... [**da ne** (sâm mu ja)]___pokazvala.*[43]

FUT-AUX-imp,1sg ... [CONJ-PRT NEG-PRT (CL-AUX-1sg DAT-3sg,masc ACC-3sg,fem)]_show-pcp,sg,fem

'I would not have shown her to him.'

From a prosodic viewpoint, any combination of proclitics that can legitimately cooccur in the accentual unit behaves like a compound proclitic. Therefore, I propose a special term, **proclitic group**, to refer to such proclitic sequences. Quite like single clitics, exclusively proclitic groups are also a special case of clitic clusters. But, as has already been mentioned, clitic clusters containing a combination of core or both core and peripheral clitics are of primary concern in the current analysis. The examples in (26) illustrate constructions where no core clitics are involved, while strictly proclitic elements are present in the respective accentual unit and stand in the prescribed order.

(26)

a. *Štjax* {***da ne***}___*četa.*[44]

FUT-AUX-imp,1sg {CONJ-PRT NEG-PRT}_read-pres,1sg

'I would not read.'

b. {***Ne šte***}___*četeš.*

{NEG-PRT FUT-PRT}_read-pres,2sg

'You will not read.'

Within the accentual unit, the proclitics proper (and, of course, the proclitic groups) appear to be sensitive to non-prosodic information of a certain type—the morphosyntactic category of the clitic-cluster components and / or of the host. In particular, when proclitics participate in a prosodically autonomous AuxC, they may optionally be detached from the core by the hosting head auxiliary verb. The hosting head, however, is allowed to occur only at the periphery-core border.

Now let us separately consider each of the proclitics proper in an attempt to illustrate their individual interaction with the core and to further explain the observed regularities.

The future-tense particle *šte* is a form-building morphological element in the indicative verb forms for future (*šte čete*—'(he) will read') or future

[43]Examples (25c,e) are relatively rare but still grammatically acceptable variants of the respective verb forms; the more frequently used variants take advantage of the negative future auxiliary *njama* (i.e. *njama da sâm mu ja pokazvala* and *njamaše da sâm mu ja pokazvala*).

[44]Exclusively proclitic groups are enclosed within braces.

perfect (*šte e čel // šte bâde čel*—'(he) will have read'). If there is a clitic auxiliary verb in the accentual unit, as in (27a), or if the host is a full verb (27b), *šte* must join the clitic cluster.

(27)

a. ***[Šte** (si ni ja)]___pokazval*.

[FUT-PRT (CL-AUX-2sg DAT-1pl ACC-3sg,fem)]_show-pcp,sg,masc

'You will have shown her to us.'

b. ***[Šte** (ni ja)]___pokazvaš*.

[FUT-PRT (DAT-1pl ACC-3sg,fem)]_show-pres,2sg

'You will show her to us.'

This inclusion, however, seems arbitrary with verb forms containing the auxiliary verb *bâda*, as in (28). Here the accentual unit coincides with AuxC, which actually means that the head auxiliary hosts the verbal clitics. The host may split the clitic sequence right on the border between peripheral and core clitics, attaching the latter as enclitics.

(28)

a. ***[Šte** (ni ja)]___bâdeš pokazval*.

[FUT-PRT (DAT-1pl ACC-3sg,fem)]_BE2-2sg show-pcp,sg,masc

'You will have shown her to us.'

b. ***Šte**___bâdeš___(ni ja) pokazval*.

FUT-PRT_BE2-2sg_(DAT-1pl ACC-3sg,fem) show-pcp,sg,masc

Constructions that involve a proclitic group offer an even better illustration of this option in the parallel examples in (29).

(29)

a. *[Ne **šte** (ni ja)]___bâdeš pokazval*.

[NEG-PRT FUT-PRT (DAT-1pl ACC-3sg,fem)]_BE2-2sg show-pcp,sg,masc

'You will not have shown her to us.'

b. {*Ne **šte***}*___bâdeš___(ni ja) pokazval*.

{NEG-PRT FUT-PRT}_BE2-2sg_(DAT-1pl ACC-3sg,fem) show-pcp,sg,masc

Da is a morphological component of verb forms that include a non-clitic auxiliary verb of the type *šta* or *njama*[45]. In its form-building function, I consider *da* to be a **conjunctive particle**, though there is no consensus in

[45] *Da* occurs also in imperative verb forms of the 2nd person containing the negative imperative auxiliary *nedej / nedejte*, e.g.,

Nedej da mu ja pokazvaš.
NEG-IMPER-2sg CONJ-PRT DAT-1pl ACC-3sg,fem show-pres,2sg
'Don't show her to him.'

the literature as to the nature of this lexical item in general, and particularly in compound verbal constructions[46]. However, these are side issues for the present analysis which is primarily concerned with word-order phenomena. The substantial point is that this proclitic element is always a verbal clitic, which further implies that the word-order properties of *da* are identical for all its conceivable functional variants (e.g., auxiliary particle, conjunction / complementiser, modal particle). In other words, the lack of a precise categorial specification of this element does not affect generalisations about its verbal-clitic behaviour.

In an accentual unit hosted by the main verb, the conjunctive particle always occurs as the leftmost peripheral clitic and determines the overall proclitic nature of the clitic cluster—cf. (30).

(30)

a. *Šteše* ***[da*** *(ni gi)]___pokazvaš.*

FUT-AUX-imp,2sg [CONJ-PRT (DAT-1pl ACC-3pl)]_show-pres,2sg

'You would show them to us.'

b. *Šteše* ***[da*** *ne (ni gi)]___pokazvaš.*

FUT-AUX-imp,2sg [CONJ-PRT NEG-PRT (DAT-1pl ACC-3pl)]_show-pres,2sg

'You would not show them to us.'

c. *Šteše* ***[da*** *(ni gi e)]___pokazval.*

FUT-AUX-imp,2sg [CONJ-PRT (DAT-1pl ACC-3pl CL-AUX-3sg)]_show-pcp,sg,masc

'He would have shown them to us.'

d. *Šteše* ***[da*** *ne (ni gi e)]___pokazval.*

FUT-AUX-imp,2sg [CONJ-PRT NEG-PRT (DAT-1pl ACC-3pl CL-AUX-3sg)]_show-pcp,sg,masc

'He would not have shown them to us.'

When the host is a (non-clitic) auxiliary verb[47], the clitic sequence may optionally be discontinuous; in fact, (31a') and (31b') are marginal.

(31)

a. *Šteše* ***[da*** *(ni gi)]___bâde pokazal.*

FUT-AUX-imp,2sg [CONJ-PRT (DAT-1pl ACC-3pl)_BE2-3sg show-pcp,sg,masc

'He would have shown them to us.'

a'.? *Šteše* ***da___bâde___(ni gi)*** *pokazal.*

FUT-AUX-imp,2sg CONJ-PRT_BE2-3sg_(DAT-1pl ACC-3pl) show-pcp,sg,masc

---

[46] For example, some treat it as a conjunction, e.g., *Pašov 1989*, others as a particle *Hauge 1976* or AUX *Rudin 1986*.

[47] This auxiliary verb is, of course, different from *šta* and *njama*.

b. *Šteše* ***[da*** *ne (ni gi)]___bâde* *pokazal.*

FUT-AUX-imp,2sg [CONJ-PRT NEG-PRT (DAT-1pl ACC-3pl)_BE2-3sg show-pcp,sg,masc

'He would not have shown them to us.'

b'.? *Šteše* ***{da*** *ne}___bâde___(ni gi)* *pokazal.*

FUT-AUX-imp,2sg {CONJ-PRT NEG-PRT}_AUX-3sg_(DAT-1pl ACC-3pl) show-pcp,sg,masc

Although the negation particle *ne* bears a stress of its own in the syntactic domain of non-verbal constituents, as a verbal clitic it is never stressed and behaves as a proclitic with a notable prosodic peculiarity: the lexical item that immediately follows *ne* in the verb-complex constituent is always stressed, even if it happens to be phonologically weak—i.e., in accordance with the clitic LP-scheme, either a core clitic or *šte*.

In earlier transformational frameworks, both Rå Hauge[48] and Penčev[49] assume that the negation particle is primarily stressed in the deep structure, which is revealed on the surface if non-verbal constituents are negated, while in a negated verbal constituent, the stress shifts from the negation particle to the immediately following lexical element.

An essential trait of the process lies in the unchanged behaviour of the element that is immediately preceded by *ne* in the clitic sequence. Even though it bears a secondary accent, it never functions as a host in a prosodically autonomous accentual unit; this supplies additional evidence for regarding it as a clitic. The availability of a secondarily stressed clitic is nevertheless of significant relevance for positioning the interrogative particle *li* (discussed in the next section). The overall proclitic nature of a cluster that contains a secondarily stressed component as a result of *ne*-integration also remains constant. Let us consider some examples in (32)—the relevant data involve verb forms that do not contain *šta* or *njama*. If the negation particle is hosted by the full verb, then it is regularly included into the clitic cluster whenever one is formed.

(32)

a. ***[Ne*** *(vi gi)]___pokazvam.*

[NEG-PRT (DAT-1pl ACC-3pl)]_show-pres,1sg

'I am not showing them to you.'

---

[48] *Hauge 1976* states: "For the negation particle *ne*, we have to indicate that when it modifies a predicate, it always moves its stress over to the following word, also when this word is a clitic... *Ne* can modify other parts of the sentence apart from the predicate, but does not move its stress in such cases." (p.18)

[49] *Penčev 1984* states: "... after *ne*, every verbal element or pronominal clitic attains the stress of *ne*. The clitics (*mu, go, sâm*) have no stress of their own and the stress after *ne* is not theirs, therefore it does not prevent them from being clitics." (p.63)

b. *[**Ne** (sâm vi gi)]___pokazval.*
[NEG-PRT (CL-AUX-1sg DAT-1pl ACC-3pl)]_show-pcp,sg,masc
'I have not shown them to you.'

c. *Štjax [da **ne** (vi gi)]___pokazvam.*
FUT-AUX-imp,1sg [CONJ-PRT NEG-PRT (DAT-1pl ACC-3pl)]_show-pres,1sg
'I would not show them to you.'

d. *Štjax [da **ne** (sâm vi gi)]___pokazval.*
FUT-AUX-imp,1sg [CONJ-PRT NEG-PRT (CL-AUX-1sg DAT-1pl ACC-3pl)]_show-pcp,sg,masc
'I would not have shown them to you.'

e. *[**Ne** šte (si ni gi)]___dal.*
[NEG-PRT FUT-PRT (CL-AUX-2sg DAT-1pl ACC-3pl)]_give-pcp,sg,masc
'You will not have given them to us.'

If *ne* is hosted by a non-clitic auxiliary verb, then inclusion of the negation particle into the clitic cluster is optional. The case of *ne* and *šte* forming a proclitic group and optionally occurring detached from the core by the host (*bâda*) has already been illustrated in (29).

(33)

a. *[**Ne** (vi ja)]___bjax pokazval.*
[NEG-PRT (DAT-2pl ACC-3sg,fem)]_BE1-imp,1sg show-pcp,sg,masc
'I had not shown her to you.'

a'. ***Ne**___bjax___(vi ja) pokazval.*
NEG-PRT_BE1-imp,1sg_(DAT-2pl ACC-3sg,fem) show-pcp,sg,masc

b. *[**Ne** (vi ja)]___bil pokazval.*
[NEG-PRT (DAT-2pl ACC-3sg,fem)]_BE1-pcp,sg,masc show-pcp,sg,masc
'He has not shown her to you.' - reported

b'. ***Ne**___bil___(vi ja) pokazval.*
NEG-PRT_BE1-pcp,sg,masc_(DAT-2pl ACC-3sg,fem) show-pcp,sg,masc

c. *[**Ne** (mu)]___bila predstavena.*
[NEG-PRT (DAT-3sg,masc)]_BE1-pcp,sg,fem introduce-pass-pcp,sg,fem
'She was not introduced to him.' - reported

c'. ***Ne**___bila___mu predstavena.*
NEG-PRT_BE1-pcp,sg,fem_DAT-3sg,masc introduce-pass-pcp,sg,fem

It is worth noting that the reported verb forms in (33b-c) are in the 3rd person, and hence do not contain a clitic auxiliary verb. With the corresponding 1st and 2nd person verb forms, a clitic auxiliary verb is available in the accentual unit, and therefore *ne* is preferably included in

the clitic cluster in (34a-b), since otherwise the AuxC headed by the clitic auxiliary would become discontinuous, as in (34a'-b').

(34)

a. *[**Ne** (sâm vi ja)]___bil pokazval.*

[NEG-PRT (CL-AUX-1sg DAT-2pl ACC-3sg,fem)]_BE1-pcp,sg,masc show-pcp,sg,masc

'I have not shown her to you.' - reported

a'.*? **Ne**___bil___(sâm vi ja) pokazval.*

NEG-PRT_BE1-pcp,sg,masc_(CL-AUX-1sg DAT-2pl ACC-3sg,fem) show-pcp,sg,masc

b. *[**Ne** (si im ja)]___bil pokazval.*

[NEG-PRT (CL-AUX-2sg DAT-3pl ACC-3sg,fem)]_BE1-pcp,sg,masc show-pcp,sg,masc

'You have not shown her to them.' - reported

b'.*? **Ne**___bil___(si im ja) pokazval.*

NEG-PRT_BE1-pcp,sg,masc_(CL-AUX-2sg DAT-3pl ACC-3sg,fem) show-pcp,sg,masc

3.2.2.4 Contribution of the Enclitic Proper

The position of the enclitic proper *li*, a particle marking interrogation, depends basically on the type of the modified (questioned) constituent[50]. As a verbal clitic, the interrogative particle immediately follows **the first** (i.e. leftmost) **stressed** (primarily or secondarily) element of the verb complex. In (35) all stressed elements are given in SMALL CAPS:

(35)

a. *Šte ste ni ja POKAZALI **li** navreme?*

FUT-PRT CL-AUX-2pl DAT-1pl ACC-3sg,fem show-pcp,pl Q in-time

'Will you have shown her to us in time?'

b. *POKAZALI **li** ste ni ja?*

show-pcp,pl Q CL-AUX-2pl DAT-1pl ACC-3sg,fem

'Have you shown her to us?'

c. *ŠTELI **li** ste da ni ja BÂDETE POKAZALI?*

FUT-AUX-pcp,pl Q CL-AUX-2pl CONJ-PRT DAT-1pl ACC-3sg,fem BE2-2pl show-pcp,pl

'Would you have shown her to us?' - reported

d. *Ne STE **li** ni ja POKAZALI?*

NEG-PRT CL-AUX-2pl Q DAT-1pl ACC-3sg,fem show-pcp,pl

'Haven't you shown her to us?'

---

[50] *Hauge 1976* states: "... *li* follows all non-verbal constituents modified by it, while it is in second position only if it modifies the V constituent." (p.22)

e. *Ne VI **li** ja BJAX POKAZAL?*

NEG-PRT DAT-2pl Q ACC-3sg,fem BE1-imp,1sg show-pcp,sg,masc

'Hadn't I shown her to you?'

f. *Ne BJAX **li** vi ja POKAZAL?*

NEG-PRT BE1-imp,1sg Q DAT-2pl ACC-3sg,fem show-pcp,sg,masc

'Hadn't I shown her to you?'

If the negation particle occupies the leftmost position in a verb complex, the immediately following word will be the first stressed element, and will possibly bear only secondary stress as, e.g., *ste* in (35d) and *vi* in (35e). So, if this word happens to be a clitic, then *li* must immediately follow this clitic, and is consequenly included into the clitic cluster. It is important to note that the relative order of the other clitics in the cluster remains unchanged, as illustrated also in (36), where the clitic clusters are correspondingly marked.[51]

(36)

a. *[Ne sâm **li** vi ja]___pokazval?*

[NEG-PRT CL-AUX-1sg Q DAT-2pl ACC-3sg,fem]_show-pcp,sg,masc

'Haven't I shown her to you?'

b. *[Ne šte **li** ni]___kažeš?*[52]

[NEG-PRT FUT-PRT Q DAT-1pl]_tell-pres,2sg

'Won't you tell us?'

The question particle is the only peripheral verbal enclitic. In order for *li* to make the overall nature of the clitic cluster enclitic, it must be the leftmost peripheral clitic attached to the core. This condition is met only if no proclitics proper are involved in the accentual unit. In (24b) which is repeated here as (37), prosodically autonomous lexical material precedes the verb complex, which helps us to ascertain that the clitic cluster is enclitic.[53]

---

[51] This is also pointed out by Rå Hauge: "The relative order of the other clitic elements remains the same, with *li* interlocated in the second slot after *ne*." *Hauge 1976* (p.20)

[52] A more frequently used wording for (36b) is, of course, *Njama li da ni kažeš?*

[53] A remark is needed at this point. It might seem that such an enclitic cluster precedes the verb host only if *li* is semantically and intonationally linked to the preceding word, rather than to the verb, e.g.:

(i) *Sestrata **li** ni go e pokazvala?*

In fact, the prosodic articulation of this example excludes the question particle from the verb complex:

(ii) *Sestrata **li** (ni go e) pokazvala?*

'Was it the nurse who has shown him to us?'

(37)
*Sestrata pokazvala___[**li** (ni go e)]?*
nurse show-pcp,sg,fem_[Q (DAT-1pl ACC-3sg,masc CL-AUX-3sg)]
'Has the nurse shown him to us?'

Obviously, the cluster here is enclitic not as a result of clause-initial position of the verb complex, since the verb complex is not in this position, but rather as a result of the interaction between *li* and the core.

3.2.3 Aspects of Formalisation

The external behaviour of a cluster is determined by the syntactic and prosodic specifications of the clitics forming it. Therefore, it is necessary to ensure that the prosodic and syntactic requirements of a group of clitics are being composed or unified as they are combined with one another, so that the resulting cluster as a whole inherits the properties of its parts. This is exactly the case of the core-periphery distinction and the clitic interaction described above.

The **syntactic specification** of a clitic amounts to an indication of its syntactic domain. In order for two clitics to combine into a cluster, they must select the same domain. Specifically, as it is assumed in, e.g., *Halpern 1995*, the clitics select ("subcategorise for") the syntactic category to which they may attach, and in order for two clitics to combine into a cluster their syntactic selectional requirements must be able to unify. This means, among other things, that clitics with distinct domains are never combined into a single cluster, even if they are coincidentally adjacent.

The **prosodic specification** of a clitic indicates whether it requires a prosodic constituent of a certain type to its right or to its left. *Halpern 1995* (p.215) also discusses some Bulgarian data in terms of prosodic subcategorisation but his results are inadequate since they are based on a incorrect interpretation of the language material. Nevertheless, the basic idea of prosodic selection can be maintained, and I shall consider it in the following.

Let '~X~' indicate that the direction of the prosodic selection is not determined by the clitic X. These are the so-called endoclitics, or movable

---

Therefore, the clitic sequence can be split by an intervening sentential constituent:
(iii) *Sestrata **li** po-rano (ni go e) pokazvala?*
'Was it the nurse who has shown him to us before?'
It is important to notice that *li* is not a **verbal clitic** here, neither intonationally nor semantically. Otherwise, all of the verbal clitics in this example behave as expected.

clitics, or—as assumed in this study—core verbal clitics. At the same time, let us represent the proclitics proper as 'X_' and enclitics proper as '_X', with the underscore indicating the predetermined direction (rightwards or leftwards) of the selection of a prosodic host. Then the various clitic groupings distinguished in the previous sections can be represented as resulting from the following compositions of prosodic specifications of the involved clitics:

- **core clitic cluster**: [~X~] + [~Y~] = [~XY~]

(38)

a. *Dadox___(mu go).*

gave-1sg DAT-3sg,masc ACC-3sg,neut

'I gave it to him.'

b. *Az (mu go)___dadox.*

I DAT-3sg,masc ACC-3sg,neut gave-1sg

- **enclitic cluster** : [_X] + [~Y~] = [_XY]

(39)

Dade___[li (mu go)]?

gave-2sg Q DAT-3sg,masc ACC-3sg,neut

'Did you give it to him?'

- **proclitic group**: [X_] + [Y_] = [XY_]

(40)

{*Ne šte*}*___dojde.*

NEG-PRT FUT-PRT come-3sg

'She / he will not come.'

- **proclitic cluster**: [X_] + [~Y~] = [XY_]

(41)

[Šte (mu go)]___dam.

FUT-PRT DAT-3sg,masc ACC-3sg,neut give-1sg

'I shall give it to him.'

Since there is only one enclitic proper (*li*) among the verbal clitics in Bulgarian, no enclitic group is attested within the verb complex. A composition of enclitics resulting in an enclitic grouping is possible within the Bulgarian NP[54]:

[54]It is important to keep in mind that the possessive clitic is a prototypical enclitic within the NP, while the homonymous dative clitic is a core (i.e. "movable") clitic within the verb complex.

[_X] + [_Y] = [_XY]

(42)

Prijatelkata___<mu li> dojde?

girl-friend_def.art his-POSS-CL Q came-3sg

'Was it his girl-friend who came?'

Although theoretically possible, the following composition does not occur in Bulgarian:

* [X_] + [_Y] = [XY]

This is supposed to be a prosodically independent combination obtained by "cancelling" the prosodic subcategorisation of a proclitic and an enclitic. Historically this might however be how Bulgarian tag-question particle *nali,* which is a prosodically strong lexical item, has emerged.

3.2.4 Summary and Conclusions

In the analysis proposed here, it is assumed that the actual positioning of the verbal clitics (or clitic clusters) in modern Bulgarian is determined by the interplay of syntactic and prosodic factors.

Syntactically, the verbal clitics in Bulgarian belong to the verb-complex constituent and are subject to prosodic constraints within this syntactic domain. The clause-initial restriction and the quasi-second-position condition, as well as the placement of the question particle, are directly related to the prosodic organisation of the verb-complex constituent into accentual units (prosodic words). The syntactic assumption that the verb complex has an AuxC sub-constituent offers a natural explanation of certain prosodically inconsistent cases.

In the proposed classification, each verbal clitic is specified as being core or peripheral on the basis of predominantly prosodic criteria. Unlike peripheral clitics, core clitics appear to be obligatory components of clitic clusters. It has been demonstrated how particular clitic-cluster placements in the verb complex can be attributed to the interaction of different clitic types in conformity to more general constraints.

3.3 The Structure of the Verb Complex

In the adopted type of binary branching structures, the verb complex (VC), which is an intermediate construct between the lexical "level" of words (i.e.

synthetic word-forms) and the "level" of the syntax proper, will be represented as sketched in Figure 18. The thick lines in the structures below mark the respective nucleus head projections.

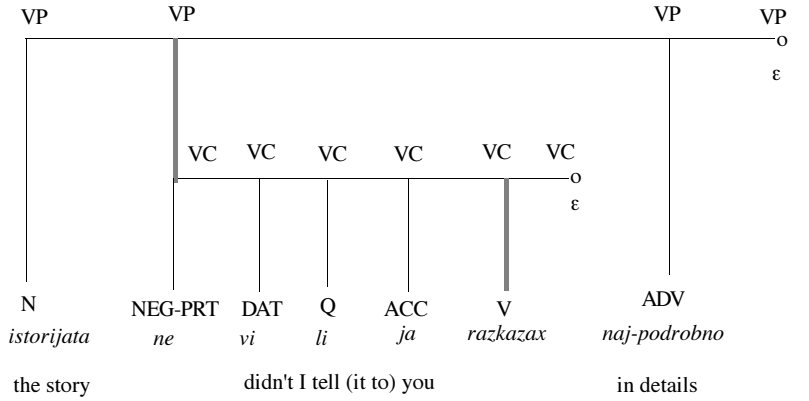


Figure 18

In the following discussion, I shall tentatively distinguish between four types of constructions according to the complexity of the involved verb form.

Compact Verb Complexes are generally based on synthetic or analytic verb forms *not involving the conjunctive particle*; such are:

- morphosyntactic verb complexes based on *synthetic* verb forms with no lexical or experiencer formants (first type);
- morphosyntactic verb complexes based on *analytic* verb forms involving no auxiliary verbs (second type);
- morphosyntactic verb complexes based on *analytic* verb forms involving at least one auxiliary verb (third type).

Composite Verb Complexes are morphosyntactic verb complexes based on *analytic* verb forms *involving the conjunctive particle* (fourth type).

The notion of an auxiliary morphosyntactic complex (AuxC) employed in the current analysis refers to an immediate constituent of the verb complex. The underlying concept is that of a marker phrase headed by an auxiliary verb. Inasmuch as marking by auxiliary verbs is not a syntactic but rather a morphosyntactic phenomenon, auxiliary verbs may form—and respectively head—only morphosyntactic constituents (AuxC's). An

auxiliary complex can contain merely functional and clitic elements: form-building or other (i.e. negative and interrogative) particles, an auxiliary verb marking the heading one, clitic pronouns, etc.—cf. Figure 19.

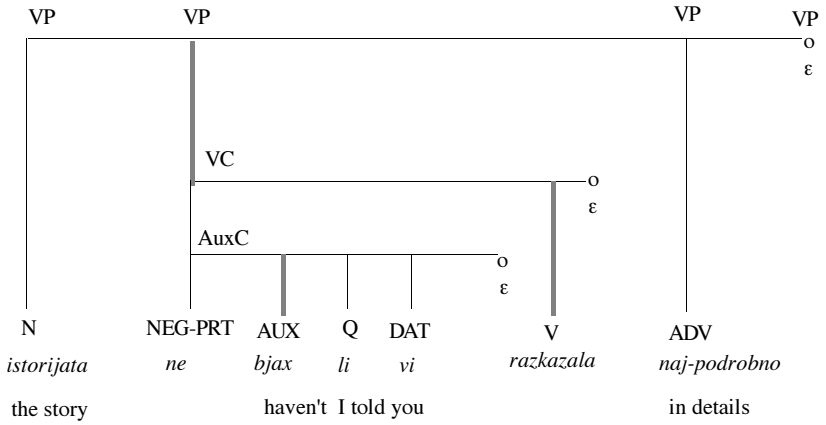


Figure 19

3.3.1 Compact Verb Complexes

No material can be extracted from within a compact verb complex and, in general, no sentential constituents can be inserted into it[55]. Only one clitic cluster can be formed in compact verb complexes. Formants, clitic pronouns, particles and clitic auxiliaries are possible clitic-cluster components. The prosodic aspect of their placement and the observed clustering effects have already been discussed in Section 3.2.2.

[55]An exception to this might be considered the marginal occurrence of a "short" adverbial (always semantically related to the way (or time) in which the action denoted by the main verb is performed) inside a compact verb complex containing an auxiliary verb and a main-verb participle:

(i) ***Šte sâm ti ja*** *veče* ***dal***.
FUT-PRT CL-AUX-1sg DAT-2sg ACC-3sg,fem already give-pcp,sg,masc
'I shall have already given it to you.'

(ii) ***Ne ste mu bili*** *navreme* ***predstaveni***.
NEG-PRT CL-AUX-2pl DAT3-sg,masc BE1-pcp,pl in time introduce-pass-pcp,pl
'You were not introduced to him in time.'

3.3.1.1 Verb Complexes Based on Synthetic Verb Forms

A compact verb complex including the main verb as the exclusive component of the verb form is structurally simple. It optionally contains personal or reflexive clitic pronouns, which in this case are to be licensed by the valence requirements of the particular verb and therefore considered object pronominal clitics rather than lexical or morphological formants, and the particles *ne* and *li* (43). In these simple constructions, the contingent clitic elements always form a cluster with the notable exception of the case where *ne* and *li* cooccur but no other clitics are involved (44): the cluster is not formed, due to the impossibility for these particles to stand adjacently, as discussed in Section 3.2.2.

(43)

a. *Istorijata* ***vi ja razkazax*** *naj-podrobno.*

story_def.art DAT-2pl ACC-3sg,fem tell-aor,1sg most-detailed

'I told you the story in details.'

a. ***Razkazax vi ja.***

tell-aor,1sg DAT-2pl ACC-3sg,fem

'I told it (the story) to you.'

b. ***Razkazax li vi ja****?*

tell-aor,1sg Q DAT-2pl ACC-3sg,fem

'Did I tell it to you?'

c. ***Ne vi li ja razkazax****?*

NEG-PRT DAT-2pl Q ACC-3sg,fem tell-aor,1sg

'Didn't I tell it to you?'

(44)

a. *Roditelite* ***ne davat li*** *bonboni na decata?*

parents_def.art NEG-PRT give-pres,3pl Q candies to children_def.art

'Don't the parents give candies to the children?'

b. ***Ne vali li*** *vân?*

NEG-PRT rain-pres,impers. Q outdoors

'Isn't it raining outdoors?'

The discussed simplest type of a compact verb complex can be based on any of the following synthetic (i.e. one-word) verb forms provided these are available in the paradigm of the chosen verb lexeme. Let us take for the sake of illustration the synthetic verb forms of the imperfective transitive verb *četa* 'to read':

- active indicative non-reported:

present (*četa*), aorist (*četox*), imperfect (*četjax*)
- 3rd person active indicative reported:
  present or imperfect (*četjal, četeli*), aorist (*čel, čeli*)
- 2nd person imperative (*četi, četete*)
- shortened infinitive (*čete*)
- synthetic conditional (? *četvam*)

3.3.1.2 Verb Complexes Based on Analytic Verb Forms Involving No Auxiliary Verbs

Let us consider now compact verb complexes based on analytic verb forms involving no auxiliary verbs. As in the previous case, all contained lexical elements are *immediate* constituents of the verb complex.

The analyticity of the verb forms has two different "sources". The first is the analyticity of the verb lexeme due to an obligatory lexical (as in *smeja se* 'to laugh') or experiencer (as in *trese me* 'I shiver') formant. The second is the morphological analyticity of verb forms containing only form-building particles: the reflexive-passive particle *se*, the future-tense particle *šte* and *da* as a modal particle introducing analytic imperative forms—all these are illustrated in (45).

(45)

a. *Tazi kniga* ***se čete*** *ot vseki lingvist.* 'This book is read by every linguist.'

b. *Vseki lingvist* ***šte čete*** *tazi kniga.* 'Every linguist will read this book.'

c. *Vseki lingvist* ***da čete*** *tazi kniga.* 'Let every linguist read this book.'

Pronominal clitics—in accordance with the valence of the verb—as well as the negative and the interrogative particles can certainly occur also in this type of construction.

3.3.1.3 Verb Complexes Based on Analytic Verb Forms Involving at least One Auxiliary Verb but No Conjunctive Particle

Two types of auxiliary verbs can participate in compact verb complexes[56]:

- ***sâm***—in its clitic present-tense forms which I refer to in the glosses of the examples as CL-AUX, as well as in its non-clitic imperfect (*bjax*), aorist (*bix*) and participle (*bil*) forms referred to as BE1 with the respective specification (imp, aor or pcp);

[56]When the negative imperative auxiliary *nedej / nedejte* is combined with the shortened infinitive of the main verb, it can also occur in this type of verb complex, e.g., *nedej / nedejte otiva* 'don't go'.

- ***bâda***—with only non-clitic present-tense (*bâda*) and participle (*bâdel*) forms referred to in the glosses of the examples as BE2 with the respective form specification (pres or pcp).

For capturing generalisations in the following discussion, an auxiliary complex headed by one of these auxiliaries will be referred to as be-auxiliary morphosyntactic complex (or: be-AuxC).

Let us consider some examples illustrating the syntactic properties of the be-AuxC. For the whole verb complex to be grammatical, the pronominal clitics[57] have to stay adjacently to the auxiliary verb. Any separating them from the auxiliary verb, as in (46d-e) or in (47d), even if this is not combined with a violation of the quasi-second-position condition (which is the case of, e.g., (46d)), would render the wording unacceptable. The negative particle always joins the AuxC—cf. the contrast between (47a-b), on the one hand, and the ungrammatical (47c), on the other hand. In the latter case where the prosodic requirements of the interrogative particle are formally satisfied—*li* immediately follows the first stressed element in this verb complex, the unacceptability results mainly from the fact that the negative particle is not included in the be-AuxC.

(46)

a. ... *mi ja bjaxa pokazali*.
DAT-1sg ACC-3sg,fem BE1-3pl show-pcp,pl
'They have shown her to me.'

b. *Bjaxa mi ja pokazali*.
BE1-3pl DAT-1sg ACC-3sg,fem show-pcp,pl

c. *Pokazali mi ja bjaxa*.
show-pcp,pl DAT-1sg ACC-3sg,fem BE1-3pl

d. *... *mi ja pokazali bjaxa*.
DAT-1sg ACC-3sg,fem show-pcp,pl BE1-3pl

e. **Bjaxa pokazali mi ja*.
BE1-3pl show-pcp,pl DAT-1sg ACC-3sg,fem

(47)

a. *Ne ti li ja bjaxa pokazali?*
NEG-PRT DAT-2sg Q ACC-3sg,fem BE1-3pl show-pcp,pl
'Haven't they shown her to you?'

b. *Ne bjaxa li ti ja pokazali?*
NEG-PRT BE1-3pl Q DAT-2sg ACC-3sg,fem show-pcp,pl

[57]The contingent lexical and experiencer formants behave syntactically in a parallel fashion.

c. *<u>*Ne*</u> *pokazali* <u>*li ti ja bjaxa?*</u>
NEG-PRT show-pcp,pl Q DAT2-sg ACC-3sg,fem BE1-3pl

d. *<u>*Ne bjaxa li*</u> *pokazali* <u>*ti ja?*</u>
NEG-PRT BE1-3pl Q show-pcp,pl DAT-2sg ACC-3sg,fem

A be-AuxC, which functions as a marker in the verb complex, can itself be marked by another clitic auxiliary verb of the *sâm*-type or—provided further clitic elements occur—even by an entire be-AuxC headed by this auxiliary. In the reported paradigm the embedded clitic auxiliary marker is systematically dropped in the third person, so for illustrating the discussed structure we have to take non-third-person reported verb forms, as in (48a-b), or certain (more elaborate) indicative participial-passive verb forms, as those exemplified in (48c-d).[58] The corresponding structure of the matrix be-AuxC will be shaped as sketched in Figure 20; the auxiliary verb heading it is in participle form.

(48)

a. <u>*Ne sme li ti ja* ***bili***</u> *pokazali?*
NEG-PRT CL-AUX-1pl Q DAT-2sg ACC-3sg,fem BE1-pcp,pl show-pcp,pl
'Haven't we shown her to you?' - reported

b. *Pokazali* <u>*ste im ja* ***bili***</u>.
show-pcp,pl CL-AUX-2pl DAT-3pl ACC-3sg,fem BE1-pcp,pl
'You have shown her to them.' - reported

c. <u>*Ne im li e* ***bila***</u> *predstavena?*
NEG-PRT DAT-3pl Q CL-AUX-3sg BE1-pcp,sg,fem introduce-pass-pcp,sg,fem
'Wasn't she introduced to them?'

d. <u>*Šte sme im* ***bâdeli***</u> *predstaveni*.
FUT-PRT CL-AUX1-pl DAT-3pl Q BE2-pcp,pl, introduce-pass-pcp,pl
'We shall have been introduced to them.' - reported

In Figure 20 the be-AuxC [*ne im li e*] functions as a morphosyntactic marker of the auxiliary participle *bila*, and the whole be-AuxC [[*ne im li e*] *bila*] is a morphosyntactic marker with respect to the main-verb passive participle *predstavena*. It is further important to notice that the dative pronominal clitic *im* and the particles *ne* and *li* combine in a morphosyntactic unit with the "lowest" be-auxiliary.

[58]The auxiliary verb heading the be-AuxC is set **bold**.

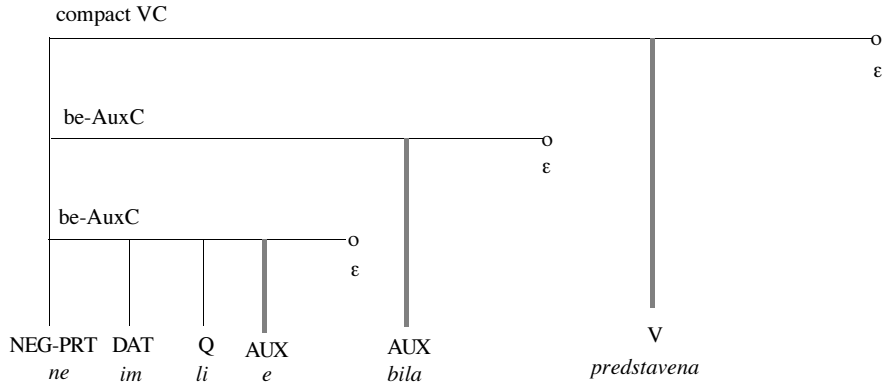


Figure 20

Even though strict prosodic constraints are to be observed, and certain scope effects with respect to alignments containing negation to be taken into consideration[59], one could basically assume that the relative order of the auxiliary complex and the main verb in compact verb complexes is *syntactically* free.

3.3.2 Composite Verb Complexes

A composite verb complex—cf. Figure 21—generally consists of two morphosyntactic sub-complexes: a detached auxiliary verb or even a detached AuxC headed by that verb, and a main-verb complex obligatorily marked by the conjunctive particle *da*. The latter will also be referred to as *da*-complex.

Figure 21

[59]A be-AuxC containing the negative particle must linearly precede the main verb—cf. the unacceptability of, e.g.,

**Sutrinta ošte pokazali ne ti ja bjaxa*

this morning still show-pcp,pl NEG-PRT DAT-2sg ACC-3sg,fem BE1-3pl

Unlike the situation with compact verb complexes, any number of sentential constituents can intervene between these two parts; also the formation of two clitic clusters is possible—one appurtenant to the detached AuxC and potentially covering the interrogative particle and a clitic auxiliary, the other one similar to the clitic cluster appropriate for the compact verb complexes and possibly containing formants, clitic pronouns, particles and a clitic auxiliary. In the ordinary case, the detached AuxC linearly precedes the *da*-complex.

In the table below, any detached AuxC can legitimately combine with any main-verb complex. For the sake of illustration, the detached AuxC's are given in their most extended version. The auxiliary participles in the parentheses occur in strongly (emphatically) reported verb forms.

detached AuxC < ... >	CONJ -PRT	main-verb complex
Štjaxte li FUT-AUX-2pl Q		*mi gi dadete?* DAT-1sg ACC-3pl give-2pl
Njamaše li NEG-FUT-AUX-3sg Q	*da*	*ste mi gi dali?* CL-AUX-2pl DAT-1sg ACC-3pl give-pcp,pl
Šteli li ste (bili) FUT-AUX-pcp,pl Q CL-AUX-2pl (BE1-pcp,pl) - reported		*mi gi bâdete predstavili?* DAT-1sg ACC-3pl BE2-2pl introduce-pcp,pl
Njamalo li (bilo) NEG-FUT-AUX-pcp,sg,neut Q (BE1-pcp, sg, neut) - reported		*ste mi bili predstaveni?* CL-AUX-2pl DAT-1sg BE1-pcp,pl introduce-pass-pcp,pl

3.3.2.1 Detached Auxiliary Complex

The detached AuxC minimally covers the head auxiliary verb, which can be:

- the positive future auxiliary *štjax* (49a),
- the impersonal negative future auxiliary *njama* (49b), or
- the negative imperative auxiliary *nedej* (49c) which has only second-person forms.

Inasmuch as the latter auxiliary verb (unlike the former two—cf. (50-51)) would never create an auxiliary complex, i.e. it is always a lexical constituent of a composite verb complex, in the following discussion, I shall

refer to the detached auxiliary complex as the future-auxiliary morpho-syntactic complex or fut-AuxC.

(49)

a. *Štjaxte* *da mi gi pokazvate*.
FUT-AUX-2pl CONJ-PRT DAT-1sg ACC-3pl show-2pl
'You would show them to me.'

b. *Njama* *da mi gi pokazvate*.
NEG-FUT-AUX-impers,fin CONJ-PRT DAT-1sg ACC-3pl show-2pl
'You would not show them to me.'

c. *Nedejte* *da mi gi pokazvate*.
NEG-IMPER-2pl CONJ-PRT DAT-1sg ACC-3pl show-2pl
'Don't show them to me.'

(50)

a. ***Šteli ste bili*** *dnes* *vie* ***da mi gi pokazvate***.
FUT-AUX-pcp,pl CL-AUX-2pl BE1-pcp,pl today you CONJ-PRT DAT-1sg ACC-3pl show-2pl
'(As I have heard) you will show them to me today.' - emphatic reported

b. *Dnes* ***ste bili šteli*** *vie* ***da mi gi pokazvate***.
today CL-AUX-2pl BE1-pcp,pl FUT-AUX-pcp,pl you CONJ-PRT DAT-1sg ACC-3pl show-2pl

c. *Dnes* ***šteli li ste bili*** *vie* ***da mi gi pokazvate****?*
today FUT-AUX-pcp,pl Q CL-AUX-2pl BE1-pcp,pl you CONJ-PRT DAT-1sg ACC-3pl show-2pl
'Will you show them to me today (as I have heard).' - emphatic reported

d. *?Dnes* ***bili li ste šteli*** *vie* ***da mi gi pokazvate****?*
today BE1-pcp,pl Q CL-AUX-2pl FUT-AUX-pcp,pl you CONJ-PRT DAT-1sg ACC-3pl show-2pl
'(As I have heard) you will show them to me today.' - emphatic reported

(51)

a. *Dnes* ***njamalo bilo*** *vie* ***da mi gi pokazvate***.
today NEG-FUT-AUX-impers,pcp BE1-pcp,sg,neut you CONJ-PRT DAT-1sg ACC-3pl show-2pl
'(As I have heard) you will not show them to me today.' - emphatic reported

b. **Dnes* ***bilo njamalo*** *vie* ***da mi gi pokazvate***.
today BE1-pcp,sg,neut NEG-FUT-AUX-impers,pcp you CONJ-PRT DAT-1sg ACC-3pl show-2pl

c. *Dnes* ***njamalo li bilo*** *vie* ***da mi gi pokazvate***.
today NEG-FUT-AUX-impers,pcp Q BE1-pcp,sg,neut you CONJ-PRT DAT-1sg ACC-3pl show-2pl
'Won't you show them to me today (as I have heard).' - emphatic reported

The prosodic constraints on the positioning of the clitic elements also in the fut-AuxC have already been described in Section 3.2.2; the unacceptability of the alignment in (51b) results, however, exclusively from the requirement for negation, which, in this case, is inherently contained in the negative

future-auxiliary verb, to be initial in the verb complex over which it scopes. The internal structure of a fut-AuxC marked by a be-AuxC will be shaped as sketched in Figure 22; this most extended variant of a detached AuxC can be headed only by a positive future-auxiliary verb.

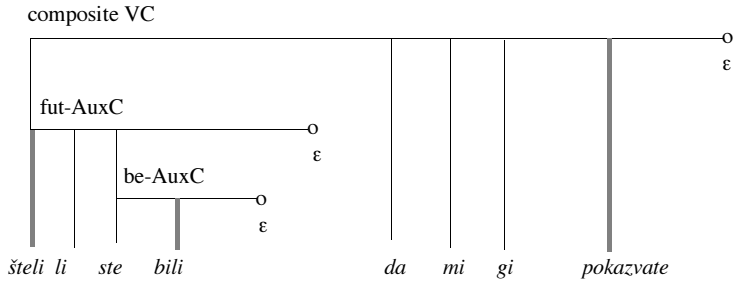


Figure 22

.c.3.3.2.2 Da-Complex

The *da*-marked sub-constituent in composite verb complexes minimally covers the conjunctive particle and the main verb. Importantly, it is in this part of composite verb complexes where the contingent formants and clitic pronouns are incorporated. The proclitic conjunctive particle is a morphosyntactic marker which is expandable only from the VC-level of the respective composite verb complex, and thus, structurally, it would never be an immediate constituent of any auxiliary complex.

The regularities within the main-verb complex marked by *da* are basically parallel to those observable in non-sentence-initial compact verb complexes—cf. (52). The inventory of the forms of the involved be-auxiliary verbs is, however, more restricted: the respective aorist and imperfect forms never occur. Inasmuch as the future particle is in complementary distribution with the fut-AuxC, it never occurs in composite verb complexes either. In certain rather rare cases, the negative particle may occur in the *da*-complex, although the parallel variants of the respective verb complex employing the negative future-auxiliary verb are definitely preferable.

(52)

a. *(Štjaxte / njama / šteli ste / njamalo / nedejte) da ni gi predstavjate*.

(fut-AuxC / NEG-IMPER-2pl) CONJ-PRT DAT-1pl ACC-3pl introduce-pres,2pl

b. *(Štjaxte / njama / šteli ste / njamalo) da ste ni gi predstavili.*
(fut-AuxC) CONJ-PRT CL-AUX-2pl DAT-1pl ACC-3pl introduce-pcp,pl

c. *(Štjaxte / njama / šteli ste / njamalo) da im bâdete predstaveni.*
(fut-AuxC) CONJ-PRT DAT-3pl BE2-2pl introduce-pass-pcp,pl

d. *(Štjaxte / njama / šteli ste / njamalo) da ste im bili predstaveni.*
(fut-AuxC) CONJ-PRT CL-AUX-2pl DAT-3pl BE1-pcp,pl introduce-pass-pcp,pl

e. *(Šteli ste / njamalo) da ste ni gi bili predstavili.*
(fut-AuxC) CONJ-PRT CL-AUX-2pl DAT-1pl ACC-3pl BE1-pcp,pl introduce-pcp,pl

f. *(Šteli ste / njamalo) da ste im bâdeli predstaveni.*
(fut-AuxC) CONJ-PRT CL-AUX-2pl DAT-3pl BE2-pcp,pl introduce-pass-pcp,pl

When the *da*-complex does not contain any auxiliary, all lexical items are immediate constituents of the verb complex—cf. Figure 22; but a be-AuxC of a high complexity may also occur as a VC-sub-constituent, as illustrated structurally in Figure 23.

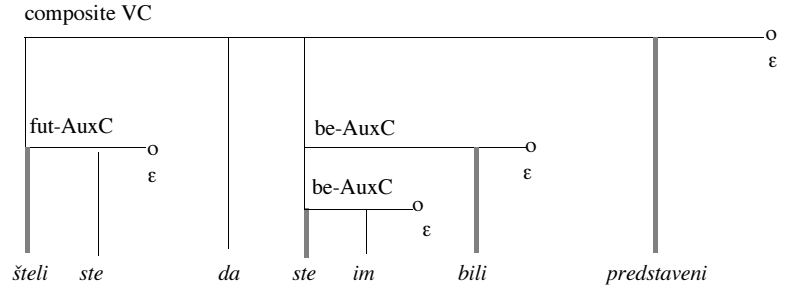


Figure 23

3.3.2.3 Order Variants

In Figure 24 the most extended structural variant of a composite verb complex is represented: the detached auxiliary complex and the be-auxiliary complex—both functioning as morphosyntactic markers of the main verb—are themselves marked by a be-AuxC.

A remarkable property of the constructions discussed here is that other sentential constituents may intervene between the two sub-constituents of the composite verb complex, provided, however, these sentential constituents are not headed by a verb.

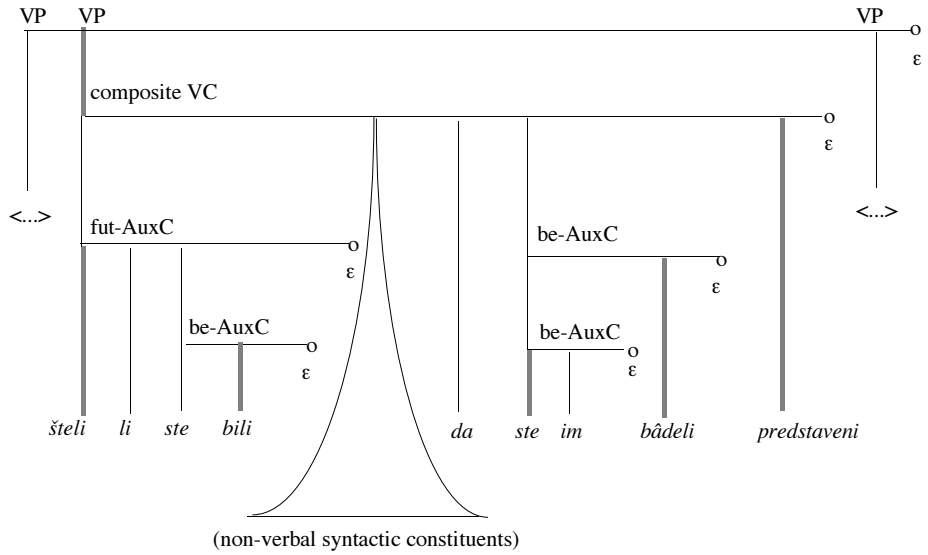


Figure 24

In the default ordering of the components of a composite verb complex, the detached auxiliary complex precedes the *da*-complex. Variants with a positive future-auxiliary complex following the *da*-complex are exceptionally possible, although they are often reported as only marginally acceptable—cf. (53).

As can be observed, an initial *da*-complex has to be relatively simple. For example, inverting a *da*-complex which contains an AuxC-sub-constituent, as illustrated in (53c-d), tends to be rather unacceptable.

(53)

a. ***Da im četa*** *na decata* ***sâm štjala*** *tazi prikazka.*

CONJ-PRT DAT-3pl read-1sg to children_def.art CL-AUX-1sg FUT-AUX-pcp,sg,fem this story

'I shall read this story to the children.' - reported

b. *?* ***Da im četa*** *na decata* ***štjax*** *tazi prikazka.*

CONJ-PRT DAT-3pl read-1sg to children_def.art FUT-AUX-1sg this story

'I would read this story to the children.'

c. * ***Da sâm im čela*** *na decata* ***sâm štjala*** *tazi prikazka.*

CONJ-PRT CL-AUX-1sg DAT-3pl read-pcp,sg,fem to children_def.art CL-AUX-1sg FUT-AUX-pcp,sg,fem this story

'I shall have read this story to the children.' - reported

d. * ***Da sâm im čela*** *na decata* ***štjax*** *tazi prikazka.*

CONJ-PRT CL-AUX-1sg DAT-3pl read-pcp,sg,fem to children_def.art FUT-AUX-1sg this story

'I would have read this story to the children.'

Due to the fact that negation has to be verb-complex-initial, no inverted variants with the negative future-auxiliary verb (54a) or the negative imperative auxiliary verb (54b) are possible.

(54)

a. ***Da im četa*** *na decata* ***njama*** *tazi prikazka.*

CONJ-PRT DAT-3pl read-1sg to children_def.art NEG-FUT-AUX-impers this story

'I shall not read this story to the children.'

b. ***Da im četeš*** *na decata* ***nedej*** *tazi prikazka.*

CONJ-PRT DAT-3pl read-2sg to children_def.art NEG-IMPER-2sg this story

'Don't read this story to the children.'

3.3.3 Summary and Conclusions

Based on the structural complexity and syntactic behaviour, two main types of verb complexes are distinguished in the proposed analysis. Compact verb complexes are characterised by strict adjacency of their components, though the mutual order of the latter is not syntactically fixed and also depends on prosodic factors, provided verbal clitics are involved. On the other hand, composite verb complexes have two loosely bound parts which usually do but need not stand in immediate adjacency, so that further sentential constituents may intervene.

Introducing the morphosyntactic domain of verb and auxiliary complexes in the description of Bulgarian allows for a comprehensible and linguistically adequate representation of structural relations from to the interface area between lexicon / morphology and clausal syntax.

3.4 The Morphosyntactic Aspect of Replication Phenomena

Although the realisation of the obligatory pronominal experiencer formant with phrasal experiencer verbs differs in nature from the replication of

determined nominal material by a pronominal verbal clitic[60], these are formally very similar phenomena.

In Section 3.1.3 I discussed the evidence for assuming the existence of an *experiencer agreement* in Bulgarian, i.e. a special type of concord between the experiencer-object, which can syntactically be dative or accusative, and the verbal head containing the respective formant in the form of a pronominal clitic. Let us now turn our attention to the morphosyntactic aspect of the *replication* phenomena observable in Bulgarian.

The main difference is that the involved pronominal clitics are no longer obligatorily required by the verbal lexeme, therefore, they cannot be regarded as formants. On the one hand, they are valid complements able to satisfy the respective valence requirements of the verb, and on the other hand, they may be just replicas of overtly realised full-fledged NP-complements, exhibiting then mainly thematicising and, on certain conditions, also object-identifying function with respect to the nominal material they replicate (cf. Section 4.3).

Traditionally, clitic pronouns are analysed as parts of the sentence (i.e. sentential constituents). For example, in *Pokazax ti ja*. ('I showed her to you.') *ja* would be considered the direct object and *ti* the indirect object. But one might also analyse such a form as being marked for *object-verb agreement*, in a parallel fashion as the *subject-verb agreement* is marked in the verb inflection. So it can be assumed that the verbal ending *-x* indicates that the subject is first person singular, the verbal clitic *ti* that the indirect object is second person singular, and the verbal clitic *ja* that the direct object is third person singular feminine. In case of coordination, the clitics are kept with (hence, repeated on) each conjunct—again, quite similarly to the verb inflection that is expressed on each of the verbs—cf. (55).

(55)

*[Pokaza**x** **ti ja**]* *i* *[**ti ja** predstavi**x**]*.

showed-1sg DAT-2sg ACC3-sg,fem and DAT-2sg ACC-3sg,fem introduced-1sg

'I showed and introduced her to you.'

When some nominal material has been replicated by a pronominal clitic, the corresponding valence requirement of the verbal head seems to be satisfied twice. This causes non-trivial problems if one assumes—as is often the case—that the pronominal clitics are sentential constituents.

[60]Cf. Section 4.1.3 for a discussion of what nominal material can be replicated in Bulgarian.

In the analysis proposed here, the problem with two-fold satisfaction of a single valence requirement does not occur, due to the fact that the clitics are not legitimate constituents on the sentential / clausal level. Predicative pronominal clitics are always integrated into the verb complex constituent, and thus occur on the intermediate morphosyntactic level. Even if they alone satisfy a valence requirement of the verbal head, they do not really occupy a position appropriate for a full sentential nominal constituent. And in the case of replication, the relevant valence requirement is "met" both on the morphosyntactic level—by the corresponding clitic—and on the clausal level—by the coreferential full-fledged NP.

The role of the pronominal clitic is to "enrich" the valence requirement by specifying the index of the respective complement. If we recall that the verb inflection also just specifies the index of the subject, and the experiencer formant actually provides the index of the experiencer object, the optionality of the respective full-NP realisations in all of these cases can be uniformly explained: *whenever the index of a valence requirement is specified, it need not be overtly realised.*

This, in turn, implies that any overt realisation of the subject or of a complement with an already specified index is not directly related to the syntactic well-formedness of the construction in question, and hence has some other, e.g., communicative, motivation.

Chapter 4

Constituent Order and Replication Phenomena on the Clausal Level

The structural model of Bulgarian sentences argued for in this thesis, wherein morphosyntactic and syntactic constituency are formally distinguished, allows for an advantageous representation of the commonly admitted "two-faced" appearance of the object clitics which are neither real morphemes nor full-fledged syntactic constituents. The intuitive background of the proposed treatment is that, in Bulgarian, a certain parallelism exists between the relation of the verb inflection to the subject NP and the relation holding between a pronominal clitic and the corresponding coreferent object NP. The information about the person, number and gender—i.e. the index—of the syntactically nominative subject NP in a sentence is basically available in the morphology of the verb. The same index information plus information about the syntactic case of the respective full-fledged NP complement is supplied by the pronominal clitics within the morphosyntactic verb complex. Thus, both mechanisms—the morphological one of verb inflexion, and the morphosyntactic one of object cliticising—deliver very similar results on the clausal level amounting, on the one hand, to syntactic optionality and, on the other hand, to mainly communicative predeterminedness of any overt realisation of the respective full-fledged nominal constituents.

If a clitic pronoun is not a lexical formant (cf. Section 3.1.3) and does not function as a replicant of a full-fledged overt nominal constituent (a situation to be discussed in detail below), it actually appears as an informationally vacuous place holder for the corresponding valence slot.

As to the clitic replication, its morphosyntactic aspect has already been addressed in Section 3.4. Here I shall concentrate on the communicatively-driven syntactic aspect of the replication phenomena observable in the Bulgarian simple sentence.

It is widely accepted that functional sentence perspective is the main factor determining word order in a Slavic sentence, and that the logical stress (a term commonly used in traditional grammars to indicate the center of intonational prominence in an utterance) is always bound to the rheme. Along with the word order and the intonational contour, which are important means of revealing the communicative load of an utterance, it has also been acknowledged for Bulgarian that the article plays an equally important role—cf. *Ivančev 1957/1978*. And along with nominal determinedness, a further crucial factor having direct impact on the possibility or impossibility of certain communicative segmentations of a given sentence is clitic replication.

Unlike the word order in the lower level constituents (such as nominal, adjectival and adverbial phrases or the verb complex), the clause-level constituent order in Bulgarian shows considerable freedom[61]; very few constraints that could be expressed in grammatical terms—like category or grammatical function—can be discerned. Often the explanation for the incompatibility of a particular choice of words with a particular alignment lies in the (pragmatic) difficulty of finding a discourse context for the utterance.

It has been commonly observed that the surface order of constituents and their grammatical relations appear to be subject to certain mutual constraints even in languages with fairly flexible constituent order. In various frameworks this has motivated the modelling of some basic ordering of grammatical relations that would also provide background, or certain point of departure, for stating constraints of different complexity in the surface order. The HPSG theory makes essential use of the relation of *relative obliqueness* that obtains between syntactic dependants of the same head. As justified in *Pollard and Sag 1987*, the ordering of the elements on the SUBCAT list (the feature modelling subcategorisation, or syntactic valence—i.e. the selectional properties of a given lexical item when it functions as a head of a phrase) corresponds not to surface order, but rather to a version of the traditional *obliqueness hierarchy*, where subjects appear first (leftmost), followed by other complements (if any) in the order primary object, secondary object, then oblique objects and verbal and / or predicative complements.[62] Thus, relative obliqueness is defined (roughly)

[61]In general, "free" with respect to Bulgarian constituent order on the clausal level can be basically used as in *Vilkuna 1989* for Finnish, i.e. meaning discourse-conditioned.

[62]As explained in *Pollard and Sag 1987* and *Sag and Pollard 1989*, the ordering they adopt is broadly similar to other proposed hierarchies of grammatical relations such as the Keenan-Comrie accessibility hierarchy, the 1-2-3-oblique ordering of relational grammar, and the

in terms of the relative order of the elements in the subcategorisation lists of predicators.

While the obliqueness hierarchy of grammatical relations is concerned only with the subcategorised modifiers of a given head, the *systemic ordering*, as employed in, e.g., *Hajičová 1973*, *Hajičová 1984a*, *Hajičová 1984b*, *Hajičová and Sgall 1985*, *Hajičová and Sgall 1987*, *Hajičová and Vrbová 1982*, *Panevová 1980*, *Sgall et al. 1986*, is a broader concept which also includes the adjunct dependants (the free modifiers) of the respective head. The next logical step is to extend this concept further in such a way that it would also cover the (default positioning of the) head itself—for example, Koktová in her word order based grammar (*Koktová to appear*) has introduced the verb directly into the systemic ordering (renamed "neutral ordering" in her approach): the verb is assumed to occupy the second position following what she refers to as the "attitude" and is considered an obligatory syntactic relation (in the context of her distinction between obligatory, optional and free syntactic relations).

In the approach adopted here, it is assumed that what is called canonical element order reflects the obliqueness hierarchy of grammatical relations (appropriate for the particular head) with the heading verb positioned immediately after the least oblique element, prototypically the subject[63], and that the prototypical positioning of free modifiers is to be captured within the broader concept of systemic ordering which in any case subsumes the obliqueness hierarchy introduced by the particular head.[64] However, since I shall concentrate on phenomena involving constituent order in its interaction with clitic replication, and since these are genuinely related to the mutual positioning of the verbal head and its obligatory modifiers, a notion of basic default ordering reflecting just the obliqueness hierarchy would be sufficient for the purposes of the present study.

SUBJ-OBJ-OBJ2 hierarchy employed in the LFG Lexical Rule of Functional Control. The HPSG obliqueness order also corresponds closely to the semantic order of arguments assumed in categorial grammar.

[63]With subjectless verbs the prototypical positioning of the head is either following the experiencer object, if such is available, or initial.

[64]Even though the respective subcategorised elements are not necessarily adjacent in the systemic ordering, their relative order would remain basically unchanged.

4.1 Determinedness and Replication Potential of Nominal Material

This section systematically approaches the problem of which nominal material has the potential to be replicated in Bulgarian. Some of the terms I use—e.g., *replication of nominal material*, *replication potential*, *replication causing factors*—are coined in *Dyer 1988*[65]. The relevant linguistic concept behind them is the process of manifesting as a clitic pronoun the **index** (i.e., the person, gender and number information) and the **case** of a direct or an indirect object NP in a Bulgarian sentence.[66]

Replication phenomena are intrinsically related to the problems of nominal determinedness. In fact, **determinedness as a broader concept** is the main prerequisite for replication of nominal material. Therefore, it is of primary interest for us to clarify what such a concept would subsume.

4.1.1 Approaching the Problem

There seems to be an emerging consensus among the authors treating Bulgarian nominal determinedness in one way or another that this is a fairly broad notion that should be understood as being, at least formally, manifested in a number of different ways.[67]

---

[65]While borrowing certain terminology, I nevertheless disagree with the interpretation of the data given in this paper. For the sake of completeness, it should be pointed out that this material has a working paper status. In a personal remark (1994) the author himself expressed certain doubts about the adequacy of the analysis he had presented there. Still, it should be acknowledged that some of Dyer's ideas—e.g., on the possibility of gauging nominal material with respect to replication potential which, in turn, may be equated with "degree of determinedness"—have been quite insightful for my analysis.

[66]It is important to distinguish clitic replication from pronominal resumption, which may but need not be realised by a clitic pronoun, as the following examples illustrate. In (i) the resumptive element is a clitic, while in (ii) and (iii) it is a full pronoun.

(i) *Kolkoto do **Ivan**, ne sâm **go** kanila.*
as to **John**, NEG-PRT be-1sg ACC-CL invited
'As to John, I have not invited him.'

(ii) *Kolkoto do **Ivan**, **nego** ne sâm kanila.*
as to **John**, **him** NEG-PRT be-1sg invited

(iii) *Kolkoto do **Ivan**, **nego** ne sâm go kanila.*
as to **John**, **him** NEG-PRT be-1sg ACC-CL invited

Both the NP and the pronoun resuming it are given in **bold**. Replication can be observed only in (iii), where the replicated nominal material and the coreferent clitic replicant are underlined.

[67]Cf., e.g., *Naylor 1983*, *Dyer 1988*, *Guentchéva 1994*, *Mayer 1988*, among others.

As a first approximation, let us consider a cross-classification of Bulgarian NPs by two features: (i) presence vs. absence of an article, and (ii) definiteness vs. indefiniteness. With non-determined NPs, both features would be negatively specified: [art -, def -]. These are lexically indefinite NPs with no overt article—e.g., *geroj* 'a hero', *dokumenti* 'papers'. A tentative interpretation would be that the object to which the NP is referring is "totally non-determined" in the sense that either it is unknown to both the speaker and the hearer or, for some (semantic, communicative, etc.) reasons, the speaker considers it so. With determined NPs, at least one of the two features would be positively specified. On the one hand, NPs with an overt indefinite article[68] would have the specification [art +, def -]—e.g., *edni dokumenti* 'certain papers'—and would be tentatively interpreted in the following way: the object to which the NP is referring is "partially determined" in the sense that either it is known to the speaker but (assumed by the speaker to be) unknown to the hearer, or the speaker considers the hearer's awareness of the object the NP is referring to to be irrelevant for the particular communicative purposes. On the other hand, lexically definite NPs with no overt article would be specified [art -, def +]—e.g., *Ivan* 'John', *toj* 'he', *lelja* 'aunt'—implying the tentative interpretation that the object to which the NP is referring is "fully determined" in the sense of being known to the speaker and known or unknown but evident (clearly feasible) to the hearer. Finally, NPs with an overt definite article would have both features positively specified [art +, def +]—e.g., *gerojat* 'the hero'; again, the tentative interpretation would be that the object to which the NP is referring is "fully determined" in the sense of being known or evident to both the speaker and the hearer. This view is summarised in the following table.

	def +	**def -**
art +	NPs with an overt definite article	NPs with an overt indefinite article
art -	lexically definite NPs with no overt article	lexically indefinite NPs with no overt article

In such a simplified classification, however, certain important semantic and communication factors governing the use or non-use of the definite-article morpheme, which are extensively discussed in *Shamray 1989*, have not been taken into consideration. Shamray's analysis makes the original assumption

[68]The existence of an indefinite article in Bulgarian, addressed, e.g., in *Friedman 1976*, is still a controversial issue and a matter of on-going linguistic discussion. In my opinion, there is strong evidence in favour of the assumption that *edin* / *edna* / *edno* / *edni* functions as an indefinite article in certain cases. This is taken into consideration in the proposed classification of articled and non-articled NPs.

that definite articles must be viewed as signals and instructions rather than markers of (e.g., old / new) information inasmuch as both articled and non-articled forms of NPs and PPs can denote (i.e. refer to) objects that may be known or unknown to the speaker and the hearer, real or imaginary, of definite or indefinite quantity. The idea is that by using the definite article, the speaker (i) **indicates** that the object is (available as) known, and (ii) **instructs** the hearer to look for it within some perimeter of presupposed common knowledge. Such an instruction has various motivations—from direct mentioning of the object in the preceding text, to the speaker's hypothesis concerning, e.g., the hearer's world knowledge, or the hearer's familiarity with the particular communicative situation. Shamray also observes that the requirements posed on the conditions of use of demonstrative pronouns (in combination with non-articled NPs / PPs) are much stricter than those concerned with the use of the definite article: it is only in the former case that the respective object has to be either mentioned before or visually perceptible. On the other hand, the speaker gives no instructions of the above-mentioned type when using non-articled NPs / PPs, which, however, does not mean that the respective object is unknown or non-evident to the hearer. It is rather the case that according to the particular communicative intentions of the speaker any special indication of the object as being known (or unknown) to the hearer is not considered relevant.

Shamray further distinguishes the deictic nature of the definite article from a meaning appropriate to all articles, which she calls "limitedness" or "lack of limitedness" (in my notation below [lim +] and [lim -], respectively).

> Značenijata limitiranost / otsâstvie na limitiranost ne otrazjavat prjako priznacite na obektite ot vânšnija svjat, a različnite načini, po koito nie mislim za tjax v zavisimost ot komunikativnite si celi.
>
> The meaning of limitedness, or lack thereof, does not reflect directly the features of the objects in the external world, but rather it reflects the different ways we think about them in accordance with our communicative goals. *Shamray 1989* (p.51)

Using a particular noun, the speaker—on the one hand—categorises or generalises, and—on the other hand—identifies or particularises the object which is referred to by this noun. The feature [lim +] would indicate then that the identifying, i.e. the particularising, aspect of the respective noun has been activated, while the feature [lim -] would mean that only the categorising, i.e. the generalising, aspect of the noun is relevant for the

communication. On such a basis, Shamray assumes that the definite article and the "zero" article are morphemes, with the former setting the value of the feature limitedness positive, and the latter negative. In order to include the Bulgarian indefinite article as well, let us develop the idea further: one could certainly also regard *edin* as a means of expressing the meaning of limitedness but in the sense of *indefinite* particularisation, in opposition to *definite* (unique) particularisation as realised by the definite article.

In her analysis, Shamray divides the sentence into a *characterised* part containing the object(s) of characterisation and a *characterising* part containing some aspect(s) of this characterisation. From such a perspective, articled NPs / PPs can occur in **both** parts, and in all these cases, the speaker consciously provides certain information about the designated objects, since any use of an article—according to Shamray—generally indicates the speaker's intentions to say something about the respective object, while non-articled NPs / PPs (which are also lexically indefinite) **always** occur in the characterising part. In the latter case, the speaker expresses no special intentions to provide any particular information about the designated objects; rather, the lack of an article indicates that the respective object is considered only to an extent that will allow certain information to be supplied about another object. Thus, the use of articled or non-articled NPs / PPs can be considered "free" in Bulgarian, in the sense that speakers—according to their intentions and goals—are free to express different attitudes toward the situation described in the sentence, to treat it in a different way preferring one articulation into a "characterised" and a "characterising" part to another.

An attempt to directly integrate a "limited vs. non-limited" feature into the classification of Bulgarian NPs is illustrated in the next table.

lim +	**def +**	**art +**	*definite-article morpheme*
		art -	*inherent definiteness*
		art none	*demonstrative pronominals*
	def -	**art +**	*indefinite article*
lim -	**def -**	**art -**	*"zero" article*
		art none	*indefinite pronominals*
		art +	*indefinite article* (intended meaning "proper", "authentic", "real", "in general")
	def +	**art +**	*definite-article morpheme* (intended meaning "in general", "all", "as such", or uniqueness)

Obviously, the **generic** uses of both the definite and the indefinite articles[69] are incompatible with the assumption that any presence of an article activates the identifying (particularising) aspect of the respective noun. Therefore, let us re-consider how exactly concepts like "limitedness", "specificity" and "generity" can be related to each other.

4.1.2 Towards a Typology of Non-Articled and Articled Noun Phrases in Bulgarian

In agreement with Shamray, let us assume that there are two basic semantic aspects behind any nominal use which can be highlighted in accordance with the communicative intentions of the speaker. My tentative terms are ***categorising*** potential, which corresponds to (the semantics of) "non-limitedness" of nominal material, and ***identifying*** potential, which corresponds to (the semantics of) "limitedness" of nominal material. I further assume that, from the perspective of the categorising potential, one could speak of *non-generic* and *generic* descriptions of objects, while, from the perspective of the identifying potential, it is possible to distinguish *specific* and *non-specific* descriptions of objects:

categorising (lim -)	non-generic
	generic
identifying (lim +)	specific
	non-specific

Thus, the dichotomy *generic* vs. *non-generic* appears to be relevant only when the categorising potential of certain nominal material is activated, and the dichotomy *specific* vs. *non-specific* makes sense only with activated identifying potential of some nominal material. How nominal material will be used in each particular case of language communication is totally speaker-oriented and depends on what "world" or "perspective" the speaker chooses to present to the hearer.

Most generally, **non-articled NPs** in Bulgarian can be viewed now as having three main uses: the prototypical ***categorising non-generic use***, the ***identifying specific use*** bound to the so-called "inherent definiteness" of, e.g., proper names, kinship terms, etc., and the ***categorising generic use*** which is revealed in restricting environments, e.g., of type "definition".

[69]The generic use of the definite-article morpheme is indicated as [lim -, def +, art +], while the generic use of the indefinite *edin* as [lim -, def -, art +].

Bulgarian **NPs with definite article**—parallel to what can be observed synchronically in many languages—have two main uses: the ***identifying specific use*** and the ***categorising generic use***. I assume that both of them are prototypical, agreeing with *Mayer 1988* who considers the latter one:

> ... a secondary, conventionalised usage where there is no reference to a specifically identifiable object (or set of objects). Rather, a general statement is made to refer to a whole class. (p.25)

Finally, I view Bulgarian **NPs with indefinite article** as having three uses: the prototypical ***identifying specific use***—in clear opposition to the same use of NPs with a definite article, the ***identifying non-specific use***, and the ***categorising generic use***—generic descriptions as statements which impart qualities or characteristics to all members of a group are not necessarily expressed by the definite article in Bulgarian.

I follow *Mayer 1988* in assuming that a definite description is one in which an object (person, thing, situation, etc.) or set of objects is presented by the speaker as being identifiable in a specific context, regardless of whether or not the object is in fact pragmatically identifiable. For this reason, the feature "definiteness"—understood as "unique identifiability"—is inappropriate (i.e. set to "none") for nominal material with activated categorising potential. On the contrary, this feature is appropriate for nominal material with activated identifying potential (and therefore, set positively or negatively for the respective cases).

Demonstrative pronouns, while sharing some of the semantic properties of the definite article, always imply deixis and are limited to anaphoric uses; therefore, they are not normally used in generic descriptions.

Proper nouns tend not to be articled because they are inherently definite (inherently identifiable) and are usually used with unique reference; their definiteness is revealed when they are preceded by adjectival modifiers by articling the adjective. NPs headed by kinship terms are not articled when used as names (except if containing preposed adjectival modifiers) but can be normally articled when used as descriptive common nouns; for this reason, all kinship terms are articled when used in the plural.[70]

---

[70]Note, however, that nicknames (e.g., ***Borimečkata*** *vleze zasmjan.* 'B. came in smiling.') and diminutive proper nouns (e.g., ***Kateto*** *i* ***Vankata*** *bjaxa mnogo iznenadani.* 'Kate and Jonny were quite surprised.') are, as a rule, articled when not used as vocatives. The unarticled form serves as vocative with such nouns (e.g., ***Borimečka***, *ela pri nas*! 'B., come and join us!' ***Kate***, ***Vanka***, *tuka li ste*? 'Kate, Jonny, are you here?').

It is also important to distinguish between *edin* used as a cardinal numeral or an indefinite pronoun and *edin* used as an indefinite article. Prosodically, as observed in, e.g., *Scatton 1984* (p.316), this element has a primary stress (´) in the former case (56a), but carries only a secondary stress (`) in the latter, i.e. as an indefinite article (56b).

(56)

a. *Dajte mi edná kniga.*

'Give me one book.'

b. *Dajte mi ednà kniga.*

'Give me a book.'

In (56a) *edin* can also be in the focus of the utterance, e.g., *Dajte mi EDNA kniga, a ne dve*. 'Give me ONE book, not two.', which, in turn, is absolutely impossible in (56b). Morphologically, as a cardinal numeral or an indefinite pronoun *edin* can take the definite article: *ednoto dete* ('the one child') or *ednite deca* ('some (of the) children'), while this is absolutely impossible when *edin* functions as an indefinite article.

All these assumptions and observations result in a classification of Bulgarian articled and non-articled NPs, which is summarised in the following table.

categorising (lim -)	non-generic	def none	no art	prototypical use	*Tuk kupuvam knigi*.'I buy books here.' *Vleze grupa studenti*.'A group of students came in.' *Târsja prijateli*. 'I am looking for friends.'

generic	def none	def.art	prototypical use	*Kuče**to** e prijatel na čoveka**a**.* 'Dog is man's friend.' *Xora**ta** sa smârtni.* 'Humans are mortal.' *Toj običa vino**to** i ženi**te**.* 'He loves wine and women.' *Običaj bližni**ja** si.* 'Love your fellow-man.' *Gotov e vsičko da razdava na xora**ta**.* 'He would give everything to the people.' *Prozorec**ât** na staja trjabva da bâde svetâl i širok.* 'A window of a room should be bright and wide.' (in restricted context) *Fizika**ta** e edna ot naj-starite nauki za prirodata.* 'Physics is one of the most ancient sciences about nature.' (uniqueness)
		indef.art	"properness" "authenticity" (highlighting basic features)	***Edna** majka vinagi šte poznae deteto si.* 'A mother would always recognise her child.' *Podobno nešto ne može da se sluči na **edin** specialist.* 'Nothing like that could happen to an expert.' ***Edin** faringolog lekuva i ušni bolesti.* 'A pharyngologist also treats ear diseases.' ***Edin** vâlk nikoga ne se rešava da umre ot glad pred **edno** stado ovci.* 'A wolf never decides to die of hunger in front of a flock of sheep.' (E. Pelin) *Ne vi li e sram, kakvo iskate ot **edni** ženi.* 'Aren't you ashamed - what do you want with women?' (J. Jovkov) ***Edno** okončanie, obšto vzeto, ne izčezva, dokato e funkcionalno neobxodimo.* 'An ending, generally speaking, does not disappear as long as it is still necessary from a functional standpoint.' (Minkov)

			no art	"definitions" (prototypical use in predicatives)	*Za proizvodstvo na xartija se izpolzva dârvesina.* 'Wood is used for paper production.' *Stepen s osnova otricatelno čislo i četen pokazatel e položitelno čislo.* 'A power with a negative base and an even exponent is a positive number.' *Zaek, kojto e ranen, e lesna pljačka za kučetata.* 'A rabbit which has been wounded is an easy mark for dogs.' *Tova životno e vâlk.* 'This animal is a wolf.'
identifying (lim +)	specific	def +	def.art	prototypical use	*Vidjaxte li knigi**te** / dve**te** knigi / mnogo**to** knigi / vsički**te** knigi / njakolko**to** knigi*? 'Did you see the books / the two books / (the) many books / all the books / (the) several books?'
			no art	demonstratives	*Tezi knigi / onezi knigi mi trjabvat.* 'I need these books / those books.'
			no art (def.art with a preposed adjective)	"inherence"	full personal non-reflexive pronouns: *toj* 'he', *nego* 'him' (cf. *gorki**jat** toj* 'poor he') proper names: *Ivan*, *Stara Planina* (cf. *glupavi**jat** Ivan* 'stupid John') kinship terms: *majka* 'mother', *tatko* 'daddy', *baba* 'granny', *lelja* 'aunt' (cf. *stara**ta** mi baba* 'my old granny') others: months, days of week, points of compass (cf. *minali**ja** januari* 'the last January')

		def -	indef.art	prototypical use	***Edni*** *knigi gi njama ošte ot včera.* 'Some books are missing since yesterday.' *Dojdoxa s* ***edna*** *(xubava) kola.* 'They arrived with a nice car.' *Tja počuka na* ***edna*** *vrata.* 'She knocked on the (literally, 'a') door.' (E. Stanev) *Trjabvat mi* ***edni*** *dokumenti.* 'I need some papers.' ***Edna*** *studentka donese cvetja.* 'A student brought flowers.' *Da ti dam* ***edin*** *balsam da si usladiš sârceto.* 'Let me give you a balm to soothe your heart.' (Georgiev) *Po livadata tiča boso* ***edno*** *dete.* 'A child is running barefoot through the meadow' (Sv. Ivančev)
	non-specific	def -	indef.art	"some / any"	*Molja dajte mi* ***edin*** *moliv.* 'Please, give me a pencil.' *Potârsi mi* ***edni*** *po-iziskani drexi za utre.* 'For tomorrow, look for some more fancy clothes for me.' *Mislexme, če ima talanta na* ***edin*** *Dostojevski.* 'We thought that he had a Dostoyevskian talent.' *Toj beše nadaren i s glasa na* ***edin*** *Stentor.* 'He was also endowed with a Stentorian voice.' (Maslov)
			no art	lexical means (inherent indefiniteness or quantification)	*njakakvi knigi, njakoi knigi, edi-koi si knigi, koi da e knigi, koito i da bilo knigi; vsjaka kniga* 'some / certain / any books; every book' *pet knigi, mnogo knigi, njakolko knigi, vsički knigi* 'five / many / several / all books'

4.1.3 Replication Potential

On the basis of the typology of articled and non-articled NPs presented in the previous section, it is possible now to formulate constraints concerning the replicability of Bulgarian NPs. Thus I argue that what can be replicated by a clitic pronoun—under the appropriate verb-lexeme specific, syntactic or communicative conditions[71]—is the nominal material that is used as an ***identifying specific*** description of a given object. This is illustrated in (57-61), where the replicated material—both in Bulgarian examples and in their English translations—is underlined, while the respective replicating clitics are double-underlined.

(57)
*Dete**to** go dovede Elena*. 'Helen brought the child.'

(58)
Tezi knigi gi kupix za teb. 'I bought these books for you.'

(59)
Nego go poznavam ot universiteta. 'I know him from the university.'

(60)
a. *Na Ivan mu izpratixme nova pokana*. 'We sent a new invitation to John.'

b. *Baba ja risuva brat mi*. 'My brother sketches Granny.'

c. *?Juli go prekaraxme na moreto*. 'We spent July at the seaside.'(colloquial)

(61)
Edni studenti gi očakvame na objad. 'We expect certain students at lunch time.'

On the other hand, non-articled NPs which are categorising or non-specific descriptions, as well as articled NPs which are generic or non-specific descriptions completely lack replication potential.

---

[71]Under verb-lexeme specific conditions, I understand the cases of an obligatory pronominal clitic which functions as an experiencer formant with verbs like *trese me* 'to shiver', *domâčnjava mi* 'to become nostalgic', etc.—cf. Section 3.1.3 for argumentation that this phenomenon is more naturally interpreted as a specific type of analytic object-verb agreement, rather than clitic replication.
The syntactic and the communicative conditions basically amount to the two major replication-causing factors in Bulgarian: surface alignment and object thematisation—cf. Section 4.3.1.

.c.4.1.4 Some Related Cliticising Effects

With direct-object NP and indirect-object *na*-NP complements of a verb certain interesting regularities with respect to (the possibilities of) cliticising can be observed. Let us consider for the sake of illustration the shortest reply, e.g., in the context: 'They ask me if you have seen _NP_.'—'I have.'

(62)
Pitat me dali si vidjal ...

a. ... *xora* (people). —*Vidjax.*
b. ... *novi xora* (new people). —*Vidjax.*
c. ... *vseki gost* (every guest). —*Vidjax.*
d. ... *nejni gosti* (her guests). —*Vidjax.*
e. ... *pet knigi* (five books). —*Vidjax.*
f. ... *knigi, koito / deto ti trjabvat* (books you need). —*Vidjax.*
g. ... *edni xora* (certain people). —*Vidjax (gi).*
h. ... *nejnite gosti* (her_def.art guests). —*Vidjax gi.*
i. ... *gostite ì* (guests_def.art her-POSS-CL). —*Vidjax gi.*
j. ... *xorata* (people_def.art). —*Vidjax gi.*
k. ... *tezi xora* (these people). —*Vidjax gi.*
l. ... *pette knigi* (five_def.art books). —*Vidjax gi.*
m. ... *Ivan* (John). —*Vidjax go.*
n. ... *majka* (mother). —*Vidjax ja.*

It can be observed that whenever an ***identifying specific definite*** NP is involved, an obligatory clitic (62h-n) occurs in the reply. If the NP is an ***identifying specific indefinite*** one, the clitic in the reply is optional (62g.). In all other cases, no clitic in the reply is possible.

All this suggests that such a query-and-reply test can be useful as a diagnostic tool for defining nominal determinedness in general[72] as well as clitic replicability of nominal material in particular.

[72]For example, even for languages having no morphologically established category of "definiteness" but a well developed clitic system, e.g., Czech, Polish, etc.

4.1.5 Summary and Conclusions

Determining which nominal material has the potential to be replicated by a clitic pronoun in a Bulgarian sentence is a complex task involving criteria like the following:

- Which semantic aspect has been activated in the particular nominal use, the categorising or the identifying potential of the nominal material?
- Does the activation of the categorising potential result in a generic or a non-generic use of articled NPs?
- Does the activation of the identifying aspect result in a specific or non-specific use of the articled NPs, and are the objects referred to by the respective nominal material regarded as uniquely identifiable (i.e. definite) or not?

The first criterion is based on *Shamray 1989* whose proposal has been developed further in the presented analysis in order to accommodate the commonly made distinction between generic and specific NP-use taking into consideration not only articled but also unarticled NPs in a systematic way. While Shamray's study focuses on the distribution of the definite article and demonstrative pronouns, *Mayer 1988* is primarily concerned with the definite article arguing that it has two "widely disparate meanings"—(a) the specifically identifiable and (b) the generic, and that the dichotomy specific vs. generic refers to "totally different functions rather than aspects of the same meaning". In the classification proposed here, the observed difference is structured in a more general way by viewing the specific use of articled NPs as resulting from activation of the identifying potential of nominal material, and the generic use of articled NPs as resulting from activation of the categorising potential of nominal material. Definiteness in the sense of *unique* identifiability of the respective object is then viewed just as a special case of specific identifiability—a broader concept also covering the identifying specific use of the indefinite article. The controversial issue of the Bulgarian indefinite article is not addressed in this study. It is rather assumed that there is sufficient empirical evidence in favour of its existence. Therefore, the current proposal accommodates the different uses of NPs with indefinite articles in Bulgarian as categorising generic descriptions, as identifying specific descriptions, and as identifying non-specific descriptions.

The main advantage of the proposed NP typology is that it allows for the statement of a general non-replicability constraint. Namely, if certain

nominal material functioning as a direct or an indirect object is used as a categorising or as an identifying non-specific description, it can never be replicated. Only nominal material which is used as identifying specific description of the respective object is replicable, i.e. has replication potential in Bulgarian.

Another interesting outcome of the presented approach is that Ivančev's distinguishing of degrees of definiteness—i.e. "close definiteness" (which is defined by anaphoric reference) and "distant definiteness" (which is defined by general context or situation)—can now also be rendered precise as being relevant only for NPs that are used as identifying specific definite descriptions of the respective objects.

In the following sections, the clitic replication will be discussed in its relation to the constituent order in the simple Bulgarian sentence; the presented understanding of replication-relevant determinedness is thus a necessary prerequisite for such a discussion.

4.2 Communicative Aspect of Utterances (Synopsis of Relevant Issues)

The communicative aspect of utterance is crucial in licensing the particular surface alignments in Bulgarian, therefore, the main goal of this section is to present the relevant concepts as they have been traditionally viewed, and possibly to refine them for the needs of the current study.

The communicative organisation of a sentence is determined by the needs of the actual information this sentence is supposed to transmit in the particular situation in which it is uttered. In any given utterance, it can basically be distinguished between an informative part and an anchoring or vehicular part. The former is assumed to represent the rheme (or focus) of the respective utterance, and the latter is usually called theme[73] (or ground, or topic). A typical generalisation concerning the relation between the theme / rheme (topic / comment) partitioning, definite article, and constituent order in Bulgarian is reflected in the position taken in, e.g., *Mayer 1988*:

[73]Among the authors working in this perspective, there is difference in the use of the term "theme". As *Vallduví 1990* observes, there are, at least, two "themes", that can be called Firbas-theme (more or less analogous to the topic in the topic-focus framework) and Halliday-theme (almost equivalent to the topic in the topic-comment framework).

> Although there is no one-to-one correspondence between definite / indefinite, topic / comment, and initial / non-initial word order, there is a tendency for topics to take the definite article, since they often represent previous mention or knowledge, and to occur in sentence-initial position. Comments can be definite or indefinite but the comment is often in non-initial position, and, therefore, those parts of the comment that present "new" information are less likely to take the definite article. (p.25)

The application of the Prague School theory of functional sentence perspective[74] to Bulgarian originates in the works of Svetomir Ivančev[75]. His observations promote both informational and grammatical principles as being important for (i) the determination of Bulgarian sentential word order, (ii) the distribution of the definite article, and (iii) the establishment of the mechanism of clitic doubling in the system of the language. In the following, I shall extensively base my analysis on the leading ideas of Ivančev's research.

- **Objective and Subjective Alignments and the Concept of Communicatively (Un)marked Word Order**

Let us (re-)introduce some basic terminology. The term *objective word order* (OWO) refers to an alignment where the thematic elements linearly precede the rhematic ones, i.e. the basic linear precedence constraint underlying this concept can be expressed as [theme +] < [rheme +]. On the other side, the term *subjective word order* (SWO) is assumed to refer to an alignment in which the rhematic elements linearly precede the thematic ones, i.e. [rheme +] < [theme +].

The concept of *communicatively unmarked word order* (CUWO), as opposed to the *communicatively marked word order* (CMWO), depends on the communicative type of the utterance. The unmarked case for interrogative, imperative and exclamative utterances is the SWO, while the OWO is the unmarked case for declarative utterances. As *Ivančev*

---

[74]Cf. *Mathesius 1907*, *Mathesius 1915*, *Mathesius 1941*, *Mathesius 1942*, *Mathesius 1947a*, *Mathesius 1947b*, *Mathesius 1975 (1939)*, *Trávníček 1937a*, *Trávníček 1937b*, *Trávníček 1951*, *Trávníček 1959*, *Trávníček 1962*, *Firbas 1957*, *Firbas 1964*, *Firbas 1971*, *Firbas 1972*, *Firbas 1974*, *Firbas 1975*, *Daneš 1959*, *Daneš 1967*, etc.

[75]Even though some basic ideas of functional sentence perspective have been proposed for the analysis of Bulgarian as early as *Szober 1933/1979*, it has come to a systematic application only in the works of Ivančev, e.g., *Ivančev 1957/1978*, *Ivančev 1967/1978*, *Ivančev 1968/1978*, *Ivančev 1971*, *Ivančev 1974/1978*, *Ivančev 1977/1978*, *Ivančev 1978*, and consequently in *Georgieva 1974*, *Penčev 1980*, *Dyer 1992*.

1957/1978 observes, the marked type of word order (e.g., the SWO in declarative clauses) is much rarer in the written language due to the fact that the emphatic "logical stress"—an important factor in signalling rheme in marked word orders—can then be determined only indirectly, i.e. on the basis of the context, particular linear ordering, presence of article, while the unmarked type of word order (e.g., the OWO in declarative clauses) is the most straightforward alignment and is, therefore, more frequent.

- **The Subject**

The highest degree of grammaticalisation in contemporary standard Bulgarian can be claimed for the thematic usage of the subject[76]. The prototypical properties of subjects are specified by *Ivančev 1957/1978* in the following way:

> Ne može da ima sâmnenie, če edno občleneno ime v sâvremennija bâlgarski ezik, koeto e upotrebeno samostojatelno bez predlog v načaloto na izrečenieto i ne nosi logičesko udarenie, e podlog v izrečenieto, ili s drugi dumi, sâotvetstvuva na imenitelnija padež vâv fleksivnite ezici. ...[T]ri momenta fiksirat sintaktičnata funkcija na podloga v sâvremennoto bâlgarsko izrečenie: 1. Načalno položenie v izrečenieto. 2. Lipsa na logičesko udarenie. 3. Naličnost na opredelitelen člen.
>
> There is no doubt that an articled noun in modern Bulgarian, which is used alone without a preposition at the beginning of the sentence and carries no logical stress, is the subject in this sentence, or in other words, that it corresponds to the nominative case in flective languages. ... Three factors fix the syntactic function of the subject in the modern Bulgarian sentence: 1. Initial position in the sentence. 2. Lack of logical stress. 3. Presence of a definite article.

It is important to realise that even in OWO-alignments in Bulgarian declarative clauses the position of the subject is not determined structurally. If the subject is thematic it normally precedes the rhematic parts of the utterance, i.e.

subject [theme +] < [rheme +],

while a rhematic subject normally follows the thematic parts in the respective utterance, i.e.

[theme +] < subject [rheme +].

[76]Cf. also the observations in *Mayer 1988* (p.71).

In the former case, the thematic subject precedes the verb and is almost always articled, while in the latter it occurs in non-initial position, or even follows the verb, and is most typically not articled. Certainly, no immediate adjacency of the subject and the verb is presupposed in either of these cases.

In SWO-alignments in declarative clauses the opposite situation can be observed: if the subject is thematic, it follows the rhematic parts of the utterance, i.e.

[rheme +] < subject [theme +],

and if it is rhematic, it precedes the thematic parts, i.e.

subject [rheme +] < [theme +].

- **On "Logical" Stress**

Words and constituents in utterances can be focused for various reasons, and are then marked by special intonational contours. In traditional grammar studies of Bulgarian, the notion of "logical" stress is frequently assumed to be inherently bound to the rheme (or focus) of an utterance. For example, *Georgieva 1974* (p.67), citing *Adamec 1966*, specifies logical stress, regardless of its intensity which can be different depending on certain circumstances, as the "dynamic, intonational culmination of a sentence"; according to her, the "linear-dynamic structure of the sentence on the level of speech" is made clear on the basis of a particular word order and the location of the logical stress, taken together. In one of the most representative works on Bulgarian intonation— *Penčev 1980*—a whole chapter is dedicated to the direct correlation between (different types of) rhematicity and intonation, whereby rhematic sentential intonation may be understood—somewhat simplified, of course—as stress. With the oral tools of intonation and / or stress, the speaker can, e.g., depart from the canonical subject-verb-object order in a sentence by choosing—with very few limitations—any collocation of elements.

Recent studies in intonational phonology—cf. *Ladd 1996*—distinguish between normal and contrastive (or emphatic) stress. The former has no meaning or function but is rather the result of the operation of phonological rules on surface syntactic realisations and is assumed to be the accent pattern (intonational contour) that can convey broad focus. At the same time, the fact that narrow focus can be conveyed by pronunciation that is not phonetically distinct from broad focus readings results in the genuine

ambiguity appropriate for normal-stress-pattern examples, as in the one in (63a).[77] Prototypically, this is the main stress of the sentence appropriate for the rightmost clausal constituent in neutral declarative utterances. Only the so-called contrastive (or emphatic) stress, which is seen as essentially unpredictable and beyond the scope of normal stress rules[78], can be said to carry meaning. These are the cases where the (traditionally dubbed logical) stress is perceived as "strong" or "extraordinary". It can occur anywhere in the sentence, attracting the focus and allowing for narrow-focus interpretations only—cf. (63b-e).

(63)[79]

a. *[Poetesata [razdade [knigi [na decata]]]].*

poetess_def.art gave books to children_def.art.

'The poetess distributed books among the children.'

b. *Poetesata razdade knigi [NA DECATA].*

c. *Poetesata razdade [KNIGI] na decata.*

d. *(?) Poetesata [RAZDADE] knigi na decata.*

e. *[POETESATA] razdade knigi na decata.*

Since unambiguous narrow focus pronunciation always involves local emphasis, this can sometimes be brought about into play as a paralinguistic (rather than linguistically structured) device to eliminate ambiguity in the breadth of focus. This can be best illustrated if the examples in (63a-b) are compared: the interpretation of the focus as unambiguously narrow is made clear by the presence of local emphasis on the item which, in both of these cases, is intonationally most prominent. As commonly observed, non-determined elements attract the stress; therefore, the verb having such a complement resists receiving stress—cf. the relative oddness of the example (63d) illustrating a narrow focus on the verb in the presence of a non-definite direct object (i.e. *knigi* 'books'), if compared to the various other communicative structurings given in (63).

---

[77]In general, "ambiguous focus" can be observed with alignments complying with the canonical (i.e. obliqueness-determined) order—cf. Section 4.3.2 for a detailed presentation of the data.

[78]Normal-stress rules can be seen as a description of where accent is placed when focus is broad.

[79]In the following examples the (possible) rhemes, i.e. the scope of the focus, are indicatied by bracketing, and the location of the emphatic stress by CAPS; thus, an utterance containing no emphatic-stress marking presupposes a normal stress pattern, i.e. a prominence contour on the rightmost accentable item.

As far as their placement with respect to the verb (complex) is concerned, phrases bearing local emphasis behave similarly to *wh*-interrogative phrases and *li*-marked interrogative phrases. In particular, if there are several preverbal syntactic constituents, only the one immediately preceding the verb can be one of these types—cf. the examples in (64). Thus the legitimate pattern here will be

(...) STRESSED CONSTITUENT V (...)

with "stressed constituent" referring to and standing for an interrogative (*wh*- or *li*-marked) phrase or for a phrase marked with (or containing) an emphatic accent[80].

(64)

a. ***Koj*** *razdade knigi na decata?*
who gave books to children_def.art

b. *Na decata knigi* ***koj*** *razdade?*
to children_def.art books who gave

c. ***Poetesata li*** *razdade knigi na decata?*
poetess_def.art Q gave books to children_def.art

d. *Na decata knigi* ***poetesata li*** *razdade?*
to children_def.art books poetess_def.art Q gave

e. ***POETESATA*** *razdade knigi na decata.*
poetess_def.art gave books to children_def.art

f. *Na decata knigi* ***POETESATA*** *razdade.*
to children_def.art books poetess_def.art gave

As (65) shows, the acceptability decreases if any other of the preverbal constituents is stressed. So, the following combination is generally problematic:

 # (...) STRESSED CONSTITUENT (...) unstressed constituent V (...)

(65)

a. ***#Koj*** *knigi razdade na decata?*
who books gave to children_def.art

b. ***#Poetesata li*** *na decata razdade knigi?*
poetess_def.art Q to children gave books

c. ***#POETESATA*** *knigi na decata razdade.*
poetess_def.art books to children_def.art gave

[80]These are set **bold** in the examples.

It is further impossible for an unstressed constituent to stand between two stressed ones, which generally excludes the following pattern:

***(...) STRESSED CONSTITUENT unstressed constituent STRESSED CONSTITUENT (...)**

Summing up, only one peak of prominence may occur—if at all—before the verb (or the verb complex), and must be contained in the phrase immediately preceding the verb (complex), which is summarised in Figure 25.

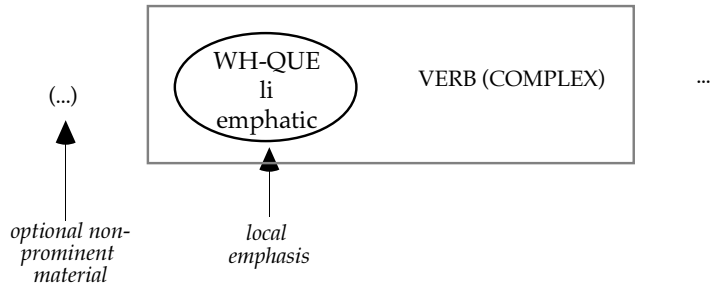


Figure 25

• The Information Packaging Approach

The dynamic perspective of the information packaging approach to communicative organisation of the sentence is recently gaining extensive ground in the (computational) linguistic community, and not at last because of its formalisability, and hence, potential computational tractability—e.g., *Engdahl 1994*. The basic idea of information packaging, a notion introduced in *Chafe 1976* which is the object of study in *Vallduví 1992*, *Vallduví 1994*, *Engdahl and Vallduví 1994*, *Vallduví and Engdahl 1995*, etc., is that speakers do not present information in an unstructured way, but they provide the hearer with detailed instructions on how to manipulate and integrate this information according to their beliefs about the hearer's knowledge and attentional state:

> Our point of departure is the assumption ... that what underlies the focus-ground distinction is a need to "package" the information conveyed by a sentence so that hearers can easily identify which part of the utterance represents an actual contribution to their information state at the time of utterance, and which part represents material that is already subsumed by this information state. *Vallduví and Engdahl 1995* (p.520)

The communicative role of an utterance is twofold: it not only conveys new information, but also contains hints linking the actual communicative contribution to the preceding discourse in order to integrate it with the context. For instance, sentences such as (66a-c) differ not in **what** they say about the world, but in **how** they say it, and hence are not mutually interchangeable in a given context[81].

(66)

a. *Ivan xaresva kartinata.*
John likes picture_def.art
'John likes the PICTURE.'

b. *Ivan XARESVA kartinata.*
'John LIKES the picture.'

c. *IVAN xaresva kartinata.*
'JOHN likes the picture.'

The actual difference is in the way the same propositional content is "packaged" in the given sentences. These different packagings are viewed in *Vallduví 1992*, *Vallduví 1994* as different instructions for information update. The truth-conditional identity of the sentences in (66) is reflected in the fact that they yield the same output information state. Differences in update potential between sentences that differ only in the scope of their focal segments are, therefore, due to the fact that they can felicitously update different information states. The knowledge of how to use and interpret information-packaging instructions is part of any speaker's linguistic competence.

Vallduví's account of information packaging yields a hierarchical articulation with three primitive notions that encompasses the two traditional bi-partitionings of utterances into focus-ground[82] and topic-comment[83]. The core distinction is made between the information (focus, a

[81]When not explicitly indicated, the main sentential stress—i.e. the intonational center marked by the relevant pitch accent—is observed within the rightmost syntactic phrase. In Bulgarian, it is perceptually weaker than the pitch accent occurring elsewhere (the latter is indicated—as it is common practice—by SMALL CAPS).

[82]In this bi-partitioning, a focus is the informative part of the sentence, the part that (speaker believes) makes some contribution to the hearer's mental state, and a ground is the non-informative part of the sentence, the part that anchors the sentence to what is already established or under discussion in (the speaker's picture of) the hearer's mental state.

[83]In this bi-partitioning, the topic—typically sentence-initial—part expresses what the sentence is about, and the comment is the part that expresses what is said about the topic. Topics are thus points of departure for what the sentence conveys, they link it to previous discourse.

rhematic part) and its anchoring (ground, a thematic part), with the latter being further divided into a link (the traditional topic) and tail—Figure 26.

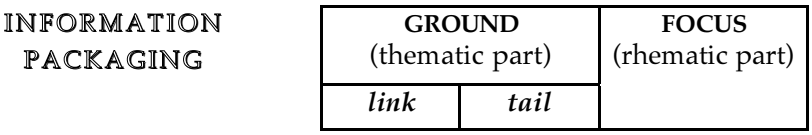


Figure 26

An important property of this partitioning is that it implies neither constituency nor continuity of the employed informational primitives.

The focus encodes the information conveyed by the sentence, which can be metaphorically described as the proposition expressed by the sentence minus the information (assumed by the speaker) already present in the hearer's model. Thus the focus is defined as the actual update potential of the respective sentence, i.e. the only contribution that (according to the speaker) this sentence makes to the information state of the hearer at the time of utterance. All sentences have a focal segment. The ground, in contrast, is already subsumed by the input information state and acts as an usher for the focus: it indicates how the information update is to be carried out by specifying the way in which the focus information fits in the hearer's model. Sentences have a ground only if the context warrants its use, i.e. if the ushering is (thought by the speaker to be) required. The two sub-segments of the ground—link and tail—contribute in their own way to the ushering role of the ground. In particular, links indicate where the focus information should go by denoting a locus of update in the hearer's model, while tails indicate how the focus information fits there. From the way the authors define the informational primitives it follows that, first, the focus-ground partition of a simple sentence is composed of discrete units that do not overlap, hence a given element can be interpreted as focus, link or tail, but may not be interpreted as both focus and tail or focus and link, and second, every non-weak (or communicatively significant, in the terminology proposed in *Avgustinova and Oliva 1995a, Avgustinova and Oliva 1995c*) element in a sentence must contribute to its information structure. It is argued in *Vallduví 1992* that weak proforms (e.g., null pronouns and pronominal clitics in Catalan, or unstressed pronouns in English) do not participate in the construction of information-packaging instructions. Generally, weak pronouns can be neither focal nor ground, they cannot display any of the structural properties associated with either foci or links

or tails (e.g., association with a rhematic or a thematic accent or placement in informationally relevant positions, etc.). Strong pronouns, on the contrary, are informationally full-fledged items and may take part in the informational segmentation of an utterance as foci, links or tails. As *Cyxun 1962* has already observed, no clitics can be used as communicatively significant parts of the Bulgarian sentence, i.e. as a theme (his "base") or as a rheme (his "kernel"), since a phonologically weak pronoun is also a communicatively "weak" form. Further observing that communicatively significant parts tend to be located at the beginning and in the end of a sentence, depending on the particular type of word order—objective or subjective, he regards these positions as communicatively "strong" positions and argues that this is most clearly manifested in the choice of strong pronominal forms for occupying them. Such a standpoint can offer a legitimate explanation for the clause-initial restriction, which prevents endoclitics (in my terminology: core verbal clitics) from occurring sentence-initially or after an intonational pause, as well as for the apparent tendency for these to attach proclitically if their verbal host is not clause-initial[84]. Also in *Nicolova 1986* (p.49) it is argued that short pronominal forms not only cannot be rhematic but they cannot be thematic either, and that they can be contained in the theme or in the rheme only in combination with the stressed word to which they lean (i.e. the verb, in modern literary Bulgarian).

There is increasing an awareness of the large degree of cross-linguistic diversity involved in the structural realisation of information packaging. Different languages choose different means—involving to a different degree prosody, word order, morphosyntax, etc.—to spell out the same informational interpretations. For example, *Vallduví and Engdahl 1995* offer a contrastive investigation of the linguistic resources for realisation of information packaging in English and Catalan. The authors motivate their language choice observing that English and Catalan vary in their structural realisation of information packaging along two important dimensions of variation: whether the language has a malleable intonational structure—intonational plasticity—and whether the focus-ground organisation affects the constituent order. While in English the focus-ground articulation is realised mostly through prosody, Catalan makes predominant use of the

[84]Cf. Section 3.2.2 for details.

word order dimension to achieve the same. This is summarised in Figure 27.[85]

In both languages focus is associated with an A-accent. While in English links tend to be associated with a B-accent realisation (indicated in the examples below by **bold** type-setting)[86], Catalan does not possess a B-accent, here links display no particular intonational prominence—instead, what identifies them is that they are obligatorily left-detached. Front placement of links is also available, albeit optional, in English. Tails display no particular structural marking in English, other than being characteristically unaccented, but must undergo right-detachment in Catalan.

language	intonation	string order
English	• plastic • A & B accents	links optionally front
Catalan	• non-plastic • A accent	ground in detachment slots[87]

Figure 27

The table given in Figure 28 offers a comparison of the information-packaging instructions employed by these typologically different languages in realising the information structure of an utterance which is licensed by the same context. The "What about NP?" is meant to trigger a link interpretation for the NP, while the *wh*-element in the subsequent question is meant to trigger a focus interpretation in the answer. [88]

---

[85]The terms A accent and B accent originate in *Jackendoff 1972*: A accent corresponds to a simplex high pitch accent generally followed by a falling boundary tone and is associated with the focus; B accent corresponds to a complex fall-rise pitch accent and, in Vallduví's approach, is associated with, e.g., English links.

[86]As the authors admit, this description of intonational facts is an idealised picture which focuses on those aspects of intonation in English that correlate most directly with the focus-ground articulation; the use of intonation to express other pragmatic or semantic aspects of interpretation may override the default prosodic realisation of foci and links.

[87]For Catalan, the authors assume that the sentential core may contain only focal material (in addition to informationally inert weak pronouns). The basic constituent order is VOS, and such is the alignment in the sentential core. If an argument of the verb is to be interpreted as non-focal, it is necessarily **clitic-dislocated**, or, equivalently, detached away from the sentential core. Phrases associated with a link interpretation are left-detached whereas phrases associated with a tail interpretation are right-detached. The A-accent is always associated with the rightmost prosodically non-weak element within the sentential core. Cf. the examples below.

[88]The abbreviations used in the information-packaging instruction types are: l (link), f (focus), and t (tail). In the examples, the focal segments are included in square brackets with a prescript F.

In sum, informational primitives are correlated with different structural realisations in Catalan and English. In the former the structural correlates are syntactic, involving both dominance and precedence. In the latter the structural correlates are (mainly) intonational, involving two types of accents.

context	English	E	C	Catalan
What about Fred? What did **he** do?	**Fred** [$_F$ ate the BEANS.]	l-f	l-f	El Pere$_1$ [$_F$ es va menjar els FESOLS _1.]
What about Fred? What did **he** eat?	**Fred** ate [$_F$ the BEANS.]	l-t-f	l-f-t	El Pere$_1$ [$_F$ _v els FESOLS _1,] es va menjar$_V$.
What about Fred? What did **he** do to the beans?	**Fred** [$_F$ ATE] the beans.	l-f-t	l-f-t	El Pere$_1$ [$_F$ se'ls$_2$ va MENJAR _2 _1,] els fesols$_2$.
What about the beans? What happened to **them**?	a. [$_F$ FRED ate] **the beans**. b. **The beans**$_1$ [$_F$ FRED ate _1.]	f-l l-f	l-f	Els fesols$_1$ [$_F$ se'ls$_1$ va menjar _1 el PERE.]
What about the beans? What did Fred to **them**?	a. Fred [$_F$ ATE] **the beans**. b. **The beans**$_1$ Fred [$_F$ ATE _1.]	t-f-l l-t-f	l-f-t	Els fesols$_1$ [$_F$ se'ls$_1$ va MENJAR _1 _2], el Pere$_2$.
What about the beans? Who ate **them**?	a. [$_F$ FRED] ate **the beans**. b. **The beans**$_1$ [$_F$ FRED] ate _1.	f-t-l l-f-t	l-f-t	Els fesols$_1$ [$_F$ _v _1 el PERE,] se'ls$_1$ va menjar$_V$.
I know what Fred cooked. But what did he **eat**?	Fred **ate** [$_F$ the BEANS.]	t-l-f	l-f-t	De menjar$_V$ [$_F$ es va menjar$_V$ els FESOLS _1,] el Pere$_1$.
I know who cooked the beans. But then, who **ate** them?	[$_F$ FRED] **ate** the beans.	f-l-t	l-f-t	De menjar$_V$ [$_F$ se'ls$_1$ va menjar _1 el PERE,] els fesols$_1$.

Figure 28

The possibility of an optional initial placement of links in English results in the availability of alternative instruction types in some of the cases. It is interesting to observe that a more "free-word-order" language like Catalan is less flexible in the order of the informational primitives which constitute information packaging instructions, i.e. on the level of communicative segmentation. The assumptions that the representation of information structure in the grammar is indepenedent of its particular structural realisations in different languages and that information packaging and the context-independent content of sentences are separate dimensions that nevertheless constrain each other have motivated the integration of an informational component into the multidimensional constrained-based

grammar architecture of HPSG in *Engdahl and Vallduví 1994*, *Vallduví and Engdahl 1995*.

In the case of Bulgarian, both prosody, i.e. manipulating intonational structure (intonation contour), and word order, i.e. reordering sentential constituents, are relevant means of realisation of the information packaging. Similarly to the situation in English, one and the same string may be assigned different intonational phrasings in order to realise different informational interpretations. As to the surface alignment, links are obligatorily initial in Bulgarian, while tails normally follow the focal segment. An attempt to include the Bulgarian counterparts in the contrastive table from Figure 28 is illustrated in Figure 29.

The link segment, corresponding basically to the notion of "contrastive topic" in *Koktová to appear*, is the prominent thematic part, and therefore it is characterised by an "anchoring" or linking intonational contour (i.e. it contains the so-called *thematic* accent—cf. *Bolinger 1972*). The tail segment, being a non-prominent thematic part, is substantially deaccented in Bulgarian. The focal segment is characterised by a special focusing intonational contour (i.e. it contains the so-called *rhematic* accent—*Bolinger 1972*). Thus it is possible to draw a functional parallel between the A accent and B accent in Germanic languages (especially English) and the rhematic (focus-associated) accent and the thematic (link-associated) accent distinguishable in Bulgarian utterances, even though—just as it can be expected cross-linguistically—there are language-specific acoustic differences as far as the particular realisations of these contours are concerned.

Links are clause-initial in Bulgarian; and since, as mentioned above, the position at the very beginning of a sentence is a communicatively strong one, it is the intonationally more prominent part in larger link segments[89], i.e. the one bearing the thematic accent, which is placed there. In fact we have a "mirror image" with respect to the situation within larger focal segments where the most prominent part—the one bearing the (non-emphatic) rhematic accent—is by default final.

[89]This is underlined in the relevant cases below.

context	Bulgarian	B	E	C
What about Andrew? What did **he** do?	**Andrej** [$_F$ izjade boba].[90]	**l-f**	l-f	l-f
What about Andrew? What did **he** eat?	**Andrej** [$_F$ BOBA] izjade . **Andrej izjade** [$_F$ BOBA].	**l-f-t** **l-f**	l-t-f	l-f-t
What about Andrew? What did **he** do to the beans?	**Andrej** [$_F$ go$_i$ IZJADE] boba$_i$. **Andrej boba$_i$** [$_F$ go$_i$ IZJADE]. [$_F$ IZJADE go$_i$] boba$_i$ Andrej .	**l-f-t** **l-f** **f-t**	l-f-t	l-f-t
What about the beans? What happened to **them**?	**Boba$_i$** [$_F$ go$_i$ izjade Andrej].	**l-f**	f-l l-f	l-f
What about the beans? What did Andrew to **them**?	**Boba$_i$** [$_F$ go$_i$ IZJADE] Andrej. **Boba$_i$ Andrej** [$_F$ go$_i$ IZJADE]. [$_F$ IZJADE go$_i$] Andrej boba$_i$.	**l-f-t** **l-f** **f-t**	t-f-l l-t-f	l-f-t
What about the beans? Who ate **them**?	**Boba$_i$** [$_F$ ANDREJ] go$_i$ izjade . **Boba$_i$ go$_i$ izjade** [$_F$ ANDREJ]. [$_F$ ANDREJ] (go$_i$) izjade boba$_i$.	**l-f-t** **l-f** **f-t**	f-t-l l-f-t	l-f-t
I know what Andrew cooked. But what did he **eat**?	[$_F$ BOBA] izjade Andrej. **Izjade** Andrej [$_F$ BOBA].	**f-t** (!) **l-t-f**	t-l-f	l-f-t
I know who cooked the beans. But then, who **ate** them?	[$_F$ ANDREJ] izjade boba. **Izjade** go$_i$ boba$_i$ [$_F$ ANDREJ].	**f-t** (!) **l-t-f**	f-l-t	l-f-t

Figure 29

No information-packaging instructions of the type **t-f-l** or **t-l-f** (which are available in, e.g., English) can be observed in Bulgarian; the type **l-t-f** (which is claimed to be quite natural for English) is extremely rare, and can only be found when it is the syntactic head (i.e. the verb) that is interpreted as a link.

Having presented in this section the (traditionally acknowledged) importance of the communicative aspect for the analysis of Bulgarian, let us now turn to the word order domain where the functional sentence perspective is assumed to be the most dominating factor. For convenience of presentation, I shall use the communicative primitives employed in the recent information packaging approach, since these allow for explicit

[90]Even though the main stress falls on the last accentable item, it is not emphatic in the broad-focus interpretation, and hence not marked by SMALL CAPS in this example. I use SMALL CAPS—as also in other places in this study—mainly to indicate intonational prominence resulting from local emphasis.

differentiation of degrees of communicative dynamism by terminologically distinguishing the more prominent part of the theme—the link—from the less communicatively dynamic thematic part—the tail.

4.3 Clitic Replication in the Simple Bulgarian Sentence

Clitic replication of nominal material (referred to also as "clitic doubling", "pronominal reprise", "reduplication", etc.) has received extensive attention mainly from the point of view of its origin, i.e. whether it is a "balkanism"[91] or a phenomenon intrinsic to the language system of modern Bulgarian—cf., e.g., *Miletič 1937*, *Sławski 1946*, *Rusek 1963*, *Mirčev 1966*, *Cyxun 1968*, *Mladenov 1968*, *Kânčev 1972*, *Orzechowska 1976*, *Lopašov 1978*, *Asenova 1989*. An attempt to analyse the nature of the "pronominal reprise" and to interpret the model of "doubled objects" in Bulgarian in the context of the overall analytic development of this language is made in *Minčeva 1968*, *Minčeva 1969*; the stylistic aspect of the problem has also been addressed by, e.g., *Popov 1962*. An important direction of research is pursued in *Guentchéva 1994*, where the phenomenon is investigated in the context of "thèmatisation de l'objet" (object thematicisation), but *Georgieva 1974* already argues that "complement reduplication" is used in certain alignments in order to include the object into the theme and should therefore be considered a "necessary structural element in building up the functional sentence perspective of the utterance"; in particular, the author concludes:

> Udvoenite dopâlnenija predstavljavat interes za teorijata na aktualnoto členenie kato xarakterna osobenost na bâlgarskija ezik, tâj kato se pojavjavat, za da balansirat napreženieto, koeto se sâzdava meždu gramatičnoto i aktualnoto členenie.
>
> Reduplicated complements, as a typical property of Bulgarian, are interesting for the theory of functional sentence perspective inasmuch as they occur to balance the tension which arises between the grammatical and the communicative segmentation. (p.76)

There is an important line of research viewing the phenomenon of replication as grammaticalised means of syntactic function identification in a language that has lost case-marking morphology in the nominal system. For example, in *Rudin 1982*, *Rudin 1986*, "clitic reduplication" is regarded as a

[91]There are striking typological parallels to this phenomenon in the Balkan languages (e.g., in Greek, Romanian, Albanian, Macedonian, partly in Serbian and Croatian dialects), while in the other modern Slavic languages which have short pronominal forms it is not attested.

way of disambiguating case roles in potentially ambiguous sentences. This agrees with *Ivančev 1978* who argues that without such a means of object-identification in Bulgarian, stronger (or even full) grammaticalisation of the word order would be unavoidable in this Slavic language.

4.3.1 Surface Alignment and Object Thematicisation as Replication-Causing Factors

The NP typology proposed in Section 4.1.2 allows us to state the general non-replicability constraint. If certain nominal material functioning as a direct or an indirect object is used as a *categorising* or as an *identifying non-specific* description, it can never be replicated. Only nominal material which is used as an *identifying specific* description is replicable, but replication takes place under certain conditions. In my opinion, the phenomenon of replication in Bulgarian has two interrelated functions. One is to identify the syntactic category of nominal material occurring in the "prototypical" subject position in the surface ordering (i.e. clause-initially, possibly immediately preceding the verb) but functioning as direct object. Such a need results from the lack of morphological case marking which would unambiguously distinguish syntactic nominative from syntactic accusative in Bulgarian non-prepositional nominal constituents, even though the opposition full vs. short definite article in the standard written language[92] basically solves this problem in the relevant cases.[93] The other function of clitic replication is communicative in nature. It is to indicate that the replicated nominal material belongs to a thematic (ground, topic) segment in the communicative organisation of the respective utterance. These two functions interact to a different degree in each particular case. But since the object-identifying function is appropriate only for the opposition subject / direct object, this interplay can be best illustrated on declarative sentences containing at least a subject, a verb and a direct object.

The subject-verb-object linear order tends to be grammaticalised cross-linguistically—although to a different degree in the particular languages—

[92]The former is appropriate for syntactically nominative masculine NPs while the latter for syntactically accusative/oblique masculine NPs. This is, however, an artificially introduced writing convention: the opposition full vs. short definite article is neither attested in dialects nor (consistently) reflected in standard spoken language.

[93]In certain Bulgarian dialects, as well as in the closely related standard Macedonian language, the object-identifying function of the clitic replication is grammaticalised to the extent that the replication of the direct object is obligatory regardless of sentence position.

and is commonly regarded as "neutral" because it potentially allows for parallel realisation of different principles among which are

- grammatical obliqueness: from subject to object;
- thematic-role hierarchy: from agent to patient;
- intonational "principle": from minimum to maximum prominence;
- functional sentence perspective: from theme to rheme (from topic to focus).

Therefore, the SVO order with preverbal thematic subject may be regarded as default for languages in which it is attested as "structurally neutral" alignment. And it is the "overriding"—for communicative reasons—of this default interpretation that underlies the object-identifying function of the accusative replication in non-SVO utterances. In cases of potential ambiguity, an isolated written sentence is preferably interpreted as SVO rather than OVS, which is a commonly made observation—e.g., *Penčev 1984*, *Penčev 1993*, *Rudin 1986* among others. Taking into consideration the tendency for the definite article in Bulgarian to function in certain conditions as a marker of subjecthood, observed already by Ivančev, *Mayer 1988* further argues that Bulgarian has preserved an accusative clitic personal pronoun form to distinguish direct objects from subjects when an article alone cannot do it. In other words, it is because NPs in initial position are generally perceived as subjects that a direct object occurring in initial position usually needs to be replicated, as (67) illustrates.

(67)

Ženata ja risuva edin xudožnik.

woman ACC-3sg,fem draws INDEF-ART artist

'Certain artist draws the woman.'

Viewed from a more general perspective, the surface alignment can be regarded as a replication-causing factor whenever the mutual order of the subject and the direct object does not comply with the obliqueness hierarchy, i.e. whenever the subject NP linearly follows the direct object NP instead of preceding it, whereby no immediate adjacency is presupposed.

At the same time, the communicative aspect of clitic replication in Bulgarian is very consistent and dominating. In the overall communicative organisation of the respective utterance, clitic replicants are indicators of the thematic (ground) status of the nominal material they replicate, even though clitics themselves are informationally inert (i.e. communicatively insignificant, as argued in, e.g., *Avgustinova and Oliva 1995b*) elements.

With pronominal-NP objects, clitics need not contribute to distinguishing the subject from the object. Full pronominal forms are morphologically marked for case[94] and are usually unambiguous in this respect. The contingent clitic replicants then have a purely thematicising function. In particular, replicated full pronouns are used as a (part of the) theme / ground, while full pronouns with no replication—as a (part of the) rheme / focus—cf. also *Nicolova 1986* (p.53). There is an apparent tendency for grammaticalisation of the thematicising function of the replication in such constructions—cf. (68a); moreover, the lack of replication is then a means of rhematisation of the respective pronominal object—cf. (68b).

(68)

a. *Nas ni pokani Kiril.*
us ACC-1pl invited Cyril
'CYRIL invited us.'

b. *NAS pokani Kiril.*
us invited Cyril
'Cyril invited US.'

The observable clitic replication, as well as some related intonational-pattern changes, can be systematically accounted for if the communicative organisation of Bulgarian utterances is viewed as an interplay of the head-lexeme specific obliqueness hierarchy and the particular surface alignment of the verb and its subcategorised dependants (i.e. subject and complements). Aiming at a classification reflecting this interplay, which—in my opinion—supplies considerable explanatory resources, I adopt the terminology of the information packaging approach.

Let us once more recall at this point that positioning of free modifiers (i.e. adjuncts) is beyond the scope of the present work, inasmuch as these cannot be cliticised in Bulgarian[95] and are not involved in the replication phenomena which the present study is concerned with.

---

[94]In fact, formally distinct non-nominative case forms are available only with pronominal nouns in the masculine gender denoting persons—cf., e.g., the discussion in *Pašov 1974*.

[95]Among the complements, the prepositional ones that are different form the syntactically dative *na*-NPs, as well as the verbal ones cannot be cliticised either.

4.3.2 Interplay of Obliqueness Hierarchy, Constituent Order and Replication in Communicative Organisation of Utterances

In this section the role of clitic replication in the information structure of the simple Bulgarian clause is investigated. The discussion is based on a classification of the various possible communicative structurings (in terms of information packaging) of different constituent-order variants of the two sample sentences given in (69). For clarity, a simplified model of accentuation is assumed allowing for only one nuclear (main) stress in an utterance, which, under various circumstances, can be either normal or emphatic (in the terminology introduced in Section 4.2) but is always bound to the focus. The proposed classification takes into consideration, on the one hand, the interplay of three factors: the lexeme-specific obliqueness ordering, the eventual surface constituent order, and the contingent clitic replication, and on the other hand, reflects the related changes in the (emphatic) accent placement.

From the point of view of information packaging, the *communicatively unmarked word order* (CUWO) in a declarative clause presupposes either an all-focus interpretation or an alignment where a link segment precedes the focal one. At the same time, *communicatively marked word order* (CMWO) in declarative clauses generally amounts to alignments in which rhematic (focal) elements linearly precede thematic (ground) ones. In particular, an alignment where a focal segment precedes the tail part of the ground is the most typical instance of a CMWO in declarative clauses. On the other hand, also the case of "split ground" where the focus is interlocated between a clause-initial link segment and a clause-final tail segment in declarative clauses is to be regarded as an instance of CMWO, since it obviously matches, even though only partly, the rheme-preceding-theme template. In this case, it seems more appropriate to speak of *emphatic word order* (EWO), since such a pattern always presupposes local emphasis (realised by emphatic stress) on the focal segment, which licenses only narrow-focus interpretations. Also a special case of a CUWO has to be distinguished which basically matches the theme-preceding-rheme template but will be called *parenthetic word order* (PWO), since it results from an information structuring where the constituent syntactically heading the utterance—the verb (complex)—is interpreted as link, bearing thus the thematic (linking) accent, and is immediately followed by the strongly

deaccented (intonationally "parenthesised") tail part of the ground[96], and only then by the focal segment containing the main (focal) accent of the utterance. Summing up, the following four types of surface alignment in terms of information packaging are available for simple Bulgarian declarative clauses:

CUWO (communicatively unmarked word order): ***all-focus*** or ***link / focus***

PWO (parenthetic word order): ***head-link / tail / focus***

CMWO (communicatively marked word order): ***focus / tail***

EWO (emphatic word order): ***link / focus / tail***

The mechanism of replication in Bulgarian is most clearly observable in simple sentences headed either by a transitive verb, and hence containing a subject NP and a direct object NP—(69a), or by a verb requiring a subject NP and two NP-complements—a direct object and an indirect object—(69b). The nominal material in the chosen examples is potentially replicable: all NPs are used as *identifying specific definite descriptions* of the respective objects (cf. Section 4.1.2 for these concepts). The direct objects in (69a-b) are replicable by the accusative feminine singular clitic *ja* and the indirect object in (69b)—by the dative plural clitic *im*.

(69)

a. *Andrej vidja kartinata.*

Andrew saw picture_def.art

'Andrew saw the picture.'

b. *Ivan izprati kuklata na decata.*

John sent doll_def.art to children_def.art

'John sent the doll to the children.'

As argued in Section 4.3.1, the replication of a **non-pronominal** direct object NP has twofold motivation if the mutual positioning of the subject and the direct object in the clause does not correspond to the head-specific obliqueness ordering of grammatical relations. Along with imposing ground interpretation on the replicated nominal material, it resolves the potential syntactic-category ambiguity by explicitly specifying that the replicated material functions as (syntactically accusative) direct object. Let us also recall that a **pronominal** direct or indirect object in Bulgarian is replicated by a clitic whenever it is supposed to be interpreted as thematic. Interestingly, a clitic replicant of a thematic pronominal object is observed

[96] All words constituting the tail segment in the PWO case tend to be pronounced with low pitch, low amplitude, and—perhaps most characteristically—with syllables of relatively short duration.

even in those alignments where it appears to be optional if the respective NP is not realised as a personal pronoun. Variations of the sample sentence *Ivan pokani neja*. ('John invited her.'), involving a pronominal direct object NP (*neja*) which is replicable by the accusative feminine singular clitic *ja*, will provide an illustration of this contrast in the relevant cases below.

It is further useful to distinguish between **anaphoric** and **cataphoric** replication. Anaphoric replication can be observed in all cases where the replicated nominal material linearly precedes its clitic replicant, while cataphoric replication takes place whenever the clitic replicant linearly precedes the nominal material it replicates. Since all replicants are verbal clitics and as such are included in the verb-complex constituent (cf. Chapter 3), their linear positioning with respect to the nominal material they replicate depends on the placement of the verb (complex) in the clause. Thus, another useful criterion can be introduced in the description of the replication phenomena on the clausal level, namely, whether the respective direct or indirect object NP is preverbal or postverbal. In the former case, the observed replication will be anaphoric, and in the latter—cataphoric.

Let us introduce some notational conventions to be used in the following data representation.

- Numbers in brackets refer to the subcategorised elements and correspond to the place the particular element occupies in the head's subcategorization list which (in HPSG style) models the obliqueness hierarchy of grammatical relations. In particular, **[1]** refers to the subject, **[2]** to the direct object, and **[3]** to the indirect object.
- The symbol '<<' indicates obliqueness order, with the less oblique element standing to the left of it.
- The symbol '<' indicates linear precedence in surface alignment.
- The symbol '/' marks borders between informational segments under the particular communicative interpretation of a given surface alignment.
- The thematic (or ground) parts of the sentences, i.e. the link and the tail, are given in standard type-setting.
- The rhematic (or focus) parts of the sentences are set in *italics*.

- The constituents characterised by (or containing) an extraordinary stress, and thus exclusively representing the entire focal segment, are given in UPPER CASE letters.
- Clitic replicants of nominal material are set **bold**.
- Parentheses '()' indicate optionality, e.g., of the clitic included in them. The context to the respective utterance is also included in parentheses.

.c.4.3.2.1 S-V-O Sentence Type: the Data

- **Surface Alignment Coinciding with Obliqueness Ordering**

[1] < [2] & [1] << [2]

No object-identifying function can be claimed for any type of replication in this case. The fact that there is a "strongly preferred reading" for the linearly earlier NP to be the subject and for the NP linearly following it to be the direct object is actually due to the fact that the relative obliqueness hierarchy is thus directly reflected in the surface alignment.

SUBJ - V - OBJ

Replication of the postverbal direct object NP is not really needed here; if it occurs at all, then only with object-thematicising function, as the obligatoriness of the clitic in the examples with a thematic pronominalised object unambiguously illustrates.[97]

all-focus

(Kakvo stana tazi sutrin?)

<1>[98] *Andrej vidja kartinata.*

link-focus

(Štom zagovorixme za Andrej, kaži kakvoto znaeš.)

<2> Andrej / *vidja kartinata.*

---

[97]The examples of "multiple foci", possibly interpreted as a single "complex focus", presuppose that the surface alignment of constituents coincides with the obliqueness ordering, e.g.,

Koj kakvo vidja? 'Who saw what?'

ANDREJ vidja KARTINATA.

[98]Each of the examples below is given a **<reference number>** for convenient identification and co-reference in the following discussion.

(Setix se za Andrej. Kakvo li e vidjal tazi sutrin, če beše tolkova razvâlnuvan?)

<3> (Andrej) vidja / *kartinata*.

link-focus-tail

(Setix se za Andrej. Trjabva da mu pokažem kartinata.)

<4> Andrej / ***(ja)*** *VIDJA* / kartinata.

(Setix se za Ivan. Trjabva da razberem dali e predpriel nešto po otnošenie na Zornica.)

<4'> Ivan / ***ja*** *POKANI* / neja.

The function of the cataphoric accusative clitic is to thematicise the direct object NP with respect to the verb which has to be interpreted as constituting the entire focus in this EWO. And it is the emphatic stress on the verb that makes the replication of the non-pronominal direct object NP optional here, since its less prominent communicative status is thus already revealed by the intonation.

focus-tail

(Ne možax da razbera koj e vidjal kartinata.)

<5> *ANDREJ* / **(ja)** vidja kartinata.

(Ne možax da razbera koj e pokanil Zornica.)

<5'> *IVAN* / **ja** pokani neja.

Occurring in its prototypical position, the subject NP in this case has to be interpreted as the exclusive focus of the utterance. The observable thematicisation via the accusative clitic of the postverbal direct object NP (which, in fact, also occurs in its prototypical position) takes place in order to make the communicative segmentation in this CMWO unambiguous. On the other hand, the intonational contour with local emphasis on the subject and substantially reduced prominence on the rest of the utterance makes the need of cataphoric replication not really pressing, and hence especially in the non-pronominal example, optional.

SUBJ - OBJ - V

link-focus

(Kato govorim za Andrej, kakvo stana s kartinata, kojato štjaxme da mu pokazvame?)

<6> Andrej kartinata / ***ja*** *VIDJA*.

The fact that the head is more prominent than its modifiers and constitutes the entire focus in this CUWO requires special indication here. The obligatory anaphoric replication of the direct object NP results from the need to explicitly thematicise the replicated nominal material with respect to the (emphatically stressed) verb.

link-focus-tail

(Njama zašto da se trevožim, ako tova, koeto e vidjal Andrej, e bronzovata statuetka. No kakâv e problemât?)

<7> Andrej / *KARTINATA* / vidja.

V - SUBJ - OBJ

focus-tail

(Dokolkoto znam, iskaxte da pokažete na Andrej kartinata. No štom ne govori za neja, predpolagam, če ošte ne ja e vidjal.)

<8> *VIDJA* ***ja*** / Andrej kartinata.

In this CMWO the obligatory cataphoric replication is also supposed to support the opposition between a rhematic verbal head and its thematic complement.

link(head)-tail-focus

(Za Andrej razbrax, če e vidjal nešto, koeto mnogo go e zainteresuvalo.)

<9> Vidja / Andrej / *kartinata*.

- **Surface Alignment Not Coinciding with Obliqueness Ordering**

[2] < [1] & [1] << [2]

The crucial assumption in the present analysis, namely that the obliqueness hierarchy and the surface alignment interact, allows us to make the general claim that whenever the mutual positioning of the constituents in the clause diverges from the obliqueness ordering of the grammatical relations that is appropriate for the respective governing verb, there arises a need for syntactic-category disambiguation between the subject and the direct object.

OBJ - V - SUBJ

This is the typical case where the anaphoric replication of the direct object NP has an object-identifying function which even dominates over the thematicising one, as the possibility of replicating focal nominal material suggests in the CMWO example below.

link-focus

(Ne možax da razbera koj e vidjal kartinata.)

<10> Kartinata **ja** vidja / *Andrej*.

link-focus-tail

(Kâde e Andrej? Ne e li vreme da mu kažem pone, če e pristignala kartinata?)

<11> Kartinata / ***ja*** *VIDJA* / Andrej.

focus-tail

(Dokolkoto razbiram, Andrej e vidjal veče nešto, no kakvo?)

<12> *KARTINATA* / **(ja)** vidja Andrej.

(Dokolkoto razbiram, Ivan e pokanil njakogo, no kogo?)

<12'> *NEJA* / pokani Ivan.

Since the clitic optionally replicating a rhematic direct object in the last case can only serve as a pure syntactic-function indicator, it will not occur in the parallel example with the pronominal direct-object NP. Syntactically, pronominal-NP objects are unambiguously identifiable due to the case distinctions kept in the Bulgarian pronominal system.[99]

OBJ - SUBJ - V

link-focus

(Njamam nikakva predstava za situacijata okolo kartinata. Ti znaeš li nešto?)

<13> Kartinata / *Andrej* ***ja*** *vidja*.

link-focus-tail

(Kakvo stana s kartinata? Koj ja vidja?)

<14> Kartinata / *ANDREJ* / **ja** vidja.

Since both object-identifying and thematicising functions of the replication are relevant for the preceding two cases, the presence of the anaphoric accusative clitic is obligatory.

[99]However, there seem to be rather strict constraints on the replicable nominal material constituting the direct object NP in this particular case. While it can be replicated without problems if used as an identifying specific definite description, which is the case in the given example as well as in (i) below, this is impossible if the direct object NP is used as an identifying indefinite description, as (ii) illustrates, even though such an NP is in principle replicable—cf. Section 4.1.3.

(i) *TAZI KARTINA (ja) donese Andrej.*
this picture (ACC-3sg,fem) brought Andrew
'Andrew brought this PICTURE.'

(ii) *EDNA KARTINA (*ja) donese Andrej.*
INDEF-ART picture (*ACC-3sg,fem) brought Andrew
'A picture Andrew brought.'

V - OBJ - SUBJ

link-focus

(Njakoj e vidjal kartinata, no koj?)

<15> Vidja **(ja)** kartinata / *ANDREJ*.

(Njakoj e pokanil Zornica, no koj?)

<15'> Pokani **ja** neja / *IVAN*.

The local emphasis on the subject NP in this CUWO allows for unproblematic thematic interpretation of the rest of the utterance, and thus results—in the non-pronominal case—in optionality of the thematicising cataphoric clitic.

link(head)-tail-focus

(Za kartinata razbrax, če njakoj ja e vidjal, no ne znam koj.)

<16> Vidja **ja** / kartinata / *ANDREJ*.

In the last case, the thematicising function of the accusative clitic becomes crucially relevant, since the replicated nominal material is the least prominent part of the ground, i.e. it has to be interpreted as tail in this PWO. Hence, the cataphoric replication is obligatory here.

link-focus-tail

(Andrej maj e vidjal nešto. Znaeš li kakvo točno?)

<17> Vidja / *KARTINATA* / Andrej.

focus-tail

(Xajde da povikame Andrej. Nali štjaxme da mu pokazvame kartinata?)

<18> *VIDJA **ja*** / kartinata Andrej.

The cataphoric accusative clitic is needed in the last CMWO-example mainly for making the contrast between the rhematic verb and its thematic complement explicit, since prototypically it is the latter that would be more prominent than the former.

4.3.2.2 Clitic Replication in the S-V-O Sentence Type

If the data presented above are classified according to the type of the contingent clitic replication, this would give us a clear picture of the regularities observable in the S-V-O sentence type.[100]

[100]For convenience, the corresponding reference numbers will be given after the respective utterances.

Almost all relevant communicative segmentations where clitic replication can or even must take place exhibit a narrow focus which is usually emphatically marked by prosodic means.

A. Obligatory replication

A.1. Preverbal direct object NP

The ***obligatory anaphoric replication*** both identifies and thematicises the direct object in (70b-e), though pure thematicising, as in (70a), is also observable.

(70)

surface realisation	communicative interpretation of RNM	type of alignment
a. Andrej kartinata / ***ja*** *VIDJA*. **<6>** b. Kartinata **ja** vidja / *Andrej*. **<10>** c. Kartinata / *Andrej* ***ja*** *vidja*. **<13>**	link	CUWO
d. Kartinata / ***ja*** *VIDJA* / Andrej. **<11>** e. Kartinata / *ANDREJ* / **ja** vidja. **<14>**	link	EWO

A.2. Postverbal direct object NP

The ***obligatory cataphoric replication*** has a mainly thematicising function—cf. (71b-c)—but it also identifies the direct object in (71a). It is noteworthy that in the relevant examples the verb (complex) linearly precedes all its subcategorised modifications.

(71)

surface realisation	communicative interpretation of RNM	type of alignment
a. *VIDJA* ***ja*** / kartinata Andrej. **<18>** b. *VIDJA* ***ja*** / Andrej kartinata. **<8>**	tail	CMWO
c. Vidja **ja** / kartinata / *ANDREJ*. **<16>**	tail	PWO

B. Optional replication

As mentioned several times before, the optionality of the thematicising replication is usually due to the availability of an emphatic stress exclusively marking the narrow focus and thus allowing for unambiguous thematic interpretation of the rest of the utterance.

B.1. Preverbal direct object NP

Since in (72) the replicated material is interpreted as focus, the ***optional anaphoric replication*** exhibits exclusively object-identifying function, i.e. this is the only case where the accusative clitic has to be regarded as a pure syntactic-function indicator.

(72)

surface realisation	communicative interpretation of RNM	type of alignment
KARTINATA / **(ja)** vidja Andrej. **<12>**	focus	CMWO

B.2. Postverbal direct object NP

The ***optional cataphoric replication*** in (73a) would both identify and thematicise the direct object, while only a thematicising function can be claimed for the clitic possibly occurring in (73b-c).

(73)

surface realisation	communicative interpretation of RNM	type of alignment
a. Vidja **(ja)** kartinata / *ANDREJ*. **<15>**	link	CUWO
b. *ANDREJ* / **(ja)** vidja kartinata. **<5>**	tail	CMWO
c. Andrej / ***(ja)*** *VIDJA* / kartinata. **<4>**	tail	EWO

So, the following generalisations can be made with respect to the particular type of alignment (or, equivalently, information-packaging instruction type) and the contingent clitic replication of the direct-object NP in the S-V-O sentence type:

- In CUWO the (optionally or obligatorily) replicated nominal material is always a link.
- In CMWO there can be both obligatory and optional cataphoric, but just optional anaphoric replication of tail nominal material.
- In EWO, anaphoric replication of link nominal material can be obligatory or optional; while cataphoric replication of tail nominal material is only optional.
- In PWO there always is obligatory cataphoric replication of tail nominal material.

4.3.2.3 S-V-O1-O2 Sentence Type: the Data

Even though the constituent order combinatorics in a double object construction is much richer than in a sentence containing, apart from the verb (complex), only a subject NP and a direct object NP, the regularities stated for the S-V-O sentence type are basically also relevant for the S-V-O1-O2 sentence type.[101] It is important to bear in mind that any replication by a dative clitic can be only thematicising, i.e. it imposes either link or tail interpretation on the indirect-object NP it replicates.

- **Surface Alignment Coinciding with Obliqueness Ordering**

[1] < [2] < [3] & [1] << [2] << [3]

SUBJ - V- OBJ1 - OBJ2

all-focus

(Kakvo stana tazi sutrin?)

<19> *Ivan izprati kuklata na decata.*

link-focus

(Štom zagovorixme za Ivan, kaži kakvoto znaeš.)

<20> Ivan / *izprati kuklata na decata.*

(Dokolkoto si spomnjam, Ivan se kaneše na njakogo da izprašta porcelanovata kukla.)

<21> Ivan **(ja)** izprati kuklata / *na decata.*

link-focus-tail

(Setix se za decata. Ivan izprati li im nešto?)

<22> Ivan **(im)** izprati / *KUKLATA* / na decata.

(Porcelanovata kukla bezsporno vpečatli decata. Kakvo napravi s neja Ivan.)

<23> Ivan / ***(im) (ja)*** *IZPRATI* / kuklata na decata.

---

[101] In this sentence type, "multiple foci" or "complex focus" interpretations also presuppose a surface alignment coinciding with the obliqueness ordering, e.g.,

(i) *Koj kakvo na kogo e izpratil?* 'Who sent what to whom?'
IVAN izprati KUKLATA na DECATA.

(ii) *Na kogo kakvo e izpratil Ivan?* 'What did John send to whom?'
Ivan izprati KUKLATA na DECATA.

(iii) *Koj na kogo e izpratil kuklata?* 'Who sent the doll to whom?'
IVAN izprati kuklata na DECATA.

focus-tail

(Seti li se njakoj da izprati porcelanovata kukla na decata?)

<24> *IVAN* / **(im) (ja)** izprati kuklata na decata.

SUBJ - OBJ1- V - OBJ2

link-focus

(Da si viždal Ivan? Ne moga da namerja porcelanovata kukla, a dokolkoto si spomnjam, toj iskaše da ja izprašta na njakogo.)

<25> Ivan kuklata **ja** izprati / *na decata.*

(Znaeš li kakvo e napravil Ivan s porcelanovata kukla?)

<26> Ivan kuklata / ***ja*** *izprati na decata.*

link-focus-tail

(Kakvo izprati Ivan na decata - knigata ili kuklata?)

<27> Ivan / KUKLATA / (**im**) izprati na decata.

(Setix se za porcelanovata kukla, kojato iskaxme da polučat decata. Podoziram obače, če Ivan vse ošte ništo ne e napravil. Ili greša?)

<28> Ivan kuklata / ***(im) ja*** *IZPRATI* / na decata.

SUBJ - OBJ1- OBJ2 - V

link-focus

(Znaeš li kak sa polučili decata tazi kukla ot Ivan?)

<29> Ivan kuklata na decata / ***im ja*** *IZPRATI.*

link-focus-tail

(Kakvo stana s porcelanovata kukla, kojato Ivan se kaneše da izprati na sestra si? Da ne si e promenil namerenieto?)

<30> Ivan kuklata / *NA DECATA* / **ja** izprati.

V - SUBJ - OBJ1- OBJ2

link-focus

(Ne moga da namerja porcelanovata kukla. Ivan trjabva da ja e izpratil na njakogo i vie najverojatno znaete na kogo.)

<31> Izprati **ja** Ivan kuklata / *na decata.*

link-focus-tail

(Razbrax, če Ivan e izpratil nešto na decata, no ne znam kakvo.)

<32> Izprati **im** Ivan / *KUKLATA* / na decata.

focus-tail

(Decata mnogo xaresaxa porcelanovata kukla. Kakvo napravi Ivan s neja?)

<33> *IZPRATI* ***im ja*** / Ivan kuklata na decata.

- **Surface Alignment Not Coinciding with Obliqueness Ordering**

[1] < [3] < [2] & [1] << [2] << [3]

SUBJ - V - OBJ2 - OBJ1

link-focus

(Znaete li kakvo točno e izpratil Ivan na decata?)

<34> Ivan **(im)** izprati na decata / *kuklata*.

link-focus-tail

(Setix se za porcelanovata kukla. Na kogo vsâšnost ja izprati Ivan - na decata ili na sestra si?)

<35> Ivan **(ja)** izprati / *NA DECATA* / kuklata.

(Decata mnogo xaresaxa porcelanovata kukla. Dali Ivan veče e napravil nešto, za da gi zaradva s neja?)

<36> Ivan / ***(im) (ja)*** *IZPRATI* / na decata kuklata.

focus-tail

(Čudja se koj se e setil da izprati na decata kuklata.)

<37> *IVAN* / **(im) (ja)** izprati na decata kuklata.

(Kakvo stana s kuklata, kojato decata tolkova xaresaxa?)

<38> *Ivan* ***im ja*** *izprati* / na decata kuklata.

SUBJ - OBJ2 - V - OBJ1

link-focus

(Kakvo napravi Ivan za decata? Da si čul nešto tazi sutrin?)

<39> Ivan na decata / ***(im)*** *izprati kuklata.*

(Izpratil li e Ivan nešto na decata ?)

<40> Ivan na decata **(im)** izprati / *kuklata*.

link-focus-tail

(Setix se za porcelanovata kukla. Dokolkoto si spomnjam, Ivan ja izprati na njakogo. Na decata ili na sestra si?)

<41> Ivan / *NA DECATA* / **(ja)** izprati kuklata.

(Kakvo napravi Ivan, za da zaradva decata s ljubimata im kukla?)

<42> Ivan na decata / ***im (ja)*** *IZPRATI* / kuklata.

SUBJ - OBJ2 - OBJ1 - V

link-focus

(Kak sa polučili decata tazi kukla ot Ivan?)

<43> Ivan na decata kuklata / ***im ja*** *IZPRATI*.

link-focus-tail

(Ivan e izpratil veče nešto na decata. Vse pak mi e interesno dali se e sprjal na knigata ili na kuklata.)

<44> Ivan na decata / *KUKLATA* / **(im)** izprati.

V- SUBJ - OBJ2 - OBJ1

link-focus

(Kolokoto do izpraštaneto, Ivan šteše da izprašta nešto na decata, no ne znam kakvo.)

<45> Izprati **(im)** Ivan na decata / *kuklata*.

link-focus-tail

(Interesno, na kogo li e izpratil Ivan kuklata?)

<46> Izprati **(ja)** Ivan / *NA DECATA* / kuklata.

focus-tail

(Kakvo napravi Ivan s porcelanovata kukla, kojato decata tolkova xaresaxa?)

<47> *IZPRATI **im ja*** / Ivan na decata kuklata.

link(head)-tail-focus

(Za Ivan razbrax, če e izpratil nešto na decata, no ne znam kakvo.)

<48> Izprati **im** / Ivan na decata / *kuklata*.

[2] < [1] < [3] & [1] << [2] << [3]

V - OBJ1 - SUBJ - OBJ2

link-focus-tail

(Kato govorim za porcelatovata kukla, čudja se koj li se e setil da ja izprati na decata.)

<49> Izprati **(im) ja** kuklata / *IVAN* / na decata.

(Ivan vinagi misli za decata. Sigurno i tozi pât im e izpratil nešto ot ljubimite im igrački.)

<50> Izprati **(im)** / *KUKLATA* / Ivan na decata.

focus-tail

(Setix se za porcelanovata kukla, kojato Ivan iskaše da podari na decata, no ne znam kakvo stana.)

<51> *IZPRATI **im ja*** / kuklata Ivan na decata.

link(head)-tail-focus

(Za porcelanovata kukla razbrax, če Ivan ja e izpratil na njakogo, no ne znam na kogo.)

<52> Izprati **ja** / kuklata Ivan / *na decata*.

OBJ1 - SUBJ - V - OBJ2

link-focus

(Dokolkoto si spomnjam, Ivan se kaneše da izpratšta na njakogo porcelanovata kukla, no na kogo li?)

<53> Kuklata Ivan **ja** izprati / *na decata.*

(Njama ja porcelanovata kukla. Kakvo e napravil Ivan s neja?)

<54> Kuklata Ivan / ***ja*** *izprati na decata.*

(Štom zagovorixme za porcelanovata kukla, kaži kakvoto znaeš.)

<55> Kuklata / *Ivan* ***(ja)*** *izprati na decata.*

link-focus-tail

(Kato govorim za kuklata, koj ja izprati na decata - Rosica ili Ivan?)

<56> Kuklata / *IVAN* / **(im) ja** izprati na decata.

(Ivan znae, če decata mnogo xaresvat porcelanovata kukla. Kakvo napravi toj s neja?)

<57> Kuklata Ivan / ***im ja*** *IZPRATI* / na decata.

OBJ1 - V - SUBJ - OBJ2

link-focus

(Kato govorim za porcelanovata kukla, na kogo ja izprati Ivan?)

<58> Kuklata **ja** izprati Ivan / *na decata.*

link-focus-tail

(Ivan znae, če decata mnogo xaresvat porcelanovata kukla. No kakvo li e stanalo s neja? Nikâde ne ja viždam)

<59> Kuklata / ***im ja*** *IZPRATI* / Ivan na decata.

(Togava vsički znaexa, če decata mnogo xaresvat porcelanovata kukla, zatova ne moga da precenja koj im ja e izpratil.)

<60> Kuklata **(im) ja** izprati / *IVAN* / na decata.

focus-tail

(Kakvo li im e izpratil Ivan na decata? Slučajno da znaeš?)

<61> *KUKLATA* / **(im) (ja)** izprati Ivan na decata.

OBJ1 - SUBJ - OBJ2 - V

link-focus

(Setix se za porcelanovata kukla, kojato Ivan iskaše da podari na decata. Njamam predstava obače kakvo stana.)

<62> Kuklata Ivan na decata / ***im ja*** *IZPRATI.*

link-focus-tail

(Kato govorim za porcelanovata kukla, mnogo me interesuva na kogo ja izprati Ivan.)

<63> Kuklata Ivan / *NA DECATA* / **(ja)** izprati.

[2] < [3] < [1] & [1] << [2] << [3]

V - OBJ1 - OBJ2 - SUBJ

link-focus

(Bjaxa mi kazali koj e izpratil porcelanovata kukla na decata, no sega ne moga da si spomnja.)

<64> Izprati **(im) ja** kuklata na decata / *Ivan*.

link-focus-tail

(Setix se decata. Kakvo li im e izpratil Ivan?)

<65> Izprati **(im)** / *KUKLATA* / na decata Ivan.

(Setix se porcelanovata kukla. Na kogo li ja e izpratil Ivan?)

<66> Izprati **ja** kuklata / *NA DECATA* / Ivan.

focus-tail

(Setix se za porcelanovata kukla, kojato iskaxme da podarim na decata. Podoziram obače, če Ivan vse ošte ništo ne e napravil. Ili greša?)

<67> *IZPRATI* **im ja** / kuklata na decata Ivan.

link(head)-tail-focus

(Za porcelanovata kukla razbrax, če njakoj ja e izpratil na decata, no ne znam koj.)

<68> Izprati **im ja** / kuklata na decata / *Ivan*.

OBJ1 - OBJ2 - SUBJ - V

link-focus

(Setix se za porcelanovata kukla, kojato tolkova mnogo se xaresa na decata, no ne znam kakvo stana sled tova.)

<69> Kuklata na decata / *Ivan **im ja** izprati*.

link-focus-tail

(Kato govorim za kuklata, kojato decata polučxa za Koleda, stana li jasno ot kogo e izpratena?)

<70> Kuklata na decata / *IVAN* / **(im) ja** izprati.

OBJ1 - OBJ2 - V - SUBJ

link-focus

(Njakoj e izpratil porcelanovata kukla na decata i sega se opitvam da razbera koj.)

<71> Kuklata na decata **(im) ja** izprati / *Ivan*.

link-focus-tail

(Setix se za porcelanovata kukla, na kogo ja izprati Ivan - na decata ili na sestra si?)

<72> Kuklata / *NA DECATA* / **ja** izprati Ivan.

(Dokolkoto znam, Ivan samo e pokazal porcelanovata kukla na decata. Da ne bi da e predpriel ošte nešto sled tova?)

<73> Kuklata na decata / ***im ja*** *IZPRATI* / Ivan.

OBJ1 - V - OBJ2 - SUBJ

link-focus

(Znae li se koj e izpratil na decata porcelanovata kukla?)

<74> Kuklata **(im) ja** izprati na decata / *Ivan*.

link-focus-tail

(Ivan edva li e zabravil, če decata xaresvat porcelanovata kukla. Da znaeš kakvo konkretno e napravil toj s neja?)

<75> Kuklata / ***im ja*** *IZPRATI* / na decata Ivan.

(Znae li se na kogo e izpratil Ivan porcelanovata kukla?)

<76> Kuklata **ja** izprati / *NA DECATA* / Ivan.

focus-tail

(Znae li se točno kakvo e izpratil na decata Ivan?)

<77> *KUKLATA* / **(im) (ja)** izprati na decata Ivan.

[3] < [1] < [2] & [1] << [2] << [3]

V - OBJ2 - SUBJ - OBJ1

link-focus

(Kato stana duma za decata, razbrax, če Ivan im e izpratil nešto. Mnogo mi e interesno kakvo imenno.)

<78> Izprati **(im)** na decata Ivan / *kuklata*.

link-focus-tail

(Kolkoto do kuklata, razbrax, če njakoj ja e izpratil na decata, no ne znam koj.)

<79> Izprati **(im) ja** na decata / *IVAN* / kuklata.

(Razbrax, če Ivan e izpratil na njakogo porcelanovata kukla. No na kogo li?)

<80> Izprati **ja** / *NA DECATA* / Ivan kuklata.

focus-tail

(Kakvo točno napravi Ivan s porcelanovata kukla?)

<81> *Izprati **ja** na decata* / Ivan kuklata.

link(head)-tail-focus

(Ot decata razbrax, če Ivan im e izpratil nešto, no te ne mi kazaxa kakvo.)

<82> Izprati **im** / na decata Ivan / *kuklata*.

OBJ2 - SUBJ - V - OBJ1

link-focus

(Kakvo novo nauči za decata?)

<83> Na decata / *Ivan **(im)** izprati kuklata.*

(Kakvo napravi Ivan za decata?)

<84> Na decata Ivan / ***(im)*** *izprati kuklata.*

(Kakvo izprati Ivan na decata?)

<85> Na decata Ivan **(im)** izprati / *kuklata*.

link-focus-tail

(Kato govorim za decata, znae li se koj im izprati porcelanovata kukla? Ivan ili Rosica?

<86> Na decata / *IVAN* / **(im) (ja)** izprati kuklata.

(Kolkoto do decata, Ivan samo im pokaza porcelanovata kukla, kogato bjaxa tuk. Kakvo stana posle?)

<87> Na decata Ivan / ***im ja*** *IZPRATI* / kuklata.

OBJ2 - V - SUBJ - OBJ1

link-focus

(Sigurna bjax, če Ivan e izpratil nešto na decata, no ne si spomnjax kakvo.)

<88> Na decata **(im)** izprati Ivan / *kuklata*.

link-focus-tail

(Kolkoto do decata, Ivan trjabva veče da e predpriel nešto s porcelanovata kukla, kojato te tolkova mnogo xaresaxa.)

<89> Na decata / ***im ja*** *IZPRATI* / Ivan kuklata.

focus-tail

(Setix se za porcelanovata kukla. Na kogo ja izprati Ivan - na decata ili na sestra si?)

<90> *NA DECATA* / **(ja)** izprati Ivan kuklata.

OBJ2 - SUBJ - OBJ1 - V

link-focus

(Decata bjaxa kazali na Ivan, če mnogo xaresvat porcelanovata kukla. Znaeš li kakvo stana sled tova?)

<91> Na decata Ivan kuklata / ***im ja*** *IZPRATI*.

link-focus-tail

(Ivan e izpratil nešto na decata, no ne znam dali kuklata ili knigata.)

<92> Na decata Ivan / *KUKLATA* / **(im)** izprati.

[3] < [2] < [1] & [1] << [2] << [3]

V - OBJ2 - OBJ1 - SUBJ

link-focus

(Stana li jasno koj e izpratil na decata porcelanovata kukla?)

<93> Izprati **(im) ja** na decata kuklata / *Ivan*.

link-focus-tail

(Ot decata razbrax, če Ivan im e izpratil nešto, no ne stana jasno dali kuklata ili knigata.)

<94> Izprati **(im)** na decata / *KUKLATA* / Ivan.

(Razbrax, če tazi sutrin Ivan e izpratil na njakogo porcelanovata kukla. Na kogo li?)

<95> Izprati **ja** / *NA DECATA* / kuklata Ivan.

focus-tail

(Rosica kaza na Ivan, če decata mnogo xaresvat porcelanovata kukla. Toj kak reagira na tova?)

<96> *IZPRATI **im ja*** / na decata kuklata Ivan.

link(head)-tail-focus

(Ot decata razbrax, če njakoj im e izpratil porcelanovata kukla, no ne znam koj.)

<97> Izprati **im ja** / na decata kuklata / *Ivan*.

OBJ2 - OBJ1 - SUBJ - V

link-focus

(Znam samo, če decata mnogo xaresaxa porcelanovata kukla.)

<98> Na decata kuklata / *Ivan **im ja** izprati*.

link-focus-tail

(Decata mi pokazaxa porcelanovata kukla, no ne mi objasnixa koj im ja e izpratil?)

<99> Na decata kuklata / *IVAN* / **(im) ja** izprati.

OBJ2 - OBJ1 - V - SUBJ

link-focus

(Opitvam se da si spomnja koj izprati na decata porcelanovata kukla.)

<100> Na decata kuklata **(im) ja** izprati / *Ivan*.

link-focus-tail

(Kakvo napravi Ivan za decata, deto xaresaxa porcelanovata kukla?)

<101> Na decata kuklata / ***im ja*** *IZPRATI* / Ivan.

(Kato govorim za decata, knigata ili kuklata im izprati Ivan?)

<102> Na decata / *KUKLATA* / **(im)** izprati Ivan.

OBJ2 - V - OBJ1 - SUBJ

link-focus

(Opitvam se da si spomnja koj izprati tazi porcelanova kukla na decata.)

<103> Na decata **(im) ja** izprati kuklata / *Ivan*.

link-focus-tail

(Ivan dali se e setil da izprati na decata njakoja ot ljubimite im igrački?)

<104> Na decata **(im)** izprati / *KUKLATA* / Ivan.

(Ivan se kaneše po vreme na sreštata da dade kniga na sestra si i kukla na decata.)

<105> Na decata / ***im ja*** *IZPRATI* / kuklata Ivan.

focus-tail

(Setix se za porcelanovata kukla, kojato kupi Ivan. Na kogo li ja e izprati?)

<106> *NA DECATA* / **ja** izprati kuklata Ivan.

4.3.2.4 Clitic Replication in the S-V-O1-O2 Sentence Type

Having the relevant data, let us now specify the conditions on which a particular type of clitic replication is possible or even obligatory in double-object constructions.

A. Exclusive replication of the direct object NP

In the discussed S-V-O1-O2 sentence type, an exclusive direct object replication presupposes focal interpretation of the indirect object NP in order to block the realisation of its replication potential.

A.1. Obligatory replication

A.1.1. Preverbal direct object NP

Obligatory anaphoric replication of the direct object is only observed with mainly link interpretation of the replicated nominal material, whereby the indirect object NP usually constitutes the entire focus.

The function of the accusative clitic in (74c,d,e,g,h) is both object-identifying and thematicising, since the mutual positioning of the subject and the direct object NPs does not follow the obliqueness ordering, while in (74a,b,f) it is just thematicising.

(74)

surface realisation	communicative interpretation of RNM	type of alignment
a. Ivan kuklata **ja** izprati / *na decata*. <25> b. Ivan kuklata / ***ja*** *izprati na decata*.<26> c. Kuklata Ivan **ja** izprati / *na decata*. <53> d. Kuklata Ivan / ***ja*** *izprati na decata*. <54> e. Kuklata **ja** izprati Ivan / *na decata*. <58>	link	CUWO
f. Ivan kuklata / *NA DECATA* / **ja** izprati. **<30>** g. Kuklata / *NA DECATA* / **ja** izprati Ivan. **<72>** h. Kuklata **ja** izprati / *NA DECATA* / Ivan. **<76>**	link	EWO

A.1.2. Postverbal direct object NP

Obligatory cataphoric replication of the direct object only can be observed when the indirect object is in the focus and the replicated nominal material is interpreted either as a link—(75a,f)—or, mainly, as a tail—(75b-e,g). The accusative clitic in (75a,c,d) has a purely thematicising function, which in (75b,e-g) is also combined with object identification.

(75)

surface realisation	communicative interpretation of RNM	type of alignment
a. Izprati **ja** Ivan kuklata / *na decata*. **<31>**	link	CUWO
b. *NA DECATA* / **ja** izprati kuklata Ivan. **<106>** c. *Izprati **ja** na decata* / Ivan kuklata. **<81>**	tail	CMWO
d. Izprati **ja** / *NA DECATA* / Ivan kuklata. **<80>** e. Izprati **ja** / *NA DECATA* / kuklata Ivan. **<95>** f. Izprati **ja** kuklata / *NA DECATA* / Ivan. **<66>**	tail link	EWO
g. Izprati **ja** / kuklata Ivan / *na decata*. **<52>**	tail	PWO

A.2. Optional replication

A.2.1. Preverbal direct object NP

Also the ***optional anaphoric replication*** of the direct object only is mainly observed with link interpretation of the replicated material combined with a focal status for the indirect object NP. As (76) suggests, the accusative clitic exhibits both an object-identifying and thematicising function. The optionality of this replication seems to result from the fact that the subject being preverbal is positioned closer to the verb (complex) than the direct object.

(76)

surface realisation	communicative interpretation of RNM	type of alignment
a. Kuklata / *Ivan* ***(ja)*** *izprati na decata.* **<55>**	link	CUWO
b. Kuklata Ivan / *NA DECATA* / **(ja)** izprati. **<63>**	link	EWO

A.2.2. Postverbal direct object NP

The thematicising function of the ***optional cataphoric replication*** here appears to be dominant. The optionality of the cataphoric replication of the direct object which is interpreted communicatively either as a link or as a tail results in this case from the fact that the thematicisation is relatively redundant here. This is usually due to the unambiguous identification, by means of local emphasis, of the entire focal segment represented by the indirect object NP. This then makes the interpretation of the rest of the utterance as thematic (ground) rather unproblematic—cf. (77).

(77)

surface realisation	communicative interpretation of RNM	type of alignment
a. Ivan **(ja)** izprati kuklata / *na decata.* **<21>**	link	CUWO
b. *NA DECATA* / **(ja)** izprati Ivan kuklata. **<90>**	tail	CMWO
c. Ivan **(ja)** izprati / *NA DECATA* / kuklata. **<35>**	tail	EWO
d. Ivan / *NA DECATA* / **(ja)** izprati kuklata. **<41>**		
e. Izprati **(ja)** Ivan / *NA DECATA* / kuklata. **<46>**	link	

B. Exclusive replication of the indirect object NP

Exclusive replication of the indirect object in the S-V-O1-O2 sentence type presupposes that the direct object NP, on the one hand, is unambiguously

identifiable with respect to its syntactic function, and on the other hand, is interpreted as belonging to the focus.

B.1. Obligatory replication

As discussed in Section 4.3.1, the object-identifying function of clitic replication in Bulgarian has evolved from the need to distinguish the syntactic accusative from the syntactic nominative, i.e. the direct object from the subject, in ambiguous cases. This, however, is not the case with the syntactic dative, i.e. the indirect object, since the full-fledged realisation of indirect objects in Bulgarian is always unambiguously marked by the preposition *na*. Therefore, dative replication can be motivated only communicatively.[102]

B.1.1. Preverbal indirect object NP

Obligatory anaphoric replication of the indirect object would presuppose that the direct object is to be interpreted as rhematic (i.e. belonging to the focus), that it is unambiguously syntactically identifiable, and—last but not least—that the indirect object NP needs special indication of its thematic status.

The latter, however, can be the case only if the indirect-object NP is pronominal, since—as discussed in Section 4.3.1—the thematicising function of the replication (and, respectively, the rhematicising function of the lack of replication) tends then to be grammaticalised. So, no such situation is attested with the discussed sample sentence.

B.1.2. Postverbal indirect object NP

The ***obligatory cataphoric replication*** only of the indirect object basically assures that the indirect object NP, which due to its postverbal positioning would have a preferably rhematic interpretation, must be interpreted as tail—cf. (78).

---

[102] *Mayer 1988* observes in a footnote (p.112) that the dative enclitic personal pronoun survives and is sometimes used to distinguish "genitive" (i.e. possessive) constructions from indirect objects, e.g.,

Az davam knigata na učitelja. 'I give (someone) the teacher's book.'

but: *Az (mu) davam knigata na učitelja.* 'I give the book to the teacher.'

This, however, is not really supported by the data, since both examples are equally ambiguous. It is the dative reading that is relevant for the current analysis, on the basic assumption that the clitic is used for indirect-object thematicisation.

(78)

surface realisation	communicative interpretation of RNM	type of alignment
a. Izprati **im** Ivan / *KUKLATA* / na decata. **<32>**	tail	EWO
b. Izprati **im** / Ivan na decata / *kuklata*. **<48>**	tail	PWO
c. Izprati **im** / na decata Ivan / *kuklata*. **<82>**		

B.2. Optional replication

B.2.1. Preverbal indirect object NP

Optional anaphoric replication only of the indirect object generally takes place in case of preverbal positioning of the indirect object NP that has to be interpreted mainly as link.

The examples (79a-b) actually exhibit the different interpretations of a single utterance with an ambiguous focus. The same holds also for (79c-e).

(79)

surface realisation	communicative interpretation of RNM	type of alignment
a. Ivan na decata / ***(im)*** *izprati kuklata*. **<39>**	link	CUWO
b. Ivan na decata **(im)** izprati / *kuklata*. **<40>**		
c. Na decata / *Ivan* ***(im)*** *izprati kuklata*. **<83>**		
d. Na decata Ivan / ***(im)*** *izprati kuklata*. **<84>**		
e. Na decata Ivan **(im)** izprati / *kuklata*. **<85>**		
f. Na decata **(im)** izprati Ivan / *kuklata*. **<88>**		
g. Ivan na decata / *KUKLATA* / **(im)** izprati. **<44>**	link	EWO
h. Na decata Ivan / *KUKLATA* / **(im)** izprati.**<92>**		
i. Na decata / *KUKLATA* / **(im)** izprati Ivan. **<102>**		
j. Na decata **(im)** izprati / *KUKLATA* / Ivan. **<104>**		

B.2.2. Postverbal indirect object NP

Optional cataphoric replication of the indirect object only presupposes either link (80a-c,h) or tail (80d-g) interpretation of the replicated nominal material.

(80)

surface realisation	communicative interpretation of RNM	type of alignment
a. Ivan **(im)** izprati na decata / *kuklata*. **<34>**	link	CUWO
b. Izprati **(im)** Ivan na decata / *kuklata*. **<45>**		
c. Izprati **(im)** na decata Ivan / *kuklata*. **<78>**		
d. Ivan **(im)** izprati / *KUKLATA* / na decata. **<22>**	tail	EWO
e. Ivan / *KUKLATA* / **(im)** izprati na decata. **<27>**		
f. Izprati **(im)** / *KUKLATA* / Ivan na decata. **<50>**		
g. Izprati **(im)** / *KUKLATA* / na decata Ivan.**<65>**		
h. Izprati **(im)** na decata / *KUKLATA* / Ivan.**<94>**	link	

C. Replication of both the direct object NP and the indirect object NP

This is the situation where both the direct and the indirect object NPs are interpreted as thematic, i.e. as belonging to the ground.

C.1. Obligatory replication

In the majority of cases of obligatory replication, it is the verb (complex) that exclusively constitutes the focal segment, although communicative interpretations with focus amounting to or including the subject are also an available option here. Typically, any object is more communicatively loaded and, by default, more prominent that the verb itself. Therefore, in order to support the more prominent status of the verb, an explicit indication of the thematic status of its objects is required, which results in obligatory replication of the respective nominal material.

C.1.1. Uniform replication with respect to obligatoriness/optionality

C.1.1.1. Preverbal object NPs

Obligatory anaphoric replication of both the direct and the indirect objects is observed in surface alignments where the verb (complex) linearly follows them, irrespective of their mutual ordering: all possible permutations of the preverbal constituents (i.e. the two objects in (81g-h), or the two objects and the subject in (81a-f)) can be observed. In (81a-c) the replication has only a thematicising function, while in (81d-h) direct-object identification by the accusative clitic also takes place.

(81)

surface realisation	communicative interpretation of RNM	type of alignment
a. Ivan kuklata na decata / ***im ja*** *IZPRATI*. **<29>** b. Ivan na decata kuklata / ***im ja*** *IZPRATI*. **<43>** c. Na decata Ivan kuklata / ***im ja*** *IZPRATI*. **<91>** d. Kuklata Ivan na decata / ***im ja*** *IZPRATI*. **<62>** e. Kuklata na decata / *Ivan* ***im ja*** *izprati*.. **<69>** f. Na decata kuklata / *Ivan* ***im ja*** *izprati*.. **<98>**	link	CUWO
g. Kuklata na decata / ***im ja*** *IZPRATI* / Ivan. **<73>** h. Na decata kuklata / ***im ja*** *IZPRATI* / Ivan. **<101>**	link	EWO

C.1.1.2. Postverbal object NPs

Obligatory cataphoric replication of both the direct and the indirect objects, on the other hand, takes place only with surface alignments where the verb (complex) linearly precedes them, also irrespective of their mutual order, e.g., all possible postverbal permutations can be observed in (82b-g). The function of the replication in (82a-b,f) is purely thematicising, while in (82c-e,g-h) object-identification by the accusative clitic can additionally be assumed.

(82)

surface realisation	communicative interpretation of RNM	type of alignment
a. *Ivan* ***im ja*** *izprati* / na decata kuklata. **<38>** b. *IZPRATI* ***im ja*** / Ivan kuklata na decata. **<33>** c. *IZPRATI* ***im ja*** / kuklata na decata Ivan. **<67>** d. *IZPRATI* ***im ja*** / na decata kuklata Ivan.**<96>** e. *IZPRATI* ***im ja*** / kuklata Ivan na decata. **<51>** f. *IZPRATI* ***im ja*** / Ivan na decata kuklata. **<47>**	tail	CMWO
g. Izprati **im ja** / kuklata na decata / *Ivan*. **<68>** h. Izprati **im ja** / na decata kuklata / *Ivan*. **<97>**	tail	PWO

C.1.2. Different types of replication with respect to obligatoriness / optionality

C.1.2.1. Preverbal direct object NP and postverbal indirect object NP

What is observable in cases of ***obligatory anaphoric replication*** of the direct object and simultaneous ***obligatory cataphoric replication*** of the indirect object is, on the one hand, the object-identifying and thematicising function of the accusative clitic with respect to nominal material interpreted as link and, on the other hand, the thematicising function of the dative clitic with respect to nominal material interpreted as tail—cf. (83).

(83)

surface realisation	communicative interpretation of RNM	type of alignment
a. Kuklata Ivan / ***im ja*** *IZPRATI* / na decata. **<57>** b. Kuklata / ***im ja*** *IZPRATI* / Ivan na decata. **<59>** c. Kuklata / ***im ja*** *IZPRATI* / na decata Ivan. **<75>**	link & tail	EWO

C.1.2.2. Preverbal indirect object NP and postverbal direct object NP

There are two cases of ***obligatory cataphoric replication*** of the direct object and simultaneous ***obligatory anaphoric replication*** of the indirect object. The first one is illustrated in (84a-b) where both clitics have a thematicising function: the accusative one with respect to nominal material interpreted as tail and the dative one with respect to nominal material interpreted as link. The second case is given in (84c): the dative clitic imposes link interpretation on the nominal material it replicates, while the accusative one imposes tail interpretation on the replicated nominal material, and in addition identifies it as the direct object in this sentence.

(84)

surface realisation	communicative interpretation of RNM	type of alignment
a. Na decata Ivan / ***im ja*** *IZPRATI* / kuklata. **<87>** b. Na decata / ***im ja*** *IZPRATI* / Ivan kuklata. **<89>** c. Na decata / ***im ja*** *IZPRATI* / kuklata Ivan. **<105>**	link & tail	EWO

C.2. Optional replication

The optionality of replication is basically due to the fact that although the thematic (ground) status of the respective object NP is identifiable, it is not

unambiguously clear. The relevant utterances have an (emphatic) narrow focus either on the verb (complex) or on the subject NP.

C.2.1. Uniform replication with respect to obligatoriness/optionality

C.2.1.1. Preverbal object NPs

Optional anaphoric replication of both the direct object and the indirect object is not attested with respect to the sample sentence.

C.2.1.2. Postverbal object NPs

The ***optional cataphoric replication*** of both the direct object and the indirect object in (85c-f), as well as of the indirect object in (85a-b), imposes tail interpretation on the replicated nominal material. The optional accusative clitic in (85a-b), however, is a pure syntactic-category indicator because the nominal material it replicates has focus interpretation.

(85)

surface realisation	communicative interpretation of RNM	type of alignment
a. *KUKLATA* / **(im) (ja)** izprati na decata Ivan.**<77>**	focus & tail	CMWO
b. *KUKLATA* / **(im) (ja)** izprati Ivan na decata. **<61>**		
c. *IVAN* / **(im) (ja)** izprati kuklata na decata. **<24>**	tail	
d. *IVAN* / **(im) (ja)** izprati na decata kuklata. **<37>**		
e. Ivan / ***(im) (ja)*** *IZPRATI* / kuklata na decata.**<23>**	tail	EWO
f. Ivan / ***(im) (ja)*** *IZPRATI* / na decata kuklata.**<36>**		

C.2.2. Different types of replication with respect to obligatoriness / optionality

C.2.2.1. Preverbal direct object NP and postverbal indirect object NP

No case of ***optional anaphoric replication*** of the direct object NP combined with ***optional cataphoric replication*** of the indirect object NP has been attested with respect to the sample sentence.

C.2.2.2. Preverbal indirect object NP and postverbal direct object NP

The ***optional anaphoric replication*** of the direct object combined with ***optional cataphoric replication*** of the indirect object—(86)—has a clearly thematicising function: the dative clitic imposes a link interpretation and

the accusative clitic—a tail interpretation on the nominal material they replicate.

(86)

surface realisation	communicative interpretation of RNM	type of alignment
Na decata / *IVAN* / **(im) (ja)** izprati kuklata. **<86>**	link & tail	EWO

C.3. Obligatory replication of the direct object NP and optional replication of the indirect object NP

The obligatoriness of replication of the direct object NP in this case is mainly due to the object-identifying function of the accusative clitic.

C.3.1. Uniform replication with respect to obligatoriness/optionality

C.3.1.1. Preverbal object NPs

The ***obligatory anaphoric replication*** of the direct and the simultaneous ***optional anaphoric replication*** of the indirect object both impose link interpretation on the respective nominal material in (87), whereby the accusative replication also has an object-identifying function.

(87)

surface realisation	communicative interpretation of RNM	type of alignment
a. Kuklata na decata **(im) ja** izprati / *Ivan*. **<71>** b. Na decata kuklata **(im) ja** izprati / *Ivan*. **<100>**	link	CUWO
c. Kuklata na decata / *IVAN* / **(im) ja** izprati. **<70>** d. Na decata kuklata / *IVAN* / **(im) ja** izprati. **<99>**	link	EWO

C.3.1.2. Postverbal object NPs

The ***obligatory cataphoric replication*** of the direct object and the ***optional cataphoric replication*** of the indirect object have a purely thematicising function only in (88d) In the rest of the examples—(88a-c)—the thematicisation interacts with direct-object identification via an accusative clitic.

(88)

surface realisation	communicative interpretation of RNM	type of alignment
a. Izprati **(im) ja** kuklata na decata / *Ivan*. **<64>** b. Izprati **(im) ja** na decata kuklata / *Ivan*. **<93>**	link	CUWO
c. Izprati **(im) ja** kuklata / *IVAN* / na decata. **<49>** d. Izprati **(im) ja** na decata / *IVAN* / kuklata. **<79>**	link & tail	EWO

C.3.2. Different types of replication with respect to obligatoriness / optionality

C.3.2.1. Preverbal direct object NP and postverbal indirect object NP

The ***obligatory anaphoric replication*** of the direct object and the ***optional cataphoric replication*** of the indirect object are only thematicising in (89d). The accusative clitic also has object-identifying function in (89a-c).

(89)

surface realisation	communicative interpretation of RNM	type of alignment
a. Kuklata **(im) ja** izprati na decata / *Ivan*. **<74>**	link	CUWO
b. Kuklata / *IVAN* / **(im) ja** izprati na decata. **<56>** c. Kuklata **(im) ja** izprati / *IVAN* / na decata. **<60>** d. Ivan kuklata / ***(im) ja*** *IZPRATI* / na decata. **<28>**	link & tail	EWO

C.3.2.2. Preverbal indirect object NP and postverbal direct object NP

There is only one example of such communicative segmentation (90) where the ***obligatory cataphoric replication*** of the direct object has an object-identifying and thematicising function and the simultaneous ***optional anaphoric replication*** of the indirect object just a thematicising one.

(90)

surface realisation	communicative interpretation of RNM	type of alignment
Na decata **(im) ja** izprati kuklata / *Ivan*. **<103>**	link	CUWO

C.4. Obligatory replication of the indirect object NP and optional replication of the direct object NP

This is a rather improbable combination due to the fact that the dative clitic has no object-identifying function to motivate its being obligatory

while, at the same time, the accusative replication is optional. So, there are no examples exhibiting the same type of replication under these conditions, as well as no examples exhibiting different types of replication provided the direct object NP is preverbal and the indirect object NP postverbal. Let us, however, indicate these facts, following the organisation of the presentation adopted in this section.

C.4.1. Uniform replication

C.4.1.1. Preverbal object NPs

(not attested)

C.4.1.2. Postverbal object NPs

(not attested)

C.4.2. Different types of replication with respect to obligatoriness / optionality

C.4.2.1. Preverbal direct object NP and postverbal indirect object NP

(not attested)

C.4.2.2. Preverbal indirect object NP and postverbal direct object NP

The ***obligatory anaphoric replication*** of the indirect object in (91) is motivated by a relatively strong need to indicate that the replicated nominal material is to be interpreted as link, while there is only ***optional cataphoric replication*** of the direct object, inasmuch as no object identification is needed.

(91)

surface realisation	communicative interpretation of RNM	type of alignment
Ivan na decata / ***im (ja)*** *IZPRATI* / kuklata. **<42>**	link & tail	EWO

Summing up the observations about the relation between the particular type of surface alignment and the contingent clitic replication in the S-V-O1-O2 sentence type, the following generalisations can be made:

- If the word order is of the type CUWO, the (optionally or obligatorily) replicated material is always link.

- If the word order is of the type CMWO, the (optionally or obligatorily) replicated material is always tail.
- If the word order is of the type PWO, there is obligatory replication, and the replicated material is always tail; moreover, the heading verb (complex) is to be interpreted as link in such alignments.
- EWO allows for most variety: the (optionally or obligatorily) replicated material may be link, tail, or, in double-clitic cases, one of the replicated objects link and the other one tail.

4.3.2.5 When Replication is Impossible

Let us now turn our attention to those cases where it can be predicted with the highest probability that no clitic replication would ever take place. All-focus sentences illustrated in (92) do not allow for any clitic replication. Here the surface order follows the obliqueness hierarchy of grammatical relations with the verb positioned after the subject. This basically excludes the need of object identification via clitic replication, since no ambiguity with respect to the syntactic function of the involved nominal categories occurs. The lack of a thematic (ground) part in such cases also makes the thematicising function of clitic replication irrelevant.

(92)

a. *Andrej vidja kartinata.* **<1>**

b. *Ivan izprati kuklata na decata.* **<19>**

Furthermore, as the examples in (93-95) show, no replication can be found in whatever type of communicative segmentation, provided the respective object-NPs are interpreted as belonging to the focus.

(93) CUWO

a. Andrej / *vidja kartinata.* **<2>**

b. (Andrej) vidja / *kartinata.* **<3>**

c. Ivan / *izprati kuklata na decata.* **<20>**

(94) EWO

a. Andrej / *KARTINATA* / vidja. **<7>**

b. Vidja / *KARTINATA* / Andrej. **<17>**

(95) PWO

Vidja / Andrej / *kartinata.* **<9>**

The conclusion to be drawn is that, as a rule, any (even potentially replicable) nominal material that is to be interpreted as belonging to the

focus of an utterance need not and, actually, cannot be replicated. Only if object identification turns out to be problematic as a result of a particular surface alignment, an optional replication of a focal direct object is eligible, which is illustrated by the examples with reference numbers **<12>**, **<61>** and **<77>**.

4.3.2.6 On the Breadth of Focus

Finally, there is another important generalisation to be made on the basis of the regularities observed so far. In any communicative organisation with broad focus interpretation, the verb (complex) has always to be included in the focal segment[103]. This is illustrated in (96) for the S-V-O sentence type, and in (97) for the S-V-O1-O2 sentence type. No local emphasis (realised as emphatic stress) can occur if the focus is interpretable as broad (or in other terms, when it can be "projected"), since that would result in an exclusive narrow-focus interpretation (cf. also Section 4.2).

(96)

a. *Andrej vidja kartinata.* **<1>**

b. Andrej / *vidja kartinata.* **<2>**

c. Kartinata / *Andrej* ***ja*** *vidja.* **<13>**

(97)

a. *Ivan izprati kuklata na decata.* **<19>**

b. Ivan / *izprati kuklata na decata.* **<20>**

c. Ivan kuklata / ***ja*** *izprati na decata.* **<26>**

d. *Ivan* ***im ja*** *izprati* / na decata kuklata. **<38>**

e. Ivan na decata / ***(im)*** *izprati kuklata.* **<39>**

f. Kuklata Ivan / ***ja*** *izprati na decata.* **<54>**

g. Kuklata / *Ivan* ***(ja)*** *izprati na decata.* **<55>**

h. Kuklata na decata / *Ivan* ***im ja*** *izprati.* **<69>**

i. *Izprati* ***ja*** *na decata* / Ivan kuklata. **<81>**

j. Na decata / *Ivan* ***(im)*** *izprati kuklata.* **<83>**

k. Na decata Ivan / ***(im)*** *izprati kuklata.* **<84>**

l. Na decata kuklata / *Ivan* ***im ja*** *izprati.* **<98>**

---

[103]A similar observation concerning the so-called focus projection in German is made in *Höhle 1982*.

The relative order within the focal segment obviously corresponds to the canonical one, i.e. to the lexeme-specific obliqueness ordering, whereby the verb (complex) has to follow the subject NP and to precede any object NP in declarative utterances.

In general, two cases of non-emphatic narrow focus can be distinguished. The first is the narrow focus interpretation in ambiguous-focus utterances, exemplified in (98). The utterances in (98a-b) can also be interpreted either as entirely rhematic (cf. (96a) and (97a)) or as having a thematic subject only (cf. (96b) and (97b)); the focal segment in (98c-d) can also cover the verb (complex) (cf. (97f,k)) and even the immediately preverbal subject (cf. (97g,j)); as to (98e-f), the verb (complex) may also be included in the focus (cf. (97c,e)).

(98)

a. Andrej vidja / *kartinata*. **<3>**

b. Ivan **(ja)** izprati kuklata / *na decata*. **<21>**

c. Kuklata Ivan **ja** izprati / *na decata*. **<53>**

d. Na decata Ivan **(im)** izprati / *kuklata*. **<85>**

e. Ivan kuklata **ja** izprati / *na decata*. **<25>**

f. Ivan na decata **(im)** izprati / *kuklata*. **<40>**

In the second case, however, there are no "broader" alternatives to the coverage of the focal segment, and thus only a narrow-focus reading is available in utterances like those given in (99-101). Since there is no immediate adjacency between the verb (complex) and the focal segment in (99), (100b-e) and (101), no "extension" or "projection" of the focus is actually possible.

(99)

a. Vidja / Andrej / *kartinata*. **<9>**

b. Izprati **ja** / kuklata Ivan / *na decata*. **<52>**

c. Izprati **im** / Ivan na decata / *kuklata*. **<48>**

d. Izprati **im** / na decata Ivan / *kuklata*. **<82>**

e. Izprati **im ja** / kuklata na decata / *Ivan*. **<68>**

f. Izprati **im ja** / na decata kuklata / *Ivan*. **<97>**

The sentence-final placement of the subject in (100) makes an additional contribution to an exclusively narrow-focus interpretation because this obviously diverges from the obliqueness ordering.

(100)

a. Kartinata **ja** vidja / *Andrej*. **<10>**

b. Izprati **(im) ja** kuklata na decata / *Ivan*. **<64>**

c. Izprati **(im) ja** na decata kuklata / *Ivan*. **<93>**

d. Kuklata **(im) ja** izprati na decata / *Ivan*. **<74>**

e. Na decata **(im) ja** izprati kuklata / *Ivan*. **<103>**

f. Kuklata na decata **(im) ja** izprati / *Ivan*. **<71>**

g. Na decata kuklata **(im) ja** izprati / *Ivan*. **<100>**

While the relative order of the subject and the two objects in (101a) reflects the obliqueness hierarchy, it is the obligatory accusative clitic that imposes ground interpretation on the direct object, and thus blocks an all-focus interpretation of this utterance. In (101b-f) the surface alignment again does not coincide with the obliqueness ordering, which is an additional factor in blocking any focus projection.

(101)

a. Izprati **ja** Ivan kuklata / *na decata.* **<31>**

b. Kuklata **ja** izprati Ivan / *na decata.* **<58>**

c. Ivan **(im)** izprati na decata / *kuklata.* **<34>**

d. Izprati **(im)** Ivan na decata / *kuklata.* **<45>**

e. Izprati **(im)** na decata Ivan / *kuklata.* **<78>**

f. Na decata **(im)** izprati Ivan / *kuklata.* **<88>**

In other words, a narrow-focus interpretation of a given element implies no emphatic stress if this element is final and the overall divergence of the surface alignment from the canonical obliqueness-determined relative ordering is clearly recognisable. Otherwise, an emphatic stress is implied if the respective element is to be interpreted as the focus proper.

4.3.3 Aspects of Formalisation

The adopted type of syntactic structures—cf. Section 2.2 for details—allows for free permutation of all syntactic constituents on the clausal level. What is needed, however, is to specify linguistically motivated linear precedence constraints that would allow the unacceptability of particular surface constituent orderings in the Bulgarian clause to be modelled on the basis of the impossibility of combining certain properties of these constituents. The syntactic verbal constituent—i.e. the clause—is an

autonomous word-order domain. Its nucleus head can be either a lexical verb or a morphosyntactic verb complex. Let us recall that the empty nonterminal occurring as the rightmost element of the adopted binary branching structure of the clause is a functional syntactic projection of the respective nucleus head, and as such inherits distinguished features from it, but also introduces new ones appropriate only for a verbal category as a syntactic constituent, i.e. for the clausal level—cf. Figure 30.

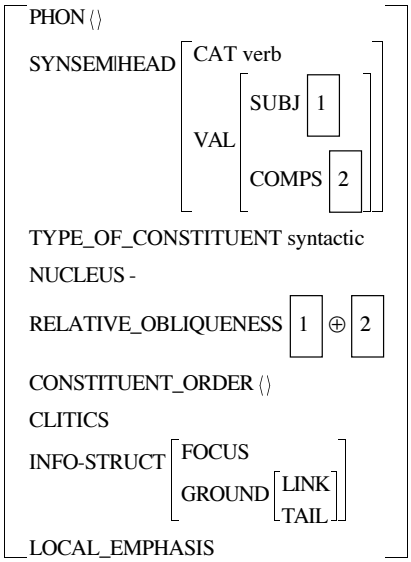


Figure 30

The nucleus head is not yet realised, therefore the feature NUCLEUS has a negative value. The CONSTITUENT_ORDER list is set to be empty at the phrase-terminating node of the structure; it is supposed to collect **all** major sentential constituents (the nucleus head with its obligatory and free modifications) in the order they occur in the utterance, so that, e.g., a check whether the relative order of the subcategorised elements diverges from or complies with the obliqueness hierarchy is possible. Both the RELATIVE_OBLIQUENESS list and the CLITICS list are constructed in the terminal node of the structure (i.e. at the functional syntactic projection of the nucleus head). The former basically concatenates the values of the

word-form specific valence features SUBJ and COMPS[104]. The latter is a list of objects of the type *replicant* given in Figure 31.

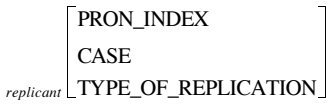


Figure 31

The information on the features PRON_INDEX and CASE of a verbal clitic is encoded in the valence information of the morphosyntactic nucleus and—in accordance with the Head Feature Principle—is thus shared also with the functional syntactic head. The third feature TYPE_OF_REPLICATION is unspecified at the functional head of the clause and can be instantiated with the values "anaphoric" or "cataphoric" as soon as the corresponding full-fledged NP is expanded in the structure, and with the value "none" if it is only the clitic pronoun that satisfies the respective valence requirement and no replication takes place.[105]

The feature INFO_STRUCT in the present shape is basically taken over from *Vallduví and Engdahl 1995*; it is unspecified at the phrase-termination node of the structure. The value of the feature LOCAL_EMPHASIS is unspecified at the functional head of the clause; it can be either instantiated as "none", presupposing a normal-stress pattern, or set to be token-identical to the element contained in the focus list, inasmuch as emphatic stress always presupposes narrow-focus interpretation. In the general architecture of the sign, this feature will correlate with certain prosodic properties and/or corresponding changes in the intonational contour, which have to be incorporated in some way in the PHON (phonology) value as appropriate features, making it thus more complex than just a mere string of phonemes. To explicitly specify what the shape of the PHON feature could be, further research on the intonational phonology of the Bulgarian sentence is required, which, however, would be beyond the scope of this thesis.

---

[104]This is necessary because the lexeme-specific SUBCAT feature cannot always be used directly. It represents the obliqueness order of grammatical relations appropriate for the particular verbal lexeme which is only canonical with active-voice morphology. In my approach, it is the valence—not the lexical subcategorisation—that changes with the passive morphology. Obviously, passive verb-forms presuppose different obliqueness hierarchy of grammatical relations.

[105]In parsing, the latter situation can be unambiguously detected only after the processing of the whole sentence has been completed.

Let us now illustrate on concrete examples—e.g., (102)—how the analysis would work.

(102)

a. Kartinata **ja** vidja / *Andrej*. **<10>**

b. Kartinata / ***ja*** *VIDJA* / Andrej. **<11>**

c. *KARTINATA* / (**ja**) vidja Andrej. **<12>**

All three examples have the same wording, and differ mainly with respect to the information structure. The syntactic analysis is sketched in Figure 32 by indicating the relevant features.

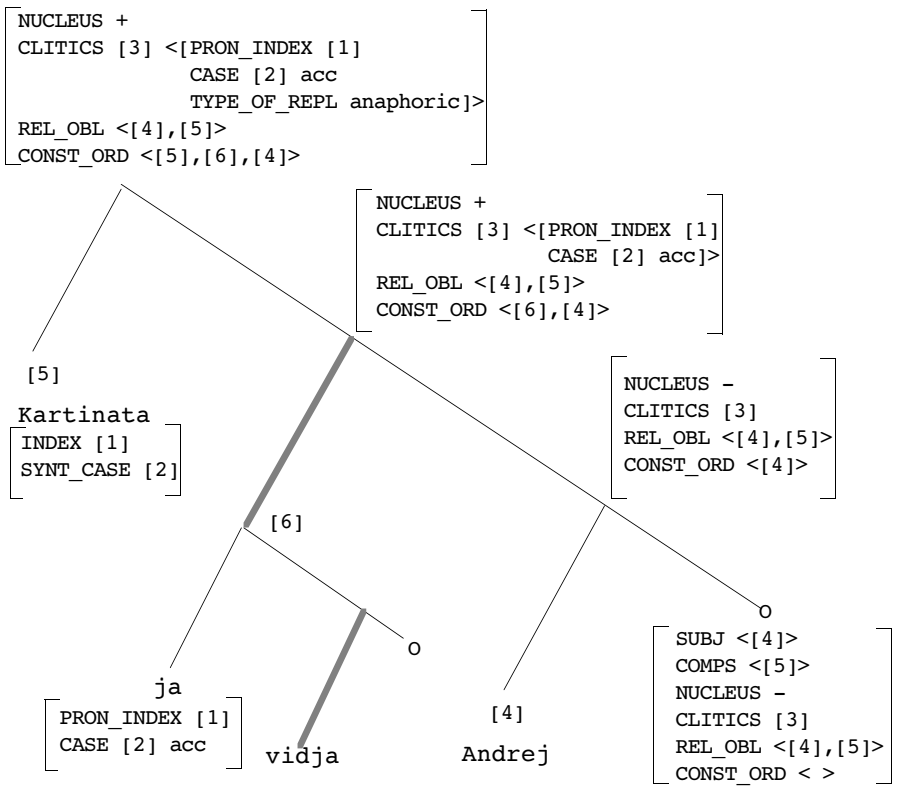


Figure 32

As to the communicative segmentation, there will be a disjunction of possible interpretations—cf. Figure 33, each corresponding to one of the utterances in (102).

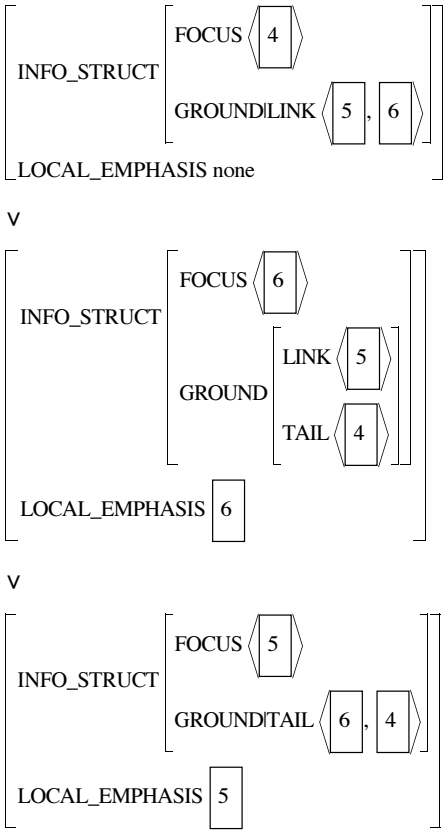


Figure 33

5 General Summary and Conclusions

Word order phenomena involving clitics have been in the focus of this thesis. I have argued that the cliticisation in Bulgarian has a morphosyntactic dimension and that verbal clitics belong to the verb-complex constituent which I regard as an intermediate construct between the lexical verb and the clause headed by this verb. Therefore, the actual placement of verbal clitics (and verbal clitic sequences) is interpreted on the level of verbal morphosyntax, whereby prosodic constraints are also taken into consideration. As a consequence, pronominal clitics—being just a part of the verb complex that heads the respective clause—are not legitimate constituents on the clausal level. An interesting fact about the adopted approach is that the problem of two-fold satisfaction of a single valence requirement—once by a pronominal clitic and once by a coreferential full-fledged NP-complement—will thus never arise. I have further argued that clitic replication of full-fledged NP-complements has a communicatively-driven syntactic dimension and deserves special attention as a factor influencing the constituent order variation in the Bulgarian sentence. In this respect, I have shown how the (canonical) lexeme-specific obliqueness order of grammatical relations, the concrete surface alignment and the contingent clitic replication interact in the communicative organisation of utterances.

A particular contribution of this thesis is the linguistic model of Bulgarian constituent structure allowing for adequate representation and interpretation of traditionally problematic data. It is possible now to make a clear distinction between morphosyntactic and syntactic phenomena in a language exhibiting a high degree of analyticity. The concept of morphosyntactic marking introduced in this study is central in the treatment of Bulgarian analytic verb forms.

In the proposed formal description of the structure and the word order properties of the Bulgarian verb complex, special attention has been paid to the placement of the verbal clitics and clitic clusters that is also subject to prosodic constraints. The clause-initial restriction and the quasi-second-position condition, as well as the placement of the question particle, are directly related to the prosodic organisation of the verbal morphosyntactic constituent into accentual units (prosodic words). Each verbal clitic is

specified as being core or peripheral on the basis of predominantly prosodic criteria. Unlike peripheral clitics, core clitics appear to be obligatory components of clitic clusters. It has been demonstrated how particular clitic-cluster placements in the verb complex can be attributed to the interaction of different clitic types in conformity to more general constraints. Based on the constituent structure and syntactic behaviour, two main types of verb complexes are distinguished—compact verb complexes characterised by strict adjacency of their components and composite verb complexes having two loosely bound parts which need not stand in immediate adjacency, so that other non-verbal sentential constituents may intervene.

With respect to constructions in which the contingent short pronominal form indicating the experiencer is an obligatory lexical formant, I have argued that it functions essentially as an agreement marker (rather than a replicant) if a full-fledged coreferential nominal constituent is present in the clause, and that this phenomenon should not be regarded as clitic replication.

As a necessary prerequisite for discussing the role of clitic replication on the clausal level, a typology of Bulgarian articled and non-articled NPs has been developed, providing criteria for determining the replication potential of nominal material. I have argued that what can be replicated by a clitic pronoun—under the appropriate verb-lexeme specific or communicative conditions—is the nominal material that is used as identifying specific description of a given object, while non-articled NPs that are categorising or non-specific descriptions, as well as articled NPs that are generic or non-specific descriptions completely lack replication potential.

The accusative clitic replication in the S-V-O sentence type, and the accusative and dative clitic replication in the S-V-O1-O2 sentence type are then modelled in their relation to the grammatical obliqueness hierarchy (the canonical element order), the particular surface alignment (fairly flexible in Bulgarian) and the information structure (the communicative segmentation of a particular utterance) with special attention to the emphatic-stress location. On the proposed approach, it can be predicted when clitic replication is impossible, when it is obligatory, and when it is only optional.

An important feature of the linguistic analysis proposed in this thesis is its computational tractability. The implemented fragment of an HPSG-style Bulgarian grammar covers in full range the verbal morphosyntax, and to a considerable extent the replication phenomena on the clausal level. The

next step, and a matter of further application-driven research, would be to test the approach on real-text corpora, which would allow for avoiding the shortcomings of using mainly theoretical sources and introspective data in building up the language description.

During the work on this dissertation, some proposals for a lexicalist treatment of Romance clitics in the HPSG framework have appeared. Therein, it is argued in favour of the affixal status of Italian (*Monachesi 1995*) and French (*Miller and Sag 1995*) clitics, and cliticisation is thus reduced to affixation. Such an approach to cliticisation employs a set of interacting lexical rules[106]. While the (multiple) inheritance eliminates the so called "vertical" redundancy in the hierarchical organisation of the lexicon, lexical rules in HPSG are considered one of the means of eliminating the so called "horizontal" redundancy. According to *Flickinger 1987*, given a word belonging to a certain type, a lexical rule predicts the existence of a corresponding word belonging to another type, with the differences and similarities between the two words captured both in the formulation of the rule and in the definition of the types of each word. Even though extensively used in some HPSG language descriptions, the lexical rules have a rather unclear formal status and are problematic for computational implementations, which has lead to their (currently dominant) abandonment in favour of various alternatives that rely mainly on the notion of type hierarchies.

My analysis is based on a variant of HPSG which provides an intermediate representation level for the analytic verb morphology and cliticisation. It is strongly motivated by both the complexity of verbal forms and the syntax of predicative clitics in Bulgarian. Once the step towards introducing morphosyntactic constituency is made, the language description gains in explanatory power and transparency with respect to a number of phenomena admittedly belonging to the vague "interface" area between the lexicon and syntax proper. As a result, a clearly defined morphosyntactic module in the grammar of Bulgarian distinguishes it, e.g., from the rest of the Slavic languages. This, in turn, shows that within HPSG, the parametrisation of linguistic variation can occur not only in the lexicon but also in the grammar. In this respect , it might be relevant to investigate whether a morphosyntactic grammar module would be justified

[106]For example, the analysis of Monachesi involves Passive Lexical Rule, Argument Composition Lexical Rule, Complement Cliticisation Lexical Rule, Impersonal si Lexical Rule, Lexical Rule for middle verbs, Lexical Rule for ergative verbs, Lexical Rule for object clitic, etc.

and beneficial in the description of other languages exhibiting phenomena that are problematic for the lexicalist HPSG approach.

References

P. Adamec. 1966. *Porjadok slov v sovremennom russkom jazyke.* Rozpravy Československé akademie věd. Řada společenských věd 76. Prague: Academia

S.R. Anderson. 1992. *A-Morphous Morphology.* Cambridge: CUP

S.R. Anderson. 1993. *Wackernagel's Revenge: Clitics, Morphology, and the Syntax of Second Position.* In: Language 69 (1): 68-98

L. Andrejčin. 1938. *Kategorie znaczeniowe koniugacji bułgarskiej.* In: Prace Komisji Językowej 26. Kraków: Polska Akademia Umiejętności

H.I. Aronson. 1967. *On Some Functions of the Opposition Definite / Non-Definite in Bulgarian.* In: Languages and Areas: Studies Presented to George V. Bobrinskoy on the Occasion of His Academic Rretirement. Chicago: UCP: 1-5

P. Asenova. 1989. *Balkansko ezikoznanie. Osnovni problemi na Balkanskija ezikov sâjuz.* Sofia: Nauka i izkustvo

T. Avgustinova. 1996. *An HPSG-style Grammar of Bulgarian (for the Purposes of a Grammar-Checker Implementation).* Final Deliverable Report, Project LATESLAV. University of Saarland

T. Avgustinova and K. Oliva. 1990. *Syntactic Description of Free Word Order Languages.* Proceedings of COLING-90, Helsinki. Vol. 3: 311-313.

T. Avgustinova and K. Oliva. 1991. *The Structure of Bulgarian Verb Complex.* CLAUS-Report No.4. University of Saarland

T. Avgustinova and K. Oliva. 1995a. *On the Nature of the Wackernagel Position in Czech.* Proceedings of FDSL-1, Leipzig (Junghanns and Zybatow (eds.) Formale Slavistik. Frankfurt am Main: Vervuert Verlag, 1997: 25-47)

T. Avgustinova and K. Oliva. 1995b. *The Position of Sentential Clitics in the Czech Clause.* CLAUS-Report No. 68. University of Saarland

T. Avgustinova and K. Oliva. 1995c. *Wackernagel Position and Related Phenomena in Czech.* In: Wiener Slavistisches Jahrbuch 41: 21-42

D.L. Bolinger. 1972. *Accent is predictable (if you are a mind reader).* In: Language 48: 633-44

J. Bresnan, ed. 1982. *The Mental Representation of Grammatical Relations.* MIT Press Series on Cognitive Theory and Mental Representation. Cambridge, MA: The MIT Press

B. Carpenter. 1992. *The Logic of Typed Feature Structures*. Cambridge Tracts in Theoretical Computer Science 32. New York: CUP

W.L. Chafe. 1976. *Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View*. In: C. N. Li (ed.) Subject and Topic. New York: 25-55

N. Chomsky. 1957. *Syntactic Structures*. The Hague: Mouton

N. Chomsky. 1981a. *Lectures on Government and Binding*. Studies in Generative Grammar 9. Dordrect: Foris

N. Chomsky. 1981b. *Principles and Parameters in Syntactic Theory.* In: N. Hornstein and D. Lightfoot (eds.) Explanation in Linguistics. London: Longman: 32-75

N. Chomsky. 1986. *Knowledge of Language. Its Nature, Origin and Use*. New York: Praeger

K. Čolakova. 1973. *Za njakoi glagoli s analitično oformena struktura i tjaxnata leksikografska obrabotka.* In: Bâlgarski ezik XXXIII (6)

G. Cyxun. 1962. *Mestoimennata enklitika i slovoredât v bâlgarskoto izrečenie*. In: Bâlgarski ezik XII (4): 283-91

G. Cyxun. 1968. *Sintaksis mestoimennyx klitik v južnoslavjanskix jazykax (Balkansko-slavjanskaja model').* Minsk

F. Daneš. 1959. *K otázce pořadku slov v slovanských jazycích.* In: Slovo a Slovesnost 20: 1-9

F. Daneš. 1967. *Order of elements and sentence intonation.* In: To honour R. Jakobson. 1

D.L. Dyer. 1988. *Intensifying RP and the Effect of Grammatical Categories on Replicated NM in Bulgarian.* In: University of Chicago Working Papers in Linguistics 4: 1-24

D.L. Dyer. 1992. *Word Order in the Simple Bulgarian Sentence: A Study in Grammar, Semantics and Pragmatics*. Amsterdam: Rodopi

D.L. Dyer and G. Fowler. 1988. *The Syntax of Deverbal Nouns in Bulgarian*. Talk on the Conference on Balkan and South Slavic Linguistics and Folklore

E. Engdahl, ed. 1994. *Integrating Information Structure into Constraint-Based and Categorial Approaches* (DYANA-2 Report R1.3.B). Amsterdam

E. Engdahl and E. Vallduví. 1994. *Information Packaging and Grammar Architecture: A Constraint-Based Approach*. In: E. Engdahl (ed.) Integrating Information Structure into Constraint-Based and Categorial Approaches (DYANA-2 Report R1.3.B). Amsterdam: 41-79

J. Engelkamp, G. Erbach and H. Uszkoreit. 1992. *Handling Linear Precedence Constraints by Unification.* Proceedings of 30th ACL Meeting. Newark, Delaware

R. Ewen. 1975. *On describing clitic ordering in Bulgarian.* In: Working Papers in Linguistics. University of Washington: 18-38

J. Firbas. 1957. *Some Thoughts in the Function of Word Order in Old English and Modern English.* In: Sborník prací filosofické fakulty Brněnské university: 72-100

J. Firbas. 1964. *On defining the theme in functional sentence perspective.* In: Travaux linguistiques de Prague 1: 267-80

J. Firbas. 1971. *On the concept of communicative dynamism in the theory of Functional Sentence Perspective.* In: Brno Studies in English 7: 12-47

J. Firbas. 1972. *On the interplay of prosodic means of functional sentence perspective (a theoretical note on teaching of English intonation).* In: V. Fried (ed.) The Prague school of linguistics and language teaching. London

J. Firbas. 1974. Some *Aspects of the Czechoslovak Approach to Problems of Functional Sentence Perspective.* In: Papers on Functional Sentence Perspective. Prague: Academia

J. Firbas. 1975. *On the thematic and non-thematic section of the sentence*. In: H. Ringbom et al. (eds.) Style and Text. Studies presented to N. E. Enkvist. Stockholm: 314-34

D. Flickinger. 1987. *Lexical Rules in the Hierarchical Lexicon.* PhD Dissertation. Stanford University

V. Friedman. 1976. *The Question of the Bulgarian Indefinite Article.* In: T. Butler (ed.) Bulgaria. Past and Present. Columbus, Ohio: 334-40

G. Gazdar, E. Klein, G.K. Pullum and I. Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwell

E. Georgieva. 1974. *Slovored na prostoto izrečenie v bâlgarskija knižoven ezik.* Sofia: BAN

Z. Guentchéva. 1994. *Thématisation de l'objet en bulgare*. Sciences pur la communication. 39. Bern-Berlin-Frankfurt/M.-Paris-Wien: Peter Lang

T. Gunji. 1987. *Japanese Phrase Structure Grammar.* Dordrecht: Riedel

E. Hajičová. 1973. *Negation and Topic vs. Comment.* In: Philologia Pragensia 16: 81-93

E. Hajičová. 1984a. *Presupposition and Allegation Revisited.* In: Journal of Pragmatics 8: 155-167

E. Hajičová. 1984b. *Topic and Focus*. In: P. Sgall (ed.) Contributions to Functional Syntax. Semantics and Language Comprehension (LLSEE 16). Amsterdam-Philadelphia: John Benjamins: 189-202

E. Hajičová and P. Sgall. 1985. *Towards an automatic identification of topic and focus.* Proceedings of 2nd EACL

E. Hajičová and P. Sgall. 1987. *The Ordering Principle*. In: Journal of Pragmatics 11: 435-454

E. Hajičová and J. Vrbová. 1982. *On the role of the hierarchy of activation in the process of natural language understanding*. Proceedings of COLING-82, Prague: 107-13

A. Halpern. 1992. *Topics in the Placement and Morphology of Clitics.* PhD Dissertation. Stanford University

A. Halpern. 1995. *On the Placement and Morphology of Clitics.* CSLI Lecture Notes. Cambridge: CUP

K.R. Hauge. 1976. *The word order of predicate clitics in Bulgarian.* Meddelelser 10. University of Oslo

K.R. Hauge. 1995. *A Short Grammar of Contemporary Bulgarian*. University of Oslo, Department of East European and Oriental Studies

T. Höhle. 1982. *Explikationen für normale Betonung und normale Wortstellung*. In: W. Abraham (ed.) Satzglieder im Deutschen. Tübingen: Gunter Narr Verlag.

S. Ivančev. 1957/1978. *Nabljudenija vârxu upotrebata na člena v bâlgarskija ezik*. In: P. Pašov (ed.) Pomagalo po bâlgarska morfologija. Imena. Sofia: Nauka i izkustvo: 186-211

S. Ivančev. 1967/1978. *Kâm vâprosa za generično upotrebenija podlog*. In: Prinosi v bâlgarskoto i slavjanskoto ezikoznanie. Sofia: Nauka i izkustvo: 152-7

S. Ivančev. 1968/1978. *Problemi na aktualnoto členenie na izrečenieto.* In: Prinosi v bâlgarskoto i slavjanskoto ezikoznanie. Sofia: Nauka i izkustvo: 157-73

S. Ivančev. 1971. *Problemi na aspektualnostta v slavjanskite ezici.* Sofia: BAN

S. Ivančev. 1974/1978. *Semantika i funkcionirane na pokazatelnoto mestoimenie tozi / toja v sâvremennija bâlgarski ezik*. In: Prinosi v bâlgarskoto i slavjanskoto ezikoznanie. Sofia: Nauka i izkustvo: 184-96

S. Ivančev. 1977/1978. *Za semantiko-funkcionalnata kategorija količestvenost (kvantitativnost) ili za okoličestvjavaneto (kvantifikacijata) v bâlgarskija ezik*. In: Prinosi v bâlgarskoto i slavjanskoto ezikoznanie. Sofia: Nauka i izkustvo: 98-107

S. Ivančev. 1978. *Prinosi v bâlgarskoto i slavjanskoto ezikoznanie*. Sofia: Nauka i izkustvo

R. Jackendoff. 1972. *Semantic interpretation in generative grammar*. Cambridge: MIT Press

B.D. Joseph, J.A. Nevis, D. Wanner and A.M. Zwicky. 1994. *Clitics: A Comprehensive Bibliography 1982-1991*. Library and Information Sources in Linguistics 22. Amsterdam-Philadelphia: John Benjamins

I. Kânčev. 1972. *Njakoi nabljudenija vârxu upotrebata na udvoenoto dopâlnenie v ispanski i bâlgarski ezik*. In: Ezik i literatura 1

L. Karttunen. 1986. *Radical Lexicalism*. Rerport CSLI-86-68. Stanford

M. Kay. 1985. *Parsing in functional unification grammar*. In: L. K. D.Dowty and M.Zwicky (eds.) Natural Language Parsing. Cambridge: CUP

E. Koktová. to appear. *Word Order Based Grammar*. Mouton

D.R. Ladd. 1996. *Intonational Phonology*. Cambridge: CUP

J. Lindstedt. 1985. *On the Semantics of Tense and Aspect in Bulgarian*. Slavica Helsingensia 4. Helsinki

J.A. Lopašov. 1978. *Mestoimennye povtory dopolnenija v balkanskix jazykax*. Leningrad

J. Maslov. 1963. *Morfologija glagol'nogo vida v sovremennom bolgarskom literaturnom jazyke*. Moscow-Leningrad: Izdatel'stvo Akademii Nauk SSSR

V. Mathesius. 1907. *Studie k dějinam anglického slovosledu*. In: Věstník České akademie 6. Prague

V. Mathesius. 1915. *O passivu v moderní angličtině*. In: Sborník filologický 5: 198-220

V. Mathesius. 1941. *Základní funkce pořadku slov v češtině*. In: Slovo a Slovesnost 7: 169-80

V. Mathesius. 1942. *Ze srovnávacích studií slovosledných*. In: ČMF 28: 13-102

V. Mathesius. 1947a. *K pořádku slov v hovorové češtině*. In: Čeština a obecný jazykozpyt: 353-59

V. Mathesius. 1947b. *Základní funkce pořádku slov v češtině*. In: Čeština a obecný jazykozpyt: 327-52

V. Mathesius. 1975 (1939). *On the information-bearing structure of the sentence*. Harvard Studies in Syntax and Semantics 1: 467-80

G.L. Mayer. 1988. *The Definite Article in Contemporary Standard Bulgarian*. Balkanologische Veröfentlichungen 14. Berlin: Otto Harrassowitz - Wiesbaden

L. Miletič. 1937. *Udvojavaneto na obekta v bâlgarskija ezik ne e "balkanizâm"*. In: Spisanie na BAN 56 (28): 1-21

P. Miller. 1992. *Clitics and Constituents in Phrase Structure Grammar*. Outstanding Dissertations in Linguistics. New York: Garland

P. Miller and I. Sag. 1995. *French Clitic Movement Without Clitics or Movement.* Proceedings of NELS

P. Monachesi. 1995. *A Grammar of Italian Clitics*. ITK Dissertation Series. Tilburg

A. Minčeva. 1968. *Za njakoi aspekti na reprizata na ličnite mestoimenija v bâlgarskija ezik*. In: Izvestija na instituta za bâlgarski ezik XVI: 371-9

A. Minčeva. 1969. *Opit za interpretacija na modela na udvoenite dopâlnenija v bâlgarskija ezik*. In: Izvestija na instituta za bâlgarski ezik XVII: 3-50

K. Mirčev. 1966. *Za xronologijata na osnovnite balkanizmi v bâlgarskija ezik*. In: Bâlgarski ezik 4: 284-6

K. Mirčev. 1972. *Starobâlgarski ezik.* Sofia

C. Mladenov. 1968. *Balkanizâm li e udvojavaneto na obekta v bâlgarski?* In: Izvestija na instituta za bâlgarski ezik XVI: 151-6

K. Naylor. 1983. *On Expressing 'Definiteness' in the Slavic Languages and English.* In: M. S. Flier (ed.) American Contributions to the Ninth International Congress of Slavists. 1 - Linguistics. Columbus, Ohio: Slavica: 202-20

K. Netter. 1992. *On Non-Head Non-Movement*. In: G. Görz (ed.) KONVENS-92. Berlin: Springer-Verlag: 218-227

K. Netter. 1994. *Towards a Theory of Functional Heads*. In: J. Nerbonne, K. Netter and C. Pollard (eds.) German in Head-Driven Phrase Structure Grammar. Stanford: Stanford University: 297-340

R. Nicolova. 1986. *Bâlgarskite mestoimenija.* Sofia: Nauka i izkustvo

K. Oliva. 1992a. *The Proper Treatment of Word Order in HPSG*. Proceedings of COLING-92, Nantes. Vol. 1: 184-190

K. Oliva. 1992b. *Word Order Constraints in Binary Branching Syntactic Structures.* CLAUS-Report No. 20. University of Saarland

K. Oliva. 1994. *HPSG Lexicon Without Lexical Rules.* Proceedings of COLING-94, Kyoto. Vol. 2: 823-6

H. Orzechowska. 1976. *Procesy bałkanizacji i sławizacji bułgarskiego języka literackiego XVII-XIX w. w świećie użycia klitycznych form zaimków.* Warszawa

J. Panevová. 1980. *Formy a funkce ve stavbě české věty*. Prague: Academia

P. Pašov. 1974. *Za upotrebata na njakoi mestoimenni formi*. In: L. Andrejčin (ed.) Problemi na bâlgarskata knižovna reč. Sofia: Nauka i izkystvo

P. Pašov. 1989. *Praktičeska gramatika na bâlgarskija knižoven ezik*. Sofia: Nauka i izkustvo

J. Penčev. 1972. *Refleksivnite, medialnite i pasivnite izrečenija v bâlgarski ezik.* In: Izvestija na Instituta za bâlgarski ezik XXI. Sofia: BAN: 245-77

J. Penčev. 1980. *Osnovni intonacionni konturi v bâlgarskoto izrečenie.* Sofia: Nauka i izkustvo

J. Penčev. 1984. *Stroež na bâlgarskoto izrečenie.* Sofia: Nauka i izkustvo

J. Penčev. 1993. *Bâlgarski sintaksis: upravlenie i svârzvane.* Plovdiv

C. Pollard and I. Sag. 1987. *Information-Based Syntax and Semantics.* CSLI Lecture Notes 13. Vol. 1: Fundamentals. Stanford: Stanford University

C. Pollard and I. Sag. 1988. *An Information-Based Theory of Agreement*. Report CSLI-88-132. Stanford

C. Pollard and I. Sag. 1994. *Head-driven Phrase Structure Grammar.* Chicago: UCP

K. Popov. 1962. *Stilno-gramatična upotreba na udvoenoto dopâlnenie v bâlgarskija knižoven ezik.* In: Izvestija na instituta za bâlgarski ezik VIII: 459-70

C. Rudin. 1982. *The syntax and semantics of object reduplication.* Proceedings of 3rd Balkan and South Slavic Conference. Bloomington

C. Rudin. 1986. *Aspects of Bulgarian Syntax: Complementizers and WH Constructions.* Columbus, Ohio: Slavica

J. Rusek. 1963. *Po vâprosa za xronologijata na udvojavane na dopâlnenjata v bâlgarskija ezik.* In: Bâlgarski ezik XIII (2)

I. Sag, R. Kaplan, L.Karttunen, M.Kay, C.Pollard, S.Shieber and A.Zaenen. 1986. *Unification and Grammatical Theory*. Proceedings of WCCFL

I.A. Sag and C. Pollard. 1989. *Subcategorization and Head-Driven Phrase Structure.* In: M. R. Baltin and A. S. Kroch (eds.) Alternate Conceptions of Phrase Structure. Chicago: UCP

E.A. Scatton. 1984. *A Reference Grammar of Modern Bulgarian.* Columbus, Ohio: Slavica

P. Sgall, E. Hajičová and J. Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht-Boston-Lancaster-Tokyo: Reidel

T. Shamray. 1989. *Členuvani i nečlenuvani imena v bâlgarskija ezik.* Sofia: Narodna prosveta

S. Shieber. 1986. *An Introduction to Unification-Based Approaches to Grammar.* CSLI Lecture Notes 4. Stanford

S. Szober. 1933/1979. *Edin osnoven tip nareždane dumite v bâlgarskoto izrečenie.* In: K. Popov (ed.) Pomagalo po bâlgarski sintaksis. Sofia: Nauka i izkustvo: 275-85 (originally published in: Sbornik v čest na prof. L. Miletič)

F. Sławski. 1946. *Miejsce enklityki odmiennej w dzieinach języka bułgarskiego*. Kraków

D. Tilkov. 1980. *Učastieto na klitikite v izgraždaneto na akcentnata edinica v bâlgarskija knižoven ezik*. In: H. Pârvev and V. Radeva (eds.) Pomagalo po bâlgarska fonetika. Sofia: Nauka i izkustvo: 218-29

F. Trávníček. 1937a. *Slovosled při důrazu*. In: Slovo a Slovesnost 5: 131-44

F. Trávníček. 1937b. *Základy československého slovosledu.* In: Slovo a Slovesnost 3: 78

F. Trávníček. 1951. *Mluvnice spisovné češtiny.* Vol. 2. Prague

F. Trávníček. 1959. *K postavení stálých příklonek po přestávce uvnitř věty.* In: Naše řeč XLII

F. Trávníček. 1962. *O tak zvaném aktuálním členění větném.* In: Slovo a Slovesnost 22: 163-71

H. Uszkoreit. 1986a. *Categorial Unification Grammars*. Proceedings of COLING-86, Bonn

H. Uszkoreit. 1986b. *Linear Precedence in Discontinuous Constituents: Complex Fronting in German*. Report CSLI-89-47. Stanford

E. Vallduví. 1990. *The Informational Component*. PhD Dissertation. University of Pennsylvania

E. Vallduví. 1992. The *Informational Component.* New York. Garland

E. Vallduví. 1994. *The Dynamics of Information Packaging*. In: E. Engdahl (ed.) Integrating Information Structure into Constraint-Based and Categorial Approaches (DYANA-2 Report R1.3.B). Amsterdam: 1-26

E. Vallduví and E. Engdahl. 1995. *Information Packaging and Grammar Architecture.* In: J. N. Beckman (ed.) NELS 25. Vol. 1. University of Pennsylvania: 519-533

M. Vilkuna. 1989. *Free Word Order in Finnish: Its Syntax and Discourse Functions.* Helsinki: Suomalaisen Kirjallisuuden Seura

J. Wackernagel. 1892. *Über ein Gesetz der Indogermanischen Wortstellung.* Indogermanische Forschungen 1: 333-436

H. Walter. 1965. *Zur Funktion der sog. Kurzformen der Pronomina in der modernen bulgarischen Literatursprache*. In: Zeitschrift für Slawistik X (1): 107-17

H. Zeevat, E. Klein and J. Calder. 1987. *Unification Categorial Grammar*. In: Haddock, Klein and Morrill (eds.) Categorial Grammar, Unification Grammar, and Parsing. Vol. 1. Edinburgh: Centre for Cognitive Science

A. Zwicky. 1977. *On Clitics*. Indiana University Linguistics Club. Bloomington

6 Appendix

Implementation of a Computer Grammar of Bulgarian

Of the grammatical theories using inheritance-based grammars—a wide spread tradition in the natural language processing community—the HPSG achieves the widest coverage.

HPSG-based formalisms are characterised by a systemic use of typing of feature structures for factoring out shared properties of linguistic objects, be they words, phrases, or anything else. A multiple inheritance scheme over these types is used to cross-classify linguistic objects. Typing is also used to restrict the application of general principles to the right class of linguistic structures.

In the following, I shall first consider the cross-classifications and the multiple-inheritance hierarchies used in the implementation of the computer grammar of Bulgarian. Then the grammar rules employed in the actual parser will be sketched, and some parsing results presented.

6.1 Lexical Hierarchy

There are three partitions of the type *word* (lexical sign) in the assumed hierarchy—cf. Figure 34.

According to the first, all words are classified, depending on whether they bear (primary) stress or not, into *prosodically strong* (i.e. stress-bearing, accentually autonomous, phonologically strong) and *prosodically weak* (i.e. clitic, accentually dependent, phonologically weak).

In the second partition each word is viewed as belonging either to a *substantive* or to a *functional* lexical category, which reflects the traditional distinction between major (e.g., verb, noun, adjective, etc.) and minor (e.g., prepositions, particles, conjunctions, etc.) categories.

It is important to note that there is a third partition of the lexical categories into *verbal* and *non-verbal*. Such an assumption is motivated by the different morphosyntactic behaviour of verbs and "non-verbs" in Bulgarian. One of the pioneers of Bulgarian grammatical studies, Alexander Teodorov-Balan, dubbed the verb "the elephant of Bulgarian grammar", for the verbal system in Bulgarian is notoriously complex, and a lot of the remarkable phenomena observed in the Bulgarian clause are related to the morphosyntax of the verbal categories.

Both main verbs, which are instances of the types *substantive* and *verbal*, and auxiliaries, which are assumed to be instances of the types *functional* and *verbal*, can head a morphosyntactic complex in Bulgarian. On the contrary, all *non-verbal substantive* categories—nouns, adjectives, adverbs, etc.—form only syntactic constituents. This in particular means that the conventional HPSG distinction between lexical and phrasal categories is altogether relevant for Bulgarian non-verbal categories.

The main verbs are cross-classified according to their aspect as *imperfective*, *perfective* or *biaspectual*, with respect to their finiteness as *finite* or *non-finite*, and according to their type as *full*, *restructuring* (i.e. *modal* or *phasal*) or *copula*. The full verbs can further be *personal* or *impersonal*, *phrasal-experiencer* or *non-phrasal-experiencer*, and *medial* or *non-medial*. The first of these three partitions introduces basic subcategorisation properties, whereby personal verbs are divided into *transitive* and *intransitive* (with *strictly transitive* being a subclass of transitive verbs, and *strictly intransitive*—a subclass of intransitive verbs), while *subjectless* verbs are assumed to be a subclass of impersonal verbs. The second and the third partitions are related to the presence of lexical formants. Finite synthetic verb forms are classified according to their mood as *indicative*, *imperative* or *conditional*, and according to their lexical tense as *present*, *imperfect* and *aorist*. Respectively, the *passive participle* and the different types of *active participle* (i.e. *aorist participle*, *imperfect participle* and *present participle*), as well as the *gerund*, the *shortened infinitive* and the *verbal noun*, are all regarded as non-finite verb forms.

Two basic types of auxiliary verbs are distinguished, namely *future* (with subtypes *positive* future auxiliary and *negative* future auxiliary) and *be* (with subtypes be1 = *sâm* and be2 = *bâda*); the *negative imperative* auxiliary is an independent subtype of auxiliary. Auxiliary verbs are further classified according to their form as *auxiliary-participle* and *auxiliary-finite* (the latter with subtypes *present-auxiliary*, *aorist-auxiliary* and *imperfect-auxiliary*).

As to the non-verbal substantive categories, they are cross-classified according to the type of reference (*pronominal* or *non-pronominal*), the type of category (*noun, adjective, adverb, numeral*, etc.) and the presence of the definite-article morpheme (*unarticled* or *articled*; the *full* or the *short* form of the definite article is captured as a more specific partitioning within the articled categories).

The development of Bulgarian towards an analytic language type results in a decline of the case system which was typical of Old-Bulgarian, and is existent in most modern Slavic languages nowadays. Therefore, it would be hardly adequate to directly distinguish cases for the lexical category *noun*. On the other hand, relic case forms can certainly be recognised with pronouns (and especially with the personal ones) in modern Bulgarian. In the lexical sort hierarchy proposed here, a distinction is made between *base*, *oblique* and *vocative* forms of nominal categories, further allowing for *dative* and *accusative* oblique-form subtypes.

The traditional types of pronominals are assumed to be instances of the type *pronominal*, with personal pronouns further classified with respect to possessivity (i.e. *possessive* or *non-possessive*) and reflexivity (i.e. *reflexive* or *non-reflexive*). On such an approach, the generalisation can be captured that in Bulgarian there are pronominal nouns (e.g., personal—*az*, demonstrative—*tova*, indefinite—*njakoj*, interrogative—*koj*, negative—*nikoj*, summative—*vseki*, relative—*kojto*, etc.), pronominal adjectives (e.g., personal—*moj*, *svoj*, demonstrative—*takiva*, indefinite—*njakakvi*, interrogative—*kakvi*, negative—*nikakvi*, summative—*vsjakakvi*, relative—*kakvito*, etc.), pronominal adverbs (e.g., demonstrative—*taka*, indefinite—*njakak*, interrogative—*kak*, negative—*nikak*, summative—*vsjakak*, relative—*kakto*, etc.), pronominal numerals (e.g., demonstrative—*tolkova*, indefinite—*njakolko*, interrogative—*kolko*, negative—*nikolko*, relative—*kolkoto*, etc.), etc.

Particles, conjunctions, prepositions, interjections, etc., are considered non-verbal functional categories.

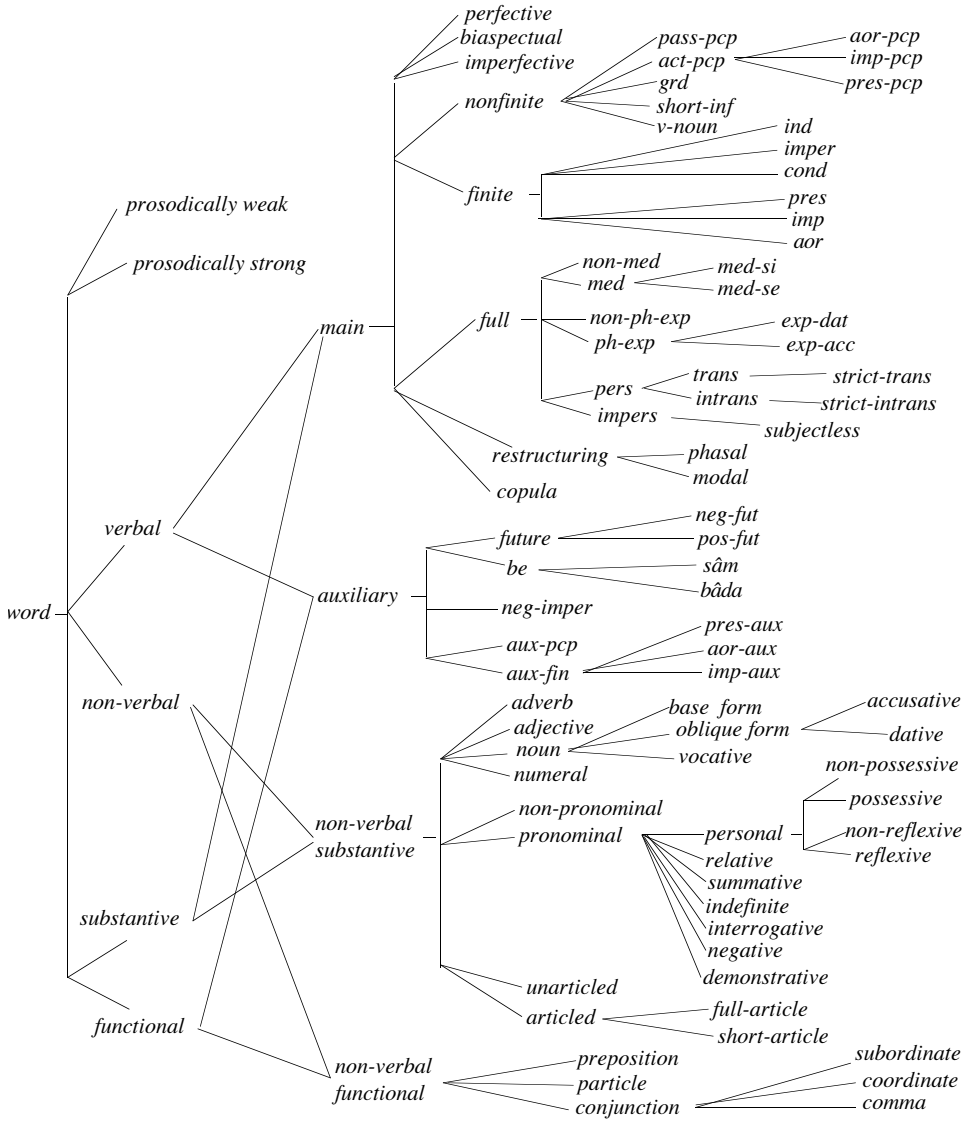
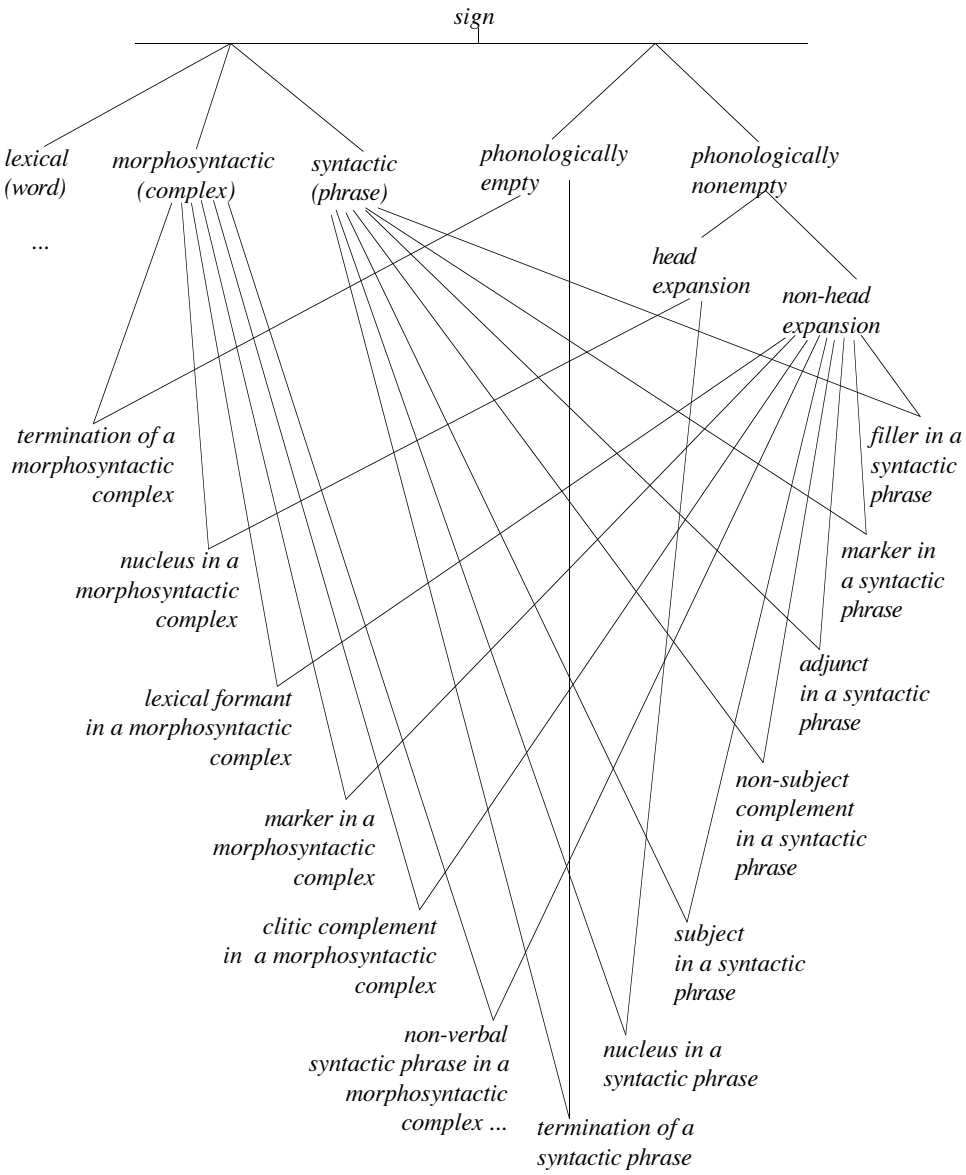

word
prosodically weak
prosodically strong
verbal
non-verbal
substantive
functional
main
auxiliary
non-verbal substantive
non-verbal functional
perfective
biaspectual
imperfective
nonfinite
finite
pass-pcp
act-pcp
grd
short-inf
v-noun
aor-pcp
imp-pcp
pres-pcp
ind
imper
cond
pres
imp
aor
full
restructuring
copula
non-med
med
med-si
med-se
non-ph-exp
ph-exp
exp-dat
exp-acc
pers
impers
trans
intrans
strict-trans
strict-intrans
subjectless
phasal
modal
future
be
neg-imper
aux-pcp
aux-fin
neg-fut
pos-fut
sâm
bâda
pres-aux
aor-aux
imp-aux
adverb
adjective
noun
numeral
base form
oblique form
vocative
accusative
dative
non-pronominal
pronominal
personal
relative
summative
indefinite
interrogative
negative
demonstrative
non-possessive
possessive
non-reflexive
reflexive
unarticled
articled
full-article
short-article
preposition
particle
conjunction
subordinate
coordinate
comma

Figure 34

6.2 Classification of Phrase-Structural Schemata



6.3 Grammar Rules

The actual parser, as implemented in a grammar-checker of Bulgarian[107], employs the following rules:[108]

```
01 Termination of a clause
02 Termination of a verb complex
03 Termination of an auxiliary complex
04 Lexical verb as nucleus in a clause
05 Verb complex as nucleus in a clause
06 Lexical verb as nucleus in a verb complex
07 Lexical auxiliary as nucleus in an auxiliary complex
08 Lexical auxiliary as marker in an auxiliary comlex
09 Auxiliary complex as marker in an auxiliary complex
10 Lexical auxiliary as marker in a verb complex
11 Auxiliary complex as marker in a verb complex
12 Imperative particle as marker in a verb complex
13 Conjunctive particle as marker in a verb complex
14 Future and negative particles as markers in a verb complex
15 Reflexive particle as marker in a verb complex
16 Experiencer formant in a verb complex
17 Pronominal clitic in a verb complex
18 Interrogative particle as marker in a verb complex
19 "Reflexive-passive" particle as marker in a verb complex
20 Future and negative particles as markers in an auxiliary complex
21 'Si'/'se' in an auxiliary complex
22 Non-reflexive pronominal clitic in an auxiliary complex
23 Interrogative particle as marker in an auxiliary complex
24 Adverb as adjunct in a clause
25 Termination of a noun phrase
26 Lexical noun as nucleus in a noun phrase
27 Lexical adjective as adjunct in a noun phrase
28 Preposition as marker in a noun phrase
29 Nominal subject in a clause
30 Clausal subject in a clause
31 Nominal complement in a clause
32 Clausal complement in a clause
33 Border element as marker in a clause
34 Subordinate conjunction as marker in a clause
```

[107]Within the framework of the joint European project LATESLAV: Language Processing Technologies for Slavic Languages—PECO 2824.

[108]Cf. *Avgustinova 1996* for a detailed documentation.

6.4 Sample Parses[109]

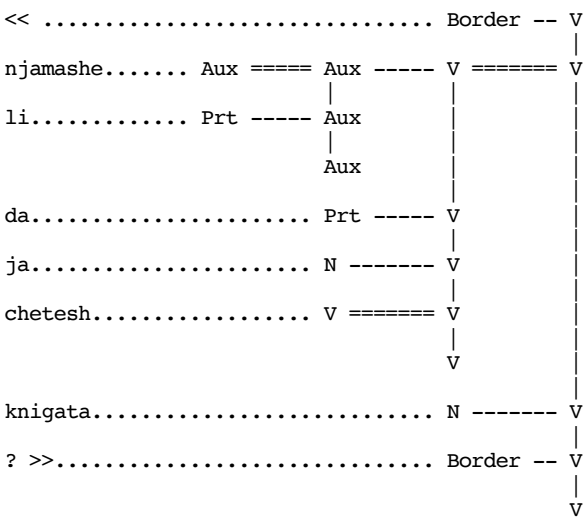
6.4.1 Parsing of Error-Free Input

INPUT:
[<< ,njamashe,li,da,ja,chetesh,knigata,? >>]

OUTPUT:

2 parse(s): future or past_future

Resulting structure:



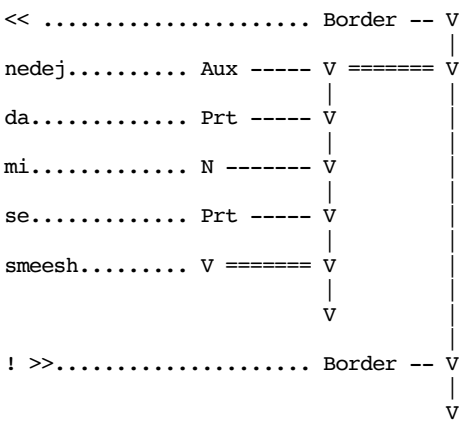
[109]The transliteration of the Cyrillic characters in the parser deviates from the one adopted in the previous chapters in the following way: č = ch; š = sh; št = sht; â = y.

INPUT:
[<< ,nedej,da,mi,se,smeesh,. >>]

OUTPUT:

1 parse(s)

Resulting structure:

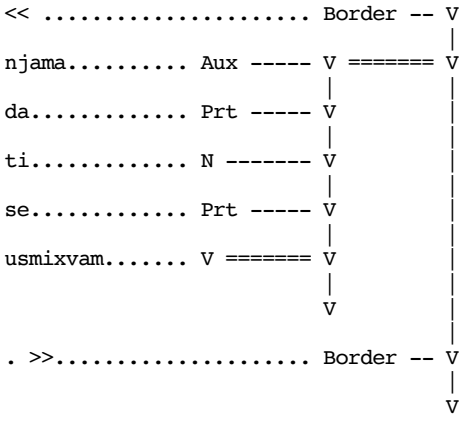


INPUT:
[<< ,njama,da,ti,se,usmixvam,. >>]

OUTPUT:

1 parse(s)

Resulting structure:

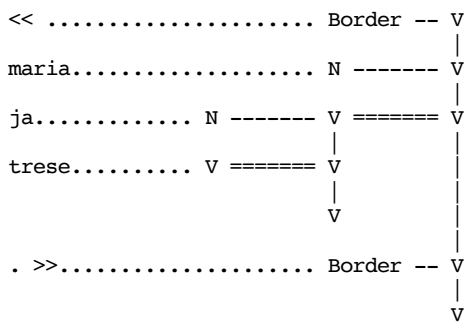


INPUT:
[<< ,maria,ja,trese,. >>]

OUTPUT:

2 parse(s): present or aorist

Resulting structure:

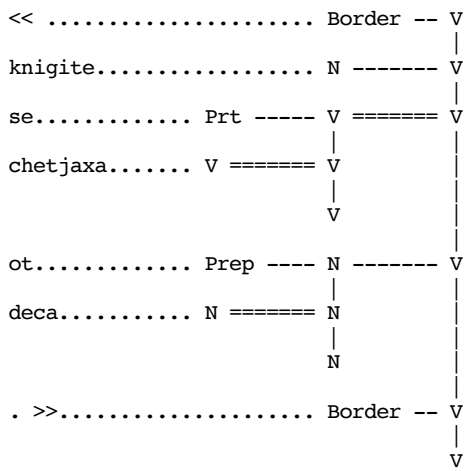


INPUT:
[<< ,knigite,se,chetjaxa,ot,deca,. >>]

OUTPUT:

1 parse(s)

Resulting structure:

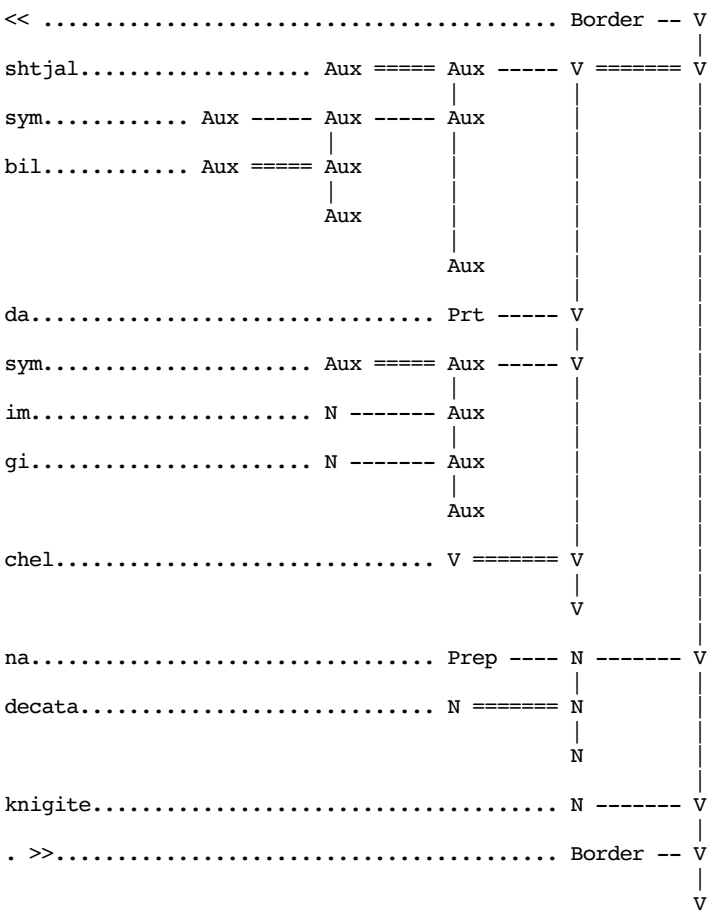


INPUT:
[<< ,shtjal,sym,bil,da,sym,im,gi,chel,na,decata,knigite,. >>]

OUTPUT:

2 parse(s): future_perfect or past_future_perfect

Resulting structure:

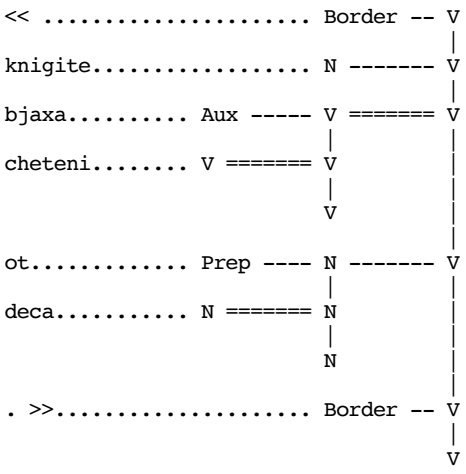


INPUT:
[<< ,knigite,bjaxa,cheteni,ot,deca,. >>]

OUTPUT:

3 parse(s): imperfect or aorist or pluperfect

Resulting structure:

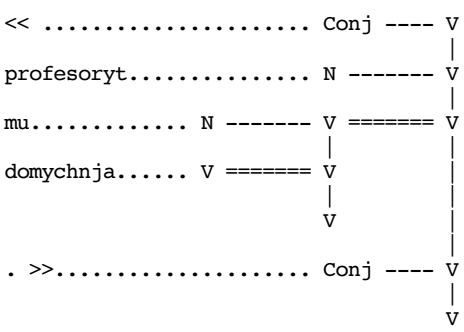


6.4.2 Parsing with Error Detection

INPUT:
[<< ,profesoryt,mu,domychnja,. >>]

OUTPUT:

Resulting structure:

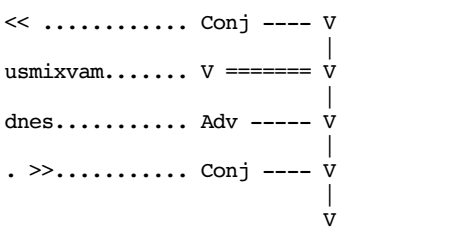


Error: [The experiencer object has possibly to be marked by the preposition "na" in combination with a short article !]
in the phrase: list([<< ,profesoryt,mu,domychnja,. >>])

INPUT:
[<< ,usmixvam,dnes,. >>]

OUTPUT:

Resulting structure:

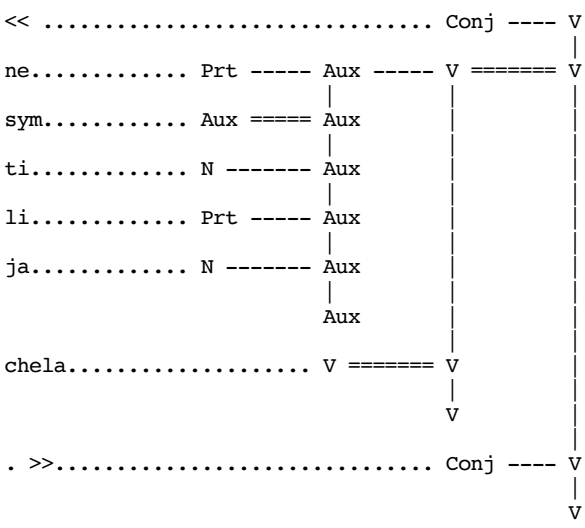


Error: [missing reflexive particle]
in the phrase: list([<< ,usmixvam,dnes,. >>])

INPUT:
[<< ,ne,sym,ti,li,ja,chela,. >>]

OUTPUT:

Resulting structure:

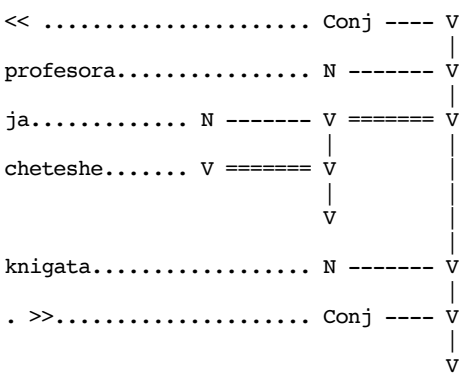


Error: [Possibly wrong placement of the interrogative particle]
in the phrase: list([<< ,ne,sym,ti,li,ja,chela,. >>])

INPUT:
[<< ,profesora,ja,cheteshe,knigata,. >>]

OUTPUT:

Resulting structure:

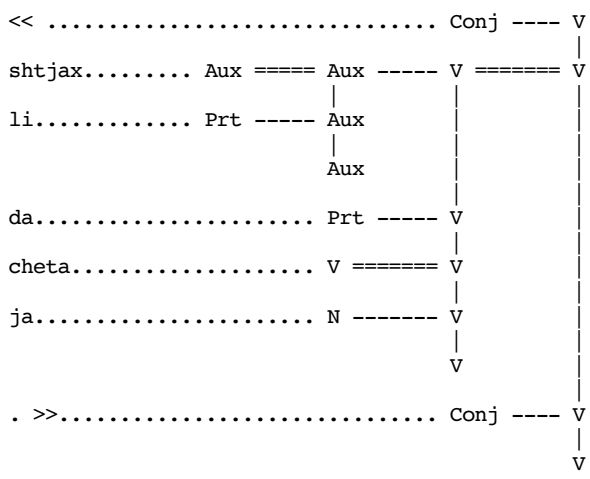


Error: [wrongly used short_form definite article in subject]
in the phrase: list([<< ,profesora,ja,cheteshe,knigata,. >>])

INPUT:
[<< ,shtjax,li,da,cheta,ja,. >>]

OUTPUT:

Resulting structure:

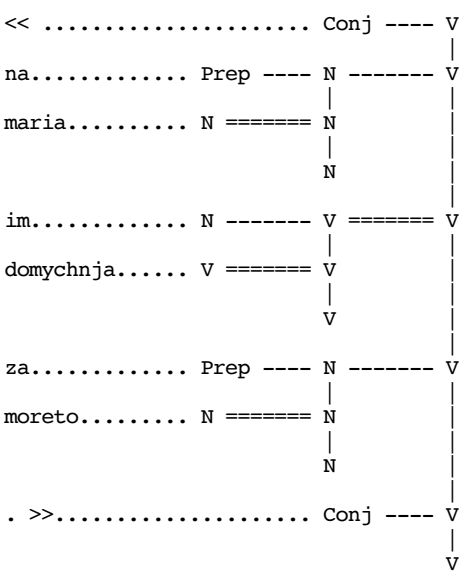


Error: [wrong word order after "da"]
in the phrase: list([shtjax,li,da,cheta,ja])

INPUT:
[<< ,na,maria,im,domychnja,za,moreto,. >>]

OUTPUT:

Resulting structure:



Error: [Possibly disagreement between an object and a clitic]
in the phrase: list([<< ,na,maria,im,domychnja,za,moreto,. >>])